REPORT

AUTOMATING SOCIETY

2020

ALGORITHM WATCH / CH

BertelsmannStiftung

ENGAGEMENT
EIN FÖRDERFONDS DER MIGROS-GRUPPE

SCHWEIZER AUSGABE
EDITION SUISSE
EDIZIONE SVIZZERA
SWISS EDITION

IMPRESSUM | INFORMATIONS LÉGALES | IMPRINT

Automating Society Report 2020

Länderausgabe Schweiz, Januar 2021

Edition suisse, Janvier 2021

Edizione svizzera, Gennaio 2021

Country issue Switzerland, January 2021

Online abrufbar unter | Disponible en ligne à | Disponibile online all'indirizzo | Available online at
https://automatingsociety.algorithmwatch.org/report2020/switzerland/

Herausgeber·innen | Éditeur·trices | Edito da | Editors
Fabio Chiusi
Sarah Fischer
Nicolas Kayser-Bril
Anna Mätzener
Matthias Spielkamp

Projektmanagement | Gestion de projet | Gestione del progetto | Project management
Fabio Chiusi

Koordination | Coordination | Coordinamento | Coordination
Marc Thümmler

Comics | Bandes dessinées | Fumetti | Comics
Samuel Daveti
Lorenzo Palloni
Allessio Ravazzani

Übersetzung | Traduction | Traduzione | Translation
Katrin Harlass (Deutsch)
Charles Robert (Français)
Fabio Chiusi (Italiano)

Gestaltung | Mise en page | Layout
Beate Autering
Beate Stangl

Redaktionsassistenz | Rédaction additionnelle | Ulteriore revisione dei testi | Additional editing
Leonard Haas

Redaktionsschluss: 30. September 2020
Date de mise en page: 30 septembre 2020
Aggiornato al 30 settembre 2020
Editorial deadline: 30 September 2020

Der Automating Society Report ist eine Gemeinschaftsproduktion von AlgorithmWatch und der Bertelsmann Stiftung.
Die Länderausgabe Schweiz wurde ermöglicht durch die finanzielle Unterstützung von Engagement Migros.

L'Automating Society Report est une production conjointe d'AlgorithmWatch et de la Bertelsmann Stiftung.
L'édition suisse a été rendue possible grâce au soutien financier d'Engagement Migros.

Il rapporto Automating Society è una produzione congiunta di AlgorithmWatch e della Bertelsmann Stiftung.
L'edizione svizzera è stata resa possibile dal sostegno finanziario di Engagement Migros.

The Automating Society Report is a collaborative production of AlgorithmWatch and the Bertelsmann Stiftung.
The Swiss country edition was made possible by the financial support of Engagement Migros.

AlgorithmWatch CH
Spindelstrasse 2
CH-8041 Zürich
https://algorithmwatch.ch/

Inhalt

Table des matières

Indice

Contents

Systeme zum automatisierten Entscheiden sind im Alltag angekommen. Wie gehen wir damit um?

Von Fabio Chiusi

Redaktionsschluss für diesen Bericht war der 30. September 2020. Spätere Entwicklungen konnten nicht berücksichtigt werden.

Es war ein bewölkter Augusttag in London, und die Schüler·innen waren wütend. Zu Hunderten strömten sie protestierend zum Parliament Square. Auf ihren Plakaten prangten Unterstützerslogans für ungewöhnliche Verbündete: ihre Lehrer·innen. Und daneben ein noch weit ungewöhnlicheres Angriffsziel: ein Algorithmus.

Wegen der Corona-Pandemie hatte das Vereinigte Königreich im März die Schulen geschlossen. Das Virus hielt auch den Sommer über ganz Europa in Atem, und den Schüler·innen war klar, dass ihre Abschlussprüfungen ausfallen und ihre Noten – irgendwie – anders ermittelt werden würden. Was sie sich allerdings nicht hatten vorstellen können, war, dass Tausende von ihnen schlechter abschneiden würden als erwartet.

Ihren Schildern und Sprechgesängen zufolge wussten die protestierenden Schüler·innen offenbar ganz genau, wer daran schuld war: das vom Office of Qualifications and Examinations Regulation (Ofqual) genutzte ADM-System (automated decision-making system). Die Behörde plante, die bestmöglichen datenbasierten Einschätzungen sowohl für General Certificates of Secondary Education (Mittlere Reife) als auch A-Level-Ergebnisse (Abiturzeugnisse) zu generieren, und zwar so, dass „die Verteilung der Noten einem Muster folgt, das dem vorangegangener Jahre ähnelt, so dass die Schüler·innen dieses Jahrgangs infolge der besonderen Umstände keine systematischen Nachteile erleiden".

Die Regierung wollte überoptimistische[1] Benotungen, die aus einem ausschliesslich menschlichen Urteil erwachsen, vermeiden: Ihrer eigenen Einschätzung zufolge wären die Abschlüsse, verglichen mit denen der Vergangenheit, zu gut ausgefallen. Doch dieser Versuch, „mit Schüler·innen, die in diesem Sommer ihre Prüfungen nicht ablegen konnten, so fair wie möglich umzugehen", scheiterte spektakulär. Und an diesem grauen Protesttag im August riss der Strom der Protestierenden nicht ab, die weiterhin Sprechchöre bildeten und ihre Plakate hochhielten, um soziale Gerechtigkeit einzufordern. Manche waren ausser sich, andere weinten.

„Stop stealing our future", stand auf einem der Plakate, eine Anleihe bei den Protesten der Klimaaktivist·innen von Fridays for Future: Hört auf, uns unsere Zukunft wegzunehmen. Andere Slogans waren etwas genauer auf die Mängel des ADM-Benotungssystems zugeschnitten: Sprüche wie „Benotet meine Arbeit, nicht meine Postleitzahl" oder „Wir sind Schüler·innen, keine statistischen Grössen" prangerten die diskriminierenden Ergebnisse an, die es geliefert hatte[2].

Schliesslich stimmte die Menge einen lauten Sprechchor an, der künftig bei Protesten zum festen Kanon gehören dürfte: „Fuck the algorithm". Aus Angst, die Regierung automatisiere eben mal so – und auf undurchsichtige Weise – ihre Zukunft, ganz gleich, ob die Benotung ihre tatsächlichen Fähigkeiten und Anstrengungen widerspiegelte oder nicht, forderten die Schüler·innen das Recht ein, sich ihre Zukunftschancen nicht über Gebühr von einem fehlerhaften Code vermasseln zu lassen. Sie wollten mitreden, und was sie zu sagen hatten, sollte gehört werden.

Algorithmen sind weder „neutral" noch „objektiv", auch wenn wir das gerne glauben wollen. Sie reproduzieren die Annahmen und Überzeugungen jener, die sich dafür entscheiden, sie einzusetzen und zu programmieren. Aus diesem Grund sind für die Auswahl guter wie schlechter Algorithmen Menschen verantwortlich (oder sollten es sein) und nicht „Algorithmen" oder ADM-Systeme. Maschinen mögen uns Angst machen, doch der Geist, der ihnen innewohnt, ist stets ein menschlicher. Und Menschen sind komplizierte Wesen, komplizierter noch als Algorithmen.

Ohnehin waren die protestierenden Schüler·innen nicht so naiv zu glauben, ihre Sorgen seien lediglich dem Versagen eines Algorithmus anzulasten. Ihre Sprechgesänge richteten sich nicht in einem Anfall technologischen Determinismus' gegen „den Algorithmus"; ihre Motivation war vielmehr das dringende Bedürfnis, soziale Gerechtigkeit zu schützen und voranzutreiben. In dieser Hinsicht ähnelt ihr Protest viel eher dem der Ludditen. Ebenso wie den Anhänger·innen der Gewerkschaftsbewegung, die im 19. Jahrhundert mechanisierte Webstühle und Strickrahmen zerstörten, ist ihnen klar, dass ADM-Systeme etwas mit Macht zu tun haben und nicht irrtümlicherweise für eine angeblich objektive Technologie gehalten werden sollten. Also riefen sie „Gerechtigkeit für die Arbeiterklasse", forderten den Rücktritt des Gesundheitsministers und brandmarkten ADM-Systeme als „Klassismus in Reinkultur" und „eklatanten Klassismus". Letztendlich hatten die Schüler·innen Erfolg. Sie erreichten eine Abschaffung des Systems, das ihren weiteren Bildungsweg und ihre Lebenschancen bedrohte: In einer spektakulären Kehrtwende verwarf die britische Regierung

1 „Die Forschungsliteratur legt nahe, dass bei der Einschätzung von Noten, die Schüler·innen aller Wahrscheinlichkeit nach erreichen, Lehrer·innen zu Optimismus neigen (allerdings nicht in allen Fällen)", schreibt Ofqual, vgl. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/909035/6656-2_-_Executive_summary.pdf

2 Für weitere Details vgl. das Kapitel United Kingdom in der Gesamtausgabe des Automating Society Reports unter https://automatingsociety.algorithmwatch.org/report2020/united-kingdom/

das fehleranfällige ADM-System und wandte die von den Lehrer·innen prognostizierte Benotung an.

Doch hinter dieser Geschichte steckt mehr als nur die Tatsache, dass die Protestierenden am Ende ihr Ziel erreichten. Dieses Beispiel verdeutlicht, wie schlecht entwickelte, implementierte und überwachte ADM-Systeme, die menschliche Voreingenommenheit und Diskriminierung reproduzieren, dazu führen, dass das ihnen innewohnende Potenzial – etwa als Hebel zu dienen, Vergleichbarkeit und Fairness zu fördern – ungenutzt bleibt.

Stärker als viele Kämpfe der Vergangenheit macht dieser Protest bewusst, dass wir längst nicht mehr nur dabei sind, die Gesellschaft zu automatisieren. Wir haben sie bereits automatisiert – und endlich hat es jemand bemerkt.

/ Von der Automatisierung der Gesellschaft zur automatisierten Gesellschaft

Als wir die erste Auflage dieses Berichts publizierten, entschieden wir uns, ihn „Automating Society" zu nennen, denn ADM-Systeme waren in Europa grösstenteils neu, in Testphasen und noch unerforscht. Vor allem aber waren sie eher die Ausnahme als die Norm.

Das hat sich dramatisch verändert. Wie die zahlreichen Fälle, die von unserem Expertin·innen-Netzwerk in diesem Bericht zusammengetragenen wurden, deutlich zeigen, hat sich der Einsatz von ADM-Systemen in nur etwas mehr als einem Jahr um ein Vielfaches erhöht. ADM-Systeme wirken sich mittlerweile nicht nur auf alle möglichen menschlichen Aktivitäten aus, sondern auch, und das ist am bedeutsamsten, auf Dienstleistungen für Millionen EU-Bürger·innen – und deren Zugang zu ihren Rechten.

Während immer mehr ADM-Systeme eingesetzt werden, bleibt die Intransparenz, die sie umgibt, bestehen. Deshalb ist es dringender denn je, mehr Informationen ans Licht zu holen. Wir haben daher die Zahl der Länder, die wir untersucht haben, erhöht: Statt zwölf, wie in der ersten Fassung dieses Berichts, sind es nun 16; Estland, Griechenland, Portugal und die Schweiz sind hinzugekommen. Das ist zwar bei Weitem nicht erschöpfend, dennoch erlaubt es uns, ein umfassenderes Bild der derzeitigen Situation von ADM-Systemen in Europa zu zeichnen. Wir sind überzeugt, dass diese Arbeit unerlässlich ist. Denn diese Systeme können Auswirkungen nicht nur auf unser Alltagsleben haben. Sie werfen auch die Frage auf, wie sich Automatisierung auf unsere Institutionen, unsere Normen und Werte, ja auf die Demokratie insgesamt auswirkt.

Dies gilt insbesondere während der COVID-19-Pandemie, in einer Zeit also, in der viele ADM-Systeme (zumeist übereilt) eingeführt wurden, die darauf ausgerichtet sind, mit Hilfe von datenbasierten Tools und Automatisierung zum Schutz der öffentlichen Gesundheit beizutragen. Wir hielten diese Entwicklung für derart bedeutend, dass wir uns entschieden haben, ihr im Rahmen des Projektes „Automating Society" eine „Sonderausgabe" zu widmen, die im September 2020 als „Vorabbericht" veröffentlicht wurde[3].

Selbst in Europa scheint es beim Einsatz von ADM-Systemen kein Halten mehr zu geben. Das wird an den in diesem Bericht aufgeführten Fällen deutlich, die zu den vielen, über die wir bereits in der Vorgängerausgabe berichtet haben, hinzugekommen sind – vom Wohlfahrtssektor über die Bildung und das Gesundheitswesen bis hin zur Justiz. Auf den folgenden Seiten geben wir zum ersten Mal ein Update zu den aktuellen Entwicklungen in diesen Fällen, und zwar auf dreierlei Weise. Erstens in journalistischen Stories; zweitens in Abschnitten, in denen verschiedene Beispiele katalogisiert werden; und drittens mit Hilfe von Comics. Unserer Auffassung nach sind diese ADM-Systeme in unser aller Leben bereits derart massgebend – und werden es zunehmend sein – dass wir unbedingt versuchen müssen, alle Zielgruppen zu erreichen, und zwar auf eine neue Weise. Wir müssen erläutern, wie sie funktionieren und was sie *mit uns machen.* Denn letztlich wirken sich ADM-Systeme auf jeden und jede Einzelne von uns aus.

Oder sie könnten es zumindest. Wir haben zum Beispiel gesehen, wie in Estland ein neuer, automatisierter, proaktiver Service Familienzuschüsse verteilt. Eltern müssen nicht einmal mehr einen Antrag stellen: Von Geburt an sammelt der Staat alle Informationen zu jedem einzelnen Neugeborenen und dessen Eltern und führt sie in Datenbanken zusammen. Dies sorgt dafür, dass Eltern automatisch alle finanziellen Hilfen erhalten, die ihnen rechtlich zustehen.

In Finnland erfolgt mit Hilfe eines vom japanischen Grosskonzern Fujitsu entwickelten Tools eine automatisierte Identifizierung individueller Risikofaktoren, die mit der gesellschaftlichen Stigmatisierung junger Erwachsener verbunden sind. In Frankreich ist es möglich, Daten aus

3 ‚Automated Decision-Making Systems in the COVID-19 Pandemic: A European Perspective', https://algorithmwatch.org/en/project/automating-society-2020-covid19/

sozialen Netzwerken zu sammeln und Algorithmen des Maschinellen Lernens mit ihnen zu füttern, um Steuerbetrug aufzudecken.

Italien experimentiert mit „voraussagender Rechtsprechung". Hierbei kommt Automatisierung zum Einsatz, die Richter·innen aufzeigt, wie frühere Gerichtsurteile zum jeweiligen Streitgegenstand tendenziell ausgefallen sind. Und in Dänemark versuchte die Regierung, jede Tastatureingabe und jeden Mausklick zu überwachen, die Student·innen bei Prüfungen auf ihren Computern vornahmen. Dies löste – auch hier – massive Proteste der Studierenden aus, die dazu führten, dass das System vorerst wieder eingestellt wurde.

/ Es ist an der Zeit, Fehlentwicklungen bei ADM-Systemen zu korrigieren

ADM-Systeme haben prinzipiell das Potenzial, das Leben der Menschen positiv zu beeinflussen – indem sie riesige Datenmengen verarbeiten, Verantwortliche in Entscheidungsfindungsprozessen unterstützen und massgeschneiderte Anwendungen bieten.

In der Praxis haben wir allerdings nur wenige überzeugende Beispiele für einen solchen positiven Einfluss gefunden.

In der Ausgabe von 2019 noch kaum erwähnt, wird Gesichtserkennung inzwischen mit alarmierender Geschwindigkeit in ganz Europa getestet und eingesetzt.

So ist etwa das System VioGén, das in Spanien seit 2007 im Einsatz ist, um Risiken in Fällen von häuslicher Gewalt abzuschätzen, zwar bei Weitem noch nicht perfekt, zeigt jedoch „vernünftige Leistungskennzahlen" und hat bereits geholfen, viele Frauen vor Gewalt zu schützen.

In Portugal hat ein zentralisiertes automatisiertes System zur Aufdeckung von Betrug im Zusammenhang mit ärztlichen Rezepten die Betrugsfälle innerhalb eines einzigen Jahres nachweislich um 80 % reduziert. In Slowenien hat sich den Aussagen der Steuerbehörden zufolge ein ähnliches System als nützlich erwiesen, das verwendet wird, um Steuerbetrug zu bekämpfen.[4]

Doch die aktuelle Bestandsaufnahme zum Einsatz von ADM-Systemen in Europa zeigt, dass positive Beispiele, die echte Vorteile bringen, rar gesät sind. Im gesamten Bericht beschreiben wir, wie die überwältigende Mehrheit der aktuellen Nutzungen Menschen tendenziell eher einem Risiko aussetzt als ihnen zu helfen. Um aber die tatsächlichen positiven und negativen Effekte beurteilen zu können, brauchen wir mehr Transparenz im Hinblick auf die Ziele, mehr Daten über die Funktionsweise der ADM-Systeme, die derzeit getestet und eingesetzt werden.

Die Botschaft an die politischen Entscheidungsträger·innen könnte klarer nicht sein: Wollen wir ihr Potenzial wirklich ausschöpfen und zugleich dafür sorgen, dass Menschenrechte und Demokratie respektiert werden, dann ist es an der Zeit, diese Systeme transparent zu machen und die Fehlentwicklungen bei ADM zu korrigieren. Jetzt.

/ Gesichtserkennung, überall Gesichtserkennung

In verschiedenen Ländern kommen verschiedene Tools zum Einsatz. Eine Technologie ist heute aber beinahe flächendeckend zu finden: Gesichtserkennung. Hierbei handelt es sich ohne Zweifel um die neueste, rasanteste und besorgniserregendste Entwicklung, die in diesem Bericht beleuchtet wird. In der Ausgabe von 2019 noch kaum erwähnt, wird Gesichtserkennung inzwischen mit alarmierender Geschwindigkeit in ganz Europa getestet und eingesetzt. Innerhalb von wenig mehr als einem Jahr seit unserem letzten Bericht ist Gesichtserkennung in Schulen und Stadien, an Flughäfen, ja selbst in Spielcasinos zu finden.

4 Weitere Details siehe Kapitel Slovenia in der Gesamtausgabe des Automating Society Reports unter https://automa tingsociety.algorithmwatch.org/report2020/slovenia/

Sie kommt ebenfalls für die vorhersagende Polizeiarbeit (Predictive Policing) zum Einsatz, um Kriminelle abzuschrecken, um gegen Rassismus vorzugehen sowie, im Hinblick auf die COVID-19-Pandemie, zur Durchsetzung des Social Distancing, und zwar sowohl in Form von Apps, wie auch durch „smarte" Videoüberwachung.

Ständig kommen neue Einsatzgebiete hinzu, obwohl die Zahl von Belegen für ihre mangelnde Genauigkeit zunimmt. Und gibt es Widerstand dagegen, dann versuchen Verfechter·innen dieser Systeme ganz einfach, ihn zu umgehen. In Belgien ist ein von der Polizei genutztes Gesichtserkennungssystem immer noch „teilweise aktiv", obwohl die Aufsichtsbehörde für Polizeilichen Informationsaustausch ein befristetes Verbot erlassen hat. Und in Slowenien wurde die Nutzung von Gesichtserkennungstechnologie durch die Polizei fünf Jahre, nachdem sie erstmals eingesetzt wurde, legalisiert.

Stellt man sich diesem Trend nicht entgegen, besteht die Gefahr, dass die Vorstellung, ununterbrochen – und verdeckt – beobachtet zu werden, zur Normalität wird, wodurch sich ein neuer Status quo von flächendeckender Massenüberwachung verfestigt. Genau aus diesem Grund hätten sich viele Akteur·innen der Bürgerrechtsbewegung eine weitaus aggressivere politische Antwort von Seiten der EU-Institutionen auf diese Entwicklungen gewünscht.[5]

Im Rahmen eines Pilotprojektes, bei dem derzeit ein ADM-System in Banken in Polen zum Einsatz kommt, spielt sogar das Lächeln eine Rolle: Je mehr die Angestellten lächeln, desto höher ist ihr Bonus. Und es werden nicht nur Gesichter überwacht. In Italien gab es den Vorschlag, in allen Fussballstadien ein Geräuschüberwachungssystem zu installieren, um Rassismus zu bekämpfen.

/ Blackboxes sind immer noch Blackboxes

Zu den alarmierenden Ergebnissen dieses Berichts gehört, dass zwar immer mehr ADM-Systeme eingesetzt werden, aber die Transparenz nicht Schritt hält. Im Jahr 2015 prägte Frank Pasquale, Professor an der Brooklyn Law School, einen berühmt gewordenen Begriff. Er nannte eine Gesellschaft, die von Netzwerken durchzogen ist, die auf intransparenten algorithmischen Systemen basieren, eine „Black Box Society". Fünf Jahre später trifft diese Metapher leider immer noch zu – und sie gilt ausnahmslos für alle Länder, die wir für diesen Bericht untersucht haben: Es gibt im Hinblick auf ADM-Systeme viel zu wenig Transparenz, sowohl im öffentlichen wie auch im privaten Sektor. Polen verfügte Intransparenz sogar von ganz oben – mit dem Gesetz zur Einführung eines automatisierten Systems zum Aufspüren von Bankkonten, die für illegale Aktivitäten genutzt werden („STIR"). Es sieht vor, dass mit bis zu 5 Jahren Gefängnis bestraft werden kann, wer Algorithmen und Risikoindikatoren offenlegt.

Zwar lehnen wir die Vorstellung, dass alle solche Systeme von Natur aus schlecht sind, nachdrücklich ab und nehmen stattdessen eine evidenzbasierte Perspektive ein. Dennoch ist es ohne Zweifel höchst problematisch, nicht in der Lage zu sein, ihre Funktionsweise und Auswirkungen auf der Grundlage echter Sachkenntnis und korrekter Fakten zu beurteilen. Schon deshalb, weil Intransparenz das Sammeln von Beweisen, die nötig sind, um überhaupt ein informiertes Urteil zur Nutzung von ADM-Systemen abgeben zu können, ernsthaft behindert.

Nimmt man dann noch die Schwierigkeiten hinzu, die sowohl unsere Wissenschaftler·innen wie auch unsere Journalist·innen hatten, wenn sie Zugang zu aussagekräftigen Daten über diese Systeme haben wollten, dann zeichnet sich hier für alle ein beunruhigendes Szenario ab, die solche Systeme kontrollieren und sicherstellen wollen, dass ihr Einsatz mit grundlegenden Rechten, gesetzlichen Regelungen und der Demokratie in Einklang steht.

/ Den algorithmischen Status quo hinterfragen

Was tut die Europäische Union in dieser Frage? Auch wenn die von der EU-Kommission unter der Leitung von Ursula von der Leyen erarbeiteten Dokumente sich eher auf „Künstliche Intelligenz" beziehen als ADM-Systeme direkt anzusprechen, enthalten sie lobenswerte Absichten: Eine „vertrauenswürdigen KI" soll gefördert und umgesetzt werden, die „die Menschen in den Vordergrund stellt"[6].

Allerdings gibt die EU, wie im Europa-Kapitel beschrieben, der vermeintlichen kommerziellen und geopolitischen Notwendigkeit, die „KI-Revolution" anzuführen, Vorrang davor, sicherzustellen, dass die Ergebnisse solcher KI demokratischer Kontrolle unterliegen.

5 Wie ausführlich im Europa-Kapitel dargelegt.

6 Vgl. das Europa-Kapitel, dort insbesondere den Abschnitt zum „Weissbuch KI" der EU-Kommission.

Zu den alarmierenden Ergebnissen dieses Berichts gehört, dass zwar immer mehr ADM-Systeme eingesetzt werden, aber die Transparenz nicht Schritt hält.

Dieser Mangel an politischem Mut wird daran am deutlichsten, dass im Weissbuch zu KI jeder Hinweis auf ein Moratorium für Systeme getilgt wurde, die an öffentlichen Orten Technologien zur Gesichtserkennung in Echtzeit einsetzen. Er ist überraschend – besonders zu einer Zeit, in der sich viele Mitgliedsstaaten im Zusammenhang mit allzu hastig implementierten ADM-Systemen, die sich negativ auf die Rechte von Büger·innen ausgewirkt haben, mit einer zunehmenden Anzahl gerichtlicher Klagen – und Niederlagen – konfrontiert sehen.

Ein besonders wegweisender Fall stammt aus den Niederlanden. Dort zogen Bürgerrechtsaktivist·innen gegen ein invasives und intransparentes automatisiertes System zur Aufdeckung von Sozialbetrug (SyRI) vor Gericht und gewannen. Das Gericht in Den Haag urteilte, dass das System eine Verletzung der Europäischen Menschenrechtskonvention darstelle, und stoppte es. Die Sache wurde auch zu einem Präzedenzfall: Dem Urteil zufolge haben Regierungen eine „besondere Verantwortung", Menschenrechte zu schützen, wenn sie solche ADM-Systeme implementieren. Für die so dringend benötigte Transparenz zu sorgen, wird als essenzieller Bestandteil dieser Verantwortung angesehen.

Seit unserem ersten Bericht haben sich die Medien und die Aktivist·innen der Zivilgesellschaft als Triebkräfte etabliert, die die Rechenschaftspflicht von ADM-Systemen einfordern. So ist es zum Beispiel in Schweden Journalist·innen gelungen, die Offenlegung des Codes zu erzwingen, der hinter dem Trelleborg-System für vollautomatisierte Entscheidungen über Anträge auf Sozialhilfe steht. In Berlin scheiterte das am Bahnhof Südkreuz durchgeführte Pilotprojekt zur Gesichtserkennung, und es kam nirgendwo in Deutschland zu einer Implementierung des Systems. Dies war auf den lautstarken Widerspruch von Aktivist·innen zurückzuführen – so lautstark, dass es ihnen gelang, Parteipositionen zu beeinflussen und damit letztendlich die politische Agenda der Regierung.

Griechische Aktivist·innen von Homo Digitalis deckten auf, dass an Versuchen mit dem griechischen Pilotprojekt eines Systems namens ‚iBorderCtrl', einem von der EU finanzierten Projekt zum Einsatz von ADM bei Grenzkontrollen, kein einziger echter Reisender teilgenommen hatte und enthüllten damit, dass die Fähigkeiten vieler solcher Systeme häufig übertrieben dargestellt werden. In Dänemark wurde zwischenzeitlich dank der Arbeit von Akademiker·innen und Journalist·innen sowie der Datenschutzbehörde (DPA) ein Profiling-System gestoppt, das Risiken erkennen sollte, die mit vulnerablen Familien und Kindern in Verbindung gebracht werden (das sogenannte „Gladsaxe-Modell").

Auch in anderen Ländern spielten die Datenschutzbehörden selbst eine wichtige Rolle. In Frankreich befand die Nationale Datenschutzbehörde sowohl ein Projekt zur Geräuschüberwachung wie auch eines zur Gesichtserkennung an Oberschulen für illegal. In Portugal weigerte sich die Datenschutzbehörde, den Einsatz eines Videoüberwachungssystems durch die Polizei in den Städten Leiria und Portimão zu genehmigen, da sie es für überdimensioniert hielt. Sein Einsatz hätte nicht nur zu einer „grossflächigen und systematischen Überwachung und Verfolgung von Menschen, ihren Gewohnheiten und ihres Verhaltens" geführt, sondern auch „zur Identifizierung von Menschen anhand von Daten im Zusammenhang mit physischen Merkmalen". Und in den Niederlanden forderte die staatliche Datenschutzbehörde mehr Transparenz bei voraussagenden Algorithmen, die von Regierungsbehörden genutzten werden.

Einige Länder wandten sich schliesslich mit der Bitte um Rat an eine Ombudsperson. In Dänemark leistete diese Form der Beratung einen Beitrag zur Entwicklung von Strategien und ethischen Richtlinien beim Einsatz von ADM-Systemen

im öffentlichen Sektor. In Finnland stufte die stellvertretende parlamentarische Ombudsperson automatisierte Steuerschätzungen als illegal ein.

Und dennoch fragt man sich angesichts des ungebremsten Einsatzes solcher Systeme in ganz Europa nach wie vor verwundert: Reicht dieses Mass an Kontrolle aus? Als die Ombudsperson in Polen die Legalität des in einer Bank genutzten (und oben erwähnten) Lächel-Detektors in Zweifel zog, verhinderte diese Entscheidung weder ein späteres Pilotprojekt in der Stadt Sopot, noch hielt es mehrere Unternehmen davon ab, ihr Interesse an einer Übernahme des Systems zu bekunden.

/ Es fehlt an allem: Kontrolle, Durchsetzung, Kompetenzen und Erläuterungen

Aktivismus ist ein grösstenteils reaktives Unterfangen. Aktivist·innen können meist nur dann reagieren, wenn ADM-Systeme in der Erprobung sind oder bereits zum Einsatz kommen. In dem Zeitraum, den Bürger·innen brauchen, um Widerstand zu organisieren, könnten ihre Rechte bereits unnötigerweise verletzt worden sein. Und dies kann trotz der Schutzmechanismen geschehen, die in den meisten Fällen durch EU-Gesetze oder Gesetze der Mitgliedsstaaten garantiert sein sollten. Aus diesem Grund sind proaktive Massnahmen zum Schutz von Bürgerrechten so überaus wichtig – bevor Pilotprojekte gestartet und Systeme eingesetzt werden.

Dennoch werden selbst in jenen Ländern, in denen Schutzgesetze existieren, diese schlicht nicht durchgesetzt. So ist etwa in Spanien „automatisiertes Verwaltungshandeln" gesetzlich geregelt, und es gelten besondere Anforderungen im Hinblick auf Qualitätskontrolle und behördliche Aufsicht. Festgeschrieben ist ebenfalls eine Auditierung des Informationssystems und seines Quellcodes. Zudem hat Spanien ein Gesetz zur Informationsfreiheit. Ungeachtet dessen geben öffentliche Einrichtungen, so schreibt unser Autor, nur selten detaillierte Informationen zu den von ihnen genutzten ADM-Systemen frei. Ähnliches gilt für Frankreich. Dort gibt es seit 2016 ein Gesetz, das algorithmische Transparenz verpflichtend macht, doch auch hier ohne jede Wirkung.

Unter Umständen reicht noch nicht einmal die gerichtliche Klage gegen einen Algorithmus auf der Grundlage spezieller Transparenzreglungen aus, um Rechte von Nutzer·innen durchzusetzen und zu schützen. Wie der Fall des Parcoursoup-Algorithmus in Frankreich zeigt, der von Universitäten genutzt wird, um Studienbewerber·innen zu sortieren, können nach Belieben Ausnahmen erarbeitet werden, um eine Verwaltung von der Rechenschaftspflicht auszunehmen.

Besonders beunruhigend ist dies dann, wenn es mit dem verbreiteten Mangel an Fähigkeiten und Kompetenzen rund um ADM-Systeme im öffentlichen Sektor einhergeht, wie er von vielen Fachleuten beklagt wird. Wie sollten Amtsträger·innen auch Auskunft geben oder für Transparenz sorgen im Zusammenhang mit Systemen, die sie nicht verstehen?

Einige Länder haben jüngst versucht, dieses Problem anzugehen. So wurde in Estland ein speziell auf ADM-Systeme ausgerichtetes Kompetenzzentrum gegründet. Dieses soll helfen, ein besseres Verständnis dafür zu entwickeln, wie solche Systeme bei der Weiterentwicklung öffentlicher Dienstleistungen zum Einsatz kommen können, und speziell dem Ministerium für Wirtschaft und Kommunikation sowie der Staatskanzlei Informationen und Grundlagen für ihr weiteres Vorgehen bei der Entwicklung von e-Government liefern. Ebenso hat die Schweiz im Rahmen ihrer breit angelegten nationalen Strategie für eine „Digitale Schweiz" ein „Kompetenznetzwerk" gefordert.

Ungeachtet dessen sind mangelhafte digitale Kompetenzen ein bekanntes Problem – mit Auswirkungen auf einen Grossteil der Bevölkerung etlicher europäischer Staaten. Abgesehen davon ist es ziemlich schwer, die Durchsetzung von Rechten einzufordern, von denen man nicht einmal weiss, dass man sie hat. Die Proteste im Vereinigten Königreich und anderswo sowie die bekannten Skandale im Zusammenhang mit der Nutzung von ADM-Systemen[7] haben mit Sicherheit das Bewusstsein für die Risiken und Chancen einer zunehmend automatisierten Gesellschaft erhöht. Allerdings ist dieses Bewusstsein, obgleich es zunimmt, in vielen Ländern nach wie vor unterentwickelt.

Die Ergebnisse unserer Recherchen sind eindeutig: ADM-Systeme nehmen bereits Einfluss auf alle möglichen Aktivitäten und Beurteilungen, doch ihr Einsatz erfolgt nach wie vor grösstenteils ohne jede vernünftige demokratische Debatte. Darüber hinaus ist es eher die Regel als die Ausnahme, dass die Durchsetzungs- und Aufsichtsmechanismen

7 Vgl. das Kapitel zu France in der Gesamtausgabe des Automating Society Reports unter https://automatingsociety.algorithmwatch.org/report2020/france/

– so sie denn überhaupt existieren – dieser Entwicklung hinterherhinken.

Selbst der Zweck dieser Systeme wird den betroffenen Bürger·innen gegenüber nicht generell gerechtfertigt oder erläutert, ganz zu schweigen von den Vorteilen, die sie bringen sollen. Man denke nur an den proaktiven Service „AuroraAI" in Finnland: Dieser soll, wie unsere finnischen Autor·innen berichten, automatisch „Lebensereignisse" identifizieren; nach den Vorstellungen seiner Befürworter·innen ist er dazu gedacht, als eine Art „Kindermädchen" zu fungieren, das Bürger·innen hilft, spezielle Anforderungen staatlicher Dienstleister im Zusammenhang mit bestimmten Lebensumständen zu erfüllen, wie etwa einem Umzug, veränderten Familienverhältnissen usw. Es könne hier „Nudging" am Werk sein, schreiben unsere Autor·innen weiter. Damit meinen sie, dass das System, anstatt Individuen zu ermächtigen, am Ende zum genauen Gegenteil führen könnte, dass es nämlich aufgrund seines speziellen Designs und seiner Architektur bestimmte Entscheidungen suggerieren oder die Optionen einer Person beschränken könnte.

Und deshalb ist es umso wichtiger zu wissen, was genau im Hinblick auf staatliche Dienstleistungen eigentlich „optimiert" werden soll: „Wird die Servicenutzung maximiert, werden Kosten minimiert, oder wird das Wohlergehen der Bürger·innen verbessert?", so die Frage der Forscher·innen. „Auf welchen Kriterien basieren diese Entscheidungen, und wer trifft sie?" Die blosse Tatsache, dass wir auf diese fundamentalen Fragen keine Antwort haben, spricht Bände über den Grad an erlaubter Beteiligung und Transparenz, selbst für ein potenziell invasives ADM-System wie dieses.

/ Die Falle des „technologischen Solutionismus"

Für all dies gibt es eine übergreifende ideologische Rechtfertigung. Sie wird „Technologischer Solutionismus" genannt und beeinflusst nach wie vor sehr stark die Art und Weise, wie viele der von uns untersuchten ADM-Systeme entwickelt werden. Auch wenn der Begriff schon lange angeprangert wird, weil er eine fehlerhafte Ideologie bezeichnet, die jedes gesellschaftliche Problem als „Fehler" begreift, der mit Hilfe von Technologie „repariert" werden muss[8], findet diese Rhetorik in den Medien und Kreisen der Politik nach wie vor breite Anwendung, um die unkritische Übernahme automatisierter Technologien ins öffentliche Leben zu rechtfertigen.

Werden sie als „Lösung" verkauft, steuern ADM-Systeme unverzüglich in das Territorium hinein, das in Arthur C. Clarkes Drittem Gesetz beschrieben wird: Magie. Magie zu regulieren, ist schwierig, wenn nicht gar unmöglich. Und diesbezüglich für Transparenz zu sorgen und Erläuterungen zu geben, noch weitaus schwieriger. Man sieht zu, wie die Hand in den Hut fasst und ein Kaninchen hervorzieht, doch der Vorgang selbst ist eine „Blackbox" und *soll es auch bleiben*.

Diesen Umstand prangerten zahlreiche am Projekt „Automating Society" beteiligte Wissenschaftler·innen als fundamentalsten Mangel in den Argumentationen an, die hinter vielen der von ihnen beschriebenen ADM-Systeme stehen. Wie im Kapitel über Deutschland dargelegt, impliziert dies auch, dass ein Grossteil der an solchen Systemen geäusserten Kritik als pauschale Ablehnung von „Innovationen" verleumdet und Befürworter·innen digitaler Rechte als „Neo-Ludditen" (Neue Maschinenstürmer·innen) dargestellt werden. Damit werden nicht nur die historischen Realitäten der Maschinenstürmerbewegung ignoriert, die sich auf Arbeitsmarktpolitik bezog und nicht auf Technologien per se. Nein, es gefährdet auch – und dies ist vielleicht noch wichtiger – die Effektivität möglicher Aufsichts- und Durchsetzungsmechanismen.

Zu einer Zeit, da die „KI"-Branche das Entstehen einer „lebhaften" Lobby-Industrie erlebt, insbesondere im Vereinigten Königreich, könnte dies dazu führen, dass Leitlinien für „ethics-washing" und anderen politischen Reaktionen verabschiedet werden, die ebenso ineffektiv wie strukturell inadäquat sind, um die Folgen von ADM-Systemen für die Menschenrechte zu bewältigen. Diese Sichtweise gipfelt letztlich in der Prämisse, dass nicht ADM-Systeme an demokratische Gesellschaften angepasst werden sollten, sondern vielmehr wir Menschen uns den ADM-Systemen anpassen sollten.

Um diesem Narrativ entgegenzutreten, sollten wir uns nicht scheuen, einige grundsätzliche Fragen zu stellen: nämlich, ob ADM-Systeme demokratiekompatibel sein und zum Wohl der Gesellschaft als Ganzes zum Einsatz kommen können, anstatt nur in Teilbereichen. So könnte es zum Beispiel sein, dass bestimmte menschliche Aktivitäten – etwa jene, die die Sozialfürsorge betreffen – nicht der Automatisierung unterworfen werden sollten, oder dass bestimmte

8 Man denke an das Debakel um den „Buona Scuola"-Algorithmus in Italien; vgl. das Kapitel Italy in der Gesamtausgabe des Automating Society Reports unter https://automatingsociety.algorithmwatch.org/report2020/italy/

Technologien – namentlich Gesichtserkennung in Echtzeit an öffentlichen Plätzen – nicht beim endlosen Streben nach „KI-Leadership“ gefördert, sondern stattdessen verboten werden sollten.

Noch wichtiger ist , dass wir jedes ideologische Framing, das uns hindert, solche Fragen zu stellen, ablehnen sollten. Im Gegenteil: Was es jetzt braucht, ist ein echter Politikwandel, um für eine strengere Überprüfung dieser Systeme zu sorgen. Im folgenden Abschnitt listen wir die Kernforderungen auf, die sich aus unseren Rechercheergebnissen ableiten. Wir hoffen, dass sie breit diskutiert und am Ende umgesetzt werden.

Nur durch eine informierte, inklusive und evidenzbasierte demokratischen Debatte wird es uns gelingen, die richtige Balance zu finden zwischen den Vorteilen, die ADM-Systeme im Hinblick auf Schnelligkeit, Effizienz, Fairness, bessere Prävention und besseren Zugang zu staatlichen Dienstleistungen bieten können – und dies auch tun – und den Gefahren, die sie für die Rechte von uns allen darstellen.

Handlungs-empfehlungen

Basierend auf den Forschungsergebnissen unseres Berichts „Automating Society 2020“, empfehlen wir den Entscheidungsträger·innen im EU-Parlament und den Parlamenten der Mitgliedsstaaten, der EU-Kommission, den Nationalregierungen, Wissenschaftler·innen, Organisationen der Zivilgesellschaft (Interessenvertretungen, Stiftungen, Gewerkschaften usw.) sowie dem privatwirtschaftlichen Sektor (Unternehmen und Branchenverbänden) die folgenden Massnahmen. Diese Empfehlungen sollen helfen, sicherzustellen, dass ADM-Systeme, die derzeit in ganz Europa im Einsatz sind oder eingeführt werden sollen, tatsächlich mit Menschenrechten und Demokratie vereinbar sind:

1. Transparenz von ADM-Systemen erhöhen

Ohne die Möglichkeit, genau zu wissen, wie, warum und zu welchem Zweck ADM-Systeme zum Einsatz kommen, sind alle anderen Anstrengungen, solche Systeme mit Grundrechten in Einklang zu bringen, zum Scheitern verurteilt.

/ Öffentliche Register für ADM-Systeme einführen, die von der öffentlichen Hand genutzt werden

Mitgliedsstaaten sollten durch EU-weite Regelungen verpflichtet werden, öffentliche Register für ADM-Systeme einzurichten, die im öffentlichen Sektor genutzt werden.

Damit verbunden sein muss die gesetzliche Verpflichtung, dass die für ein ADM-System verantwortlichen Stellen bzw. Personen den Zweck des Systems offenlegen und dokumentieren, ergänzt um eine Erläuterung des Modells (einschliesslich seiner Logik) sowie Informationen darüber, wer das System entwickelt hat. Diese Informationen müssen auf einfach zu verstehende und leicht zugängliche Weise verfügbar gemacht werden, einschliesslich strukturierter digitaler Daten, die auf einem standardisierten Protokoll basieren.

Staatliche Behörden haben eine besondere Verantwortung, die Betriebseigenschaften von ADM-Systemen, die in der öffentlichen Verwaltung zum Einsatz kommen, transparent zu machen. Unterstrichen wird dies von der Entscheidung zu einer kürzlich eingereichten Verwaltungsbeschwerde in Spanien. Dort heisst es, dass „jedes von der staatlichen Verwaltung genutzte ADM-System standardmässig öffentlich gemacht werden sollte“. Hat dieses Urteil Bestand, könnte es zu einem Präzedenzfall für ganz Europa werden.

Für ADM-Systeme, die im öffentlichen Sektor zum Einsatz kommen, sollten Offenlegungsmechanismen in allen Fällen verpflichtend sein. Für den Einsatz von ADM-Systemen durch private Unternehmen sollten sie gelten, sofern ein KI-/ADM-System signifikante Auswirkungen auf eine Einzelperson, eine bestimmte Gruppe oder die Gesellschaft als Ganzes hat.

/ Rechtlich verbindliche Rahmenbedingungen für den Zugang zu Daten einführen, um Forschung im öffentlichen Interesse zu unterstützen und zu ermöglichen

Verstärkte Transparenz erfordert neben der Offenlegung von Informationen über Zweck, Logik und Entwickler·in eines Systems auch die Möglichkeit, dessen Inputs und Outputs einer tiefgreifenden Analyse und Prüfung zu unterziehen. Darüber hinaus müssen die Trainingsdaten und Datenergebnisse unabhängigen Wissenschaftler·innen, Journalist·innen und Aktivist·innen der Zivilgesellschaft für Forschung und Recherche im öffentlichen Interesse zugänglich gemacht werden.

Aus diesem Grund empfehlen wir die Einführung robuster, rechtlich verbindlicher Rahmenbedingungen für den Zugang zu Daten, die explizit darauf ausgerichtet sind, Forschung im öffentlichen Interesse zu ermöglichen und zu unterstützen, und die die Regelungen zum Datenschutz sowie die Gesetze zum Schutz der Privatsphäre respektieren.

In Anlehnung an bestehende Best Practices auf nationaler und EU-Ebene sollten solche abgestuften Rahmenwerke sowohl Sanktionssysteme wie auch Kontrollmechanismen und regelmässige Überprüfungen einschliessen. Wie private Datensharing-Partnerschaften gezeigt haben, bestehen berechtigte Bedenken im Hinblick auf die Privatsphäre der Nutzer·innen und die mögliche Deanonymisierung bestimmter Arten von Daten.

Politische Entscheidungsträger·innen sollten sich Regelungen für die gemeinsame Nutzung von Gesundheitsdaten zum Vorbild nehmen, um den privilegierten Zugang zu bestimmten Arten granularer Daten zu ermöglichen und gleichzeitig sicherzustellen, dass persönliche Daten ausreichend geschützt sind (z. B. durch sichere Betriebsumgebungen).

Rechenschaftspflichten durchzusetzen ist nur möglich, wenn es einen Zugang zu Daten der Plattformen gibt. Das ist zugleich eine Voraussetzung für die effektive Umsetzung zahlreicher Ansätze zur Überprüfung von Systemen.

2. Einen eindeutigen Rahmen für die Rechenschaftspflicht von ADM-Systeme schaffen

Wie Forschungsergebnisse aus Spanien und Frankreich gezeigt haben, führen gesetzlich verankerte Transparenzerfordernisse und/oder die Offenlegung von Informationen nicht automatisch zu verantwortungsvollem Handeln. Weitere Schritte sind notwendig, um sicherzustellen, dass Gesetze und Vorschriften auch tatsächlich durchgesetzt werden können.

/ Ansätze für die effektive Auditierung algorithmischer System entwickeln und etablieren

Um für echte Transparenz zu sorgen, muss der erste Schritt (die Einrichtung öffentlicher Register), um Prozesse ergänzt werden, die eine effektive Auditierung algorithmischer Systeme garantieren.

Der Begriff „Auditing" ist weit verbreitet, doch gibt es keine allgemeinverbindliche Definition. Wir verstehen Auditing in diesem Kontext gemäss der ISO-Definition als einen „systematischen, unabhängigen und dokumentierten Prozess, um objektive Nachweise zu erlangen und diese objektiv zu bewerten um festzustellen, inwiefern die Prüfkriterien erfüllt sind".[9]

Wir haben bisher noch keine zufriedenstellenden Antworten auf die komplexen Fragen[10], die von der Auditierung algorithmischer Systeme aufgeworfen werden. Allerdings deuten unsere Ergebnisse klar darauf hin, dass diese Antworten gefunden werden müssen, und zwar im Rahmen einer breiten Debatte, an der alle Interessengruppen beteiligt sind, sowie durch gründliche, engagierte Forschung.

Es sollten sowohl Auditkriterien, als auch angemessene Auditprozesse entwickelt werden. Dabei ist ein Ansatz zu verfolgen, der nicht nur möglichst viele Beteiligte und Betroffene einbezieht, sondern auch die unverhältnismässigen Auswirkungen von ADM-Systemen auf vulnerable Gruppen berücksichtigt und deren Mitsprache sichert.

Aus diesem Grund fordern wir die politischen Entscheidungsträger·innen auf, solche Stakeholder-Prozesse zu initiieren, um die angesprochenen Fragen zu klären, sowie Finanzierungsquellen zur Verfügung zu stellen, die die Teil-

9 https://www.iso.org/obp/ui/#iso:std:iso:19011:ed-3:v1:en

10 Beim Nachdenken über mögliche Modelle für algorithmisches Auditing tauchen diverse Fragen auf. 1) Wer/Was (Dienstleistungen/Plattformen/Produkte) soll auditiert werden? Wie können Auditing-Systeme auf die betreffende Art von Plattform/Dienstleistung zugeschnitten werden? 2) Wann sollte das Auditing von einer öffentlichen Stelle vorgenommen werden (auf EU-Ebene, nationaler Ebene, lokaler Ebene), und wann kann es von privaten Stellen/Expert·innen übernommen werden (Unternehmen, Zivilgesellschaft, Wissenschaftler·innen)? 3) Wie stellt man die Unterscheidung zwischen der Beurteilung von Auswirkungen ex-ante (d. h. in der Designphase/Entwicklungsphase) und ex-post (d. h. im Betrieb) sowie im Hinblick auf die betreffenden Herausforderungen klar? 4) Wie können Kompromisse bei den verschiedenen Vor- und Nachteilen der Überprüfbarkeit bewertet werden? (So könnten etwa Simplizität, Allgemeingültigkeit, Anwendbarkeit, Präzision, Flexibilität, Interpretierbarkeit, Datenschutz und Leistungsfähigkeit eines Auditing-Prozesses in einem Spannungsverhältnis stehen). 5) Welche Informationen müssen verfügbar sein, damit ein Audit effektiv und verlässlich ist (z. B. Quellcode, Trainingsdaten, Dokumentation)? Brauchen Prüfende physischen Zugang zu den Systemen, während sie in Betrieb sind, um einen effektiven Audit durchzuführen? 6) Welche Nachweispflichten für Händler·innen/Service-Anbieter·innen sind nötig und angemessen? 7) Wie können wir sicherstellen, dass das Auditing überhaupt möglich ist? Sollten Auditanforderungen bereits bei der Entwicklung algorithmischer Systeme Berücksichtigung finden („prüffähiges Design")? 8) Vorschriften für die Publikation der Ergebnisse: Fällt ein Audit negativ aus, und die Probleme sind nicht behoben: Wie sollten sich Prüfende verhalten; auf welche Weise kann publik gemacht werden, dass ein Versagen vorliegt? 9) Wer prüft die Prüfenden? Wie stellt man die Rechenschaftspflicht der Prüfenden sicher?

nahme bisher nicht angemessen repräsentierter Interessengruppen sichern.

Weiterhin fordern wir die Bereitstellung ausreichender Mittel zur Unterstützung/Finanzierung von Forschungsprojekten, die sich mit der Entwicklung von Modellen zur effektiven Auditierung algorithmischer Systeme beschäftigen.

/ Zivilgesellschaftliche Organisationen als Watchdogs von ADM-Systemen unterstützen

Unsere Erkenntnisse zeigen deutlich, dass für eine effektive kritische Hinterfragung undurchsichtiger ADM-Systeme die Arbeit zivilgesellschaftlicher Organisationen unabdingbar ist. Mittels Forschung und Interessenvertretung, häufig auch in Kooperation mit Journalist·innen und dem akademischen Bereich, haben sie sich in den letzten Jahren wiederholt in politische Debatten um diese Systeme eingemischt. Dabei ist es ihnen in einigen Fällen gelungen, effektiv sicherzustellen, dass öffentliche Interessen und Grundrechte sowohl vor als auch nach der Implementierung solcher Systeme in vielen europäischen Ländern angemessen berücksichtigt werden.

Zivilgesellschaftliche Akteur·innen sollten daher als Watchdogs der zunehmend automatisierten Gesellschaft unterstützt werden. Als solche gehören sie zu den integralen Bestandteilen jedes effektiven Rahmenwerks für die Rechenschaftspflicht von ADM-Systemen.

/ Gesichtserkennung verbieten, die den Weg für Massenüberwachung ebnet

Nicht alle ADM-Systeme sind gleichermassen gefährlich, und ein risikobasierter Regulierungsansatz, wie etwa in Deutschland und der EU, spiegelt diese Tatsache korrekt wider. Um jedoch eine praktikable Rechenschaftspflicht für Systeme zu gewährleisten, die als risikobehaftet eingestuft werden, müssen effektive Aufsichts- und Durchsetzungsmechanismen geschaffen werden. Dies ist umso wichtiger für solche Systeme, denen ein „hohes Risiko" zur Verletzung von Nutzer·innenrechten bescheinigt wird.

Ein vordringliches Beispiel, das sich aus unseren Erkenntnissen ergeben hat, ist die Gesichtserkennung. Wie sich gezeigt hat, stellen ADM-Systeme, die auf biometrischen Technologien beruhen, darunter Gesichtserkennung, eine ernsthafte Bedrohung für das Gemeinwohl und die Grundrechte dar, denn sie ebnen den Weg zu undifferenzierter Massenüberwachung – insbesondere angesichts der Tatsache, dass sie trotzdem weiter verbreitet werden und auf intransparente Art und Weise zum Einsatz kommen.

Wir fordern, dass alle öffentlichen Nutzungen von Gesichtserkennung, die sich zu Massenüberwachung auswachsen könnten, auf EU-Ebene bis auf Weiteres verboten werden, und zwar so schnell wie möglich.

Solche Technologien könnten innerhalb der EU sogar jetzt schon als illegal gelten, zumindest für bestimmte Anwendungen, wenn ihr Einsatz ohne „ausdrückliche Zustimmung" der gescannten Personen erfolgt. Diese juristische Auslegung wurde von den staatlichen Behörden in Belgien vorgeschlagen, die den Einsatz von Gesichtserkennung in ihrem Land in einem bahnbrechenden Urteil mit einer Strafzahlung belegten.

3. Algorithmische Kompetenzen verbessern und die öffentliche Debatte um ADM-Systeme stärken

Mehr Transparenz von ADM-Systemen ist nur dann von echtem Nutzen, wenn diejenigen, die mit ihnen konfrontiert werden, wie etwa Aufsichtsbehörden, Regierungen und Standardisierungsgremien, in der Lage sind, auf verantwortungsvolle und umsichtige Art und Weise mit diesen Systemen und ihren Auswirkungen umzugehen. Darüber hinaus müssen jene, die von diesen Systemen betroffen sind, verstehen können, wo, warum und wie diese Systeme zum Einsatz kommen. Aus diesem Grund müssen wir die Stärkung algorithmischer Kompetenz auf allen Ebenen vorantreiben, bei wichtigen Akteur·innen ebenso wie in der breiten Öffentlichkeit, und vielfältigere öffentliche Debatten über ADM-Systeme und deren Auswirkungen auf die Gesellschaft fördern.

/ Unabhängige Kompetenzzentren für ADM einrichten

Parallel zu unserer Forderung, das Auditing von Algorithmen und entsprechende Forschung zu unterstützen, plädieren wir dafür, unabhängige Kompetenzzentren für ADM auf nationaler Ebene einzurichten. Diese Zentren sollen dazu dienen, Algorithmen zu kontrollieren und zu bewerten, Forschung zu betreiben und Berichte zu erstellen. In Koordination mit Regulierungsbehörden, der Zivilgesellschaft

und der Wissenschaft sollten sie Regierung wie Industrie zu den Auswirkungen des Einsatzes von ADM-Systemen auf die Gesellschaft und Menschenrechte beraten. Die übergreifende Rolle dieser Zentren besteht darin, ein sinnvolles System der Rechenschaftspflicht zu entwickeln sowie die entsprechenden Kapazitäten zu schaffen.

Um Vertrauen, Transparenz und Kooperation zwischen allen Beteiligten aufzubauen, sollten die nationalen Kompetenzzentren zivilgesellschaftliche Organisationen, Interessengruppen sowie bereits bestehende Durchsetzungsstellen wie etwa Datenschutzbehörden und nationale Menschenrechtsgremien einbeziehen.

Als unabhängige Körperschaften des öffentlichen Rechts käme den Kompetenzzentren eine zentrale Rolle zu: Sie würden die Entwicklung politischer Konzepte und nationaler Strategien im Umgang mit ADM koordinieren sowie Kompetenzen und Fähigkeiten bei bereits existierenden Regulierungsbehörden sowie Regierungen und Standardisierungsgremien aufbauen, um auf die zunehmende Nutzung von ADM-Systemen zu reagieren.

Diese Zentren sollten keine Regulierungsbefugnisse haben, sondern unverzichtbare Expertise einbringen, die die Frage beantworten helfen, wie Grundrechte geschützt und kollektiver und gesellschaftlicher Schaden abgewendet werden können. So sollten sie etwa kleine und mittlere Unternehmen (KMU) bei der Erfüllung ihrer Sorgfaltspflichten im Hinblick auf Menschenrechtsfragen unterstützen, einschliesslich der Durchführung von Folgenabschätzungen für Menschenrechte (Human Rights Impact Assessments) oder Folgenabschätzungen zum Einsatz von Algorithmen (Algorithmic Impact Assessments), sowie bei der Registrierung von ADM-Systemen in dem oben dargestellten öffentlichen Register.

/ Eine inklusive und vielfältige demokratische Debatte rund um ADM-Systeme fördern

Es müssen nicht nur Kenntnisse und Kompetenzen bei denjenigen gestärkt werden, die ADM-Systeme zum Einsatz bringen. Ebenso unerlässlich ist es, mit Hilfe einer breiteren Debatte und vielfältiger Programme die algorithmischen Kompetenzen der Allgemeinheit zu fördern.

Unsere Erkenntnisse legen nahe, dass ADM-Systeme der breiteren Öffentlichkeit nicht nur intransparent bleiben, wenn sie im Einsatz sind, sondern dass selbst die Entscheidung, ob ein ADM-System überhaupt zum Einsatz kommen soll oder nicht, für gewöhnlich ohne Wissen oder Beteiligung der Öffentlichkeit getroffen wird.

Es besteht daher die dringende Notwendigkeit, von Anfang an die Öffentlichkeit (das öffentliche Interesse) einzubeziehen, wenn darüber entschieden wird, ADM-Systeme einzusetzen.

Oder, um es etwas allgemeiner auszudrücken: Wir brauchen eine vielfältigere öffentliche Debatte zu den Auswirkungen von ADM. Wir müssen aufhören, ausschliesslich Expert·innengruppen anzusprechen; wir müssen dafür sorgen, dass die breitere Öffentlichkeit einen besseren Zugang zu diesen Themen bekommt. Dies bedeutet, eine andere Sprache zu sprechen als einen technisch-juristischen Slang, um öffentliche Aufmerksamkeit zu erzeugen und Interesse zu wecken.

Zu diesem Zweck sollten in Ergänzung zu den bereits genannten Massnahmen detaillierte Programme zum Aufbau und zur Förderung digitaler Selbstbestimmung etabliert werden. Wenn unser Ziel darin besteht, eine verstärkte, informierte öffentliche Debatte zu führen und den Bürger·innen der EU Selbstbestimmung zu ermöglichen, dann müssen wir damit beginnen, Kenntnisse zur Digitalisierung zu vermitteln und voranzutreiben und dabei besonderes Augenmerk auf die sozialen, ethischen und politischen Konsequenzen des Einsatzes von ADM-Systemen legen.

Weichenstellung für die Zukunft von ADM in Europa

Systeme zum automatisierten Entscheiden nehmen bei der Verteilung von Rechten und Dienstleistungen in Europa eine zentrale Rolle ein, und immer mehr Institutionen auf dem Kontinent erkennen deren Bedeutung für das öffentliche Leben an – sowohl die Chancen als auch die Herausforderungen.

Von Kristina Penner und Fabio Chiusi

Seit unserem ersten Bericht im Januar 2019 – und ungeachtet der Tatsache, dass die EU nach wie vor in einer breit geführten Debatte um „vertrauenswürdige“ Künstliche Intelligenz befangen ist – haben diverse Gremien, vom EU-Parlament bis zum Europarat, Dokumente veröffentlicht, die helfen sollen, für die kommenden Jahre, wenn nicht Jahrzehnte, einen gemeinsamen Kurs im Umgang mit ADM zu finden.

Im Sommer 2019 versprach die neugewählte Präsidentin der EU-Kommission, Ursula von der Leyen, eine selbsterklärte Technikoptimistin, innerhalb von 100 Tagen nach ihrem Amtsantritt „Rechtsvorschriften mit einem koordinierten europäischen Konzept für die menschlichen und ethischen Aspekte der künstlichen Intelligenz“ und die Regulierung von Künstliche Intelligenz (KI) vorzuschlagen. Stattdessen veröffentlichte die Europäische Kommission im Februar 2020 ein „Weissbuch” zur Künstlichen Intelligenz. Das Strategiepaket enthält „Ideen und Massnahmen”, um die Bürger·innen der EU zu informieren und den Weg für künftige Gesetzgebungsverfahren zu ebnen. Darüber hinaus macht es sich für die „technologische Unabhängigkeit“ Europas stark: In von der Leyens eigenen Worten bedeutet dies „die Fähigkeit, die Europa haben muss, um seine eigenen Entscheidungen zu treffen, die auf seinen eigenen Werten basieren und seine eigenen Regeln respektieren.“ Dies werde dabei helfen, uns alle zu Tech-Optimist·innen zu machen.

Ein zweites grundlegendes Vorhaben, das Auswirkungen auf ADM in Europa hat, ist der in von der Leyens „Agenda für Europa“ angekündigte Digital Services Act (DSA). Das Gesetz soll die E-Commerce-Direktive ersetzen, die seit 2000 in Kraft ist. Es zielt darauf ab, „unsere Haftungs- und Sicherheitsregelungen für digitale Plattformen, Dienstleistungen und Produkte zu modernisieren und unseren Digitalen Binnenmarkt zu vollenden“ – was fundamentale Debatten um die Rolle von ADM bei politischen Entscheidungen zum Thema Content Moderation, Zwischenhändlerhaftung und Meinungsfreiheit im Allgemeinen auslöst[11].

Eine explizite Fokussierung auf ADM-Systeme findet sich in einer vom Ausschuss für Binnenmarkt und Verbraucherschutz des EU-Parlaments verabschiedeten Resolution sowie in einer Empfehlung „zu den menschenrechtlichen Auswirkungen algorithmischer Systeme“ des Ministerkomitees des Europarates.

11 Detaillierte Anmerkungen und Empfehlungen rund um ADM-Systeme im Kontext des DSA finden sich in den Veröffentlichungen zum Projekt ‘Governing Platforms’ von AlgorithmWatch.

VIELE BEOBACHTER·INNEN SEHEN IN DER ART UND WEISE, WIE EU-INSTITUTIONEN UND INSBESONDERE DIE KOMMISSION IHRE ÜBERLEGUNGEN UND VORSCHLÄGE ZU KI UND ADM FORMULIEREN, EINE GRUNDLEGENDE SPANNUNG ZWISCHEN WIRTSCHAFTLICHEN IMPERATIVEN UND DEN ERFORDERNISSEN ZUR WAHRUNG VON GRUNDRECHTEN.

Es hat sich gezeigt, dass insbesondere der Europarat im Verlauf des Jahres 2019 eine zunehmend wichtige Rolle in der politischen Debatte um KI eingenommen hat. Zwar bleibt abzuwarten, wie stark sein tatsächlicher Einfluss auf die Regulierungsbemühungen sein wird, doch spricht einiges dafür, dass er als „Wächter" der Menschenrechte fungieren kann. Besonders deutlich wird dies in der Empfehlung „Unboxing Artificial Intelligence: 10 steps to protect Human Rights" der Europarat-Kommissarin für Menschenrechte Dunja Mijatović sowie in der Tätigkeit des im September 2019 gegründeten Ad-hoc-Ausschusses für künstliche Intelligenz (CAHAI).

Viele Beobachter·innen sehen in der Art und Weise, wie EU-Institutionen und insbesondere die Kommission ihre Überlegungen und Vorschläge zu KI und ADM formulieren, eine grundlegende Spannung zwischen wirtschaftlichen Imperativen und den Erfordernissen zur Wahrung von Grundrechten. Auf der einen Seite will Europa „die Nutzung von und die Nachfrage nach Daten und datenbasierten Produkten und Dienstleistungen auf dem gesamten Binnenmarkt erhöhen", um zu einem „führenden Akteur" bei industriellen Anwendungen von KI zu werden und angesichts des wachsenden Drucks von Rivalen wie den USA und China die Wettbewerbsfähigkeit europäischer Unternehmen zu stärken. Das gilt in besonderer Weise für ADM, denn hier liegt die Annahme zugrunde, dass die EU mit Hilfe dieser „datenagilen" Wirtschaft zum „Vorbild für eine durch Daten ermächtigte Gesellschaft werden [kann], die bessere Entscheidungen trifft – in der Geschäftswelt wie im staatlichen Bereich". Im Weissbuch zur Künstlichen Intelligenz heisst es sinngemäss: Daten sind das Lebenselixier wirtschaftlicher Entwicklung.

Auf der anderen Seite kann die automatische Verarbeitung von Daten in den Bereichen Gesundheit, Arbeit und Sozialwesen jedoch Entscheidungen herbeiführen, die diskriminierende und unfaire Ergebnisse zur Folge haben. Diese „dunkle Seite" des Einsatzes von Algorithmen bei Prozessen der Entscheidungsfindung geht die EU mit einer Reihe von Prinzipien an. Im Fall von risikoreichen Systemen sollen Regeln garantieren, dass automatisierte Entscheidungsprozesse mit den Menschenrechten und sinnvollen demokratischen Kontrollen vereinbar sind. Dieser Ansatz wird von EU-Institutionen unter dem Begriff „menschenzentriert" zusammengefasst; er sei einzigartig und stünde in fundamentalem Gegensatz zu den in den USA (profitorientiert) und China (bestimmt von nationalen Sicherheitsinteressen und Massenüberwachung) verfolgten Ansätzen.

Es sind jedoch Zweifel laut geworden, ob Europa beide Ziele gleichzeitig erreichen kann. Ein wichtiges Beispiel ist die Gesichtserkennung: Obwohl uns – wie dieser Bericht zeigt – inzwischen viele Belege für den Einsatz ungeprüfter und intransparenter Gesichtserkennungssysteme in den meisten EU-Mitgliedsstaaten vorliegen, hat es die EU-Kommission bisher versäumt, schnell und entschlossen zu handeln, um die Rechte der EU-Bürger·innen zu schützen. Wie geleakte Entwürfe zum Weissbuch KI der Europäischen Kommission enthüllten[12], war die EU im Begriff, „biometrische Fernidentifikation" an öffentlichen Orten zu verbieten, scheute aber in letzter Sekunde davor zurück und warb stattdessen für eine „breite Debatte" zu diesem Thema.

Unterdessen wird die Nutzung kontrovers diskutierter Anwendungen von ADM bei Grenzkontrollen, sogar einschliesslich Gesichtserkennung, im Rahmen EU-finanzierter Projekte weiterhin vorangetrieben.

Politische Strategien und Debatten

/ Das Europäische Datenstrategiepaket und das Weissbuch zur Künstlichen Intelligenz

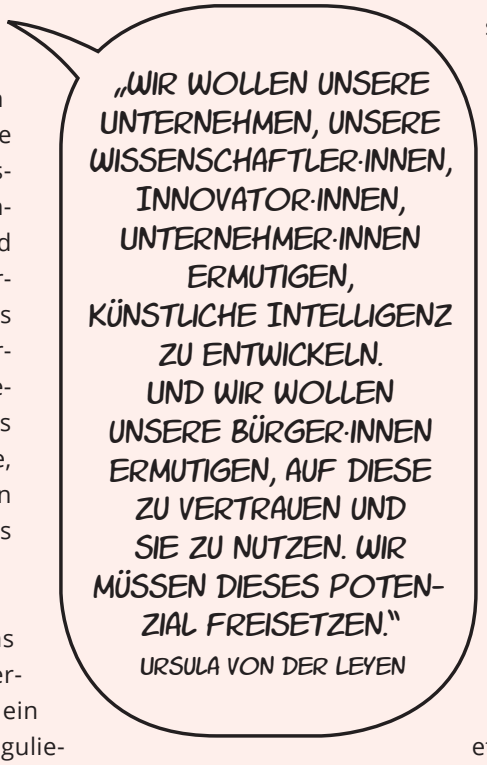
Die versprochene umfassende Gesetzgebung mit einem „koordinierten europäischen Konzept für die menschlichen und ethischen Aspekte der künstlichen Intelligenz", die von der Leyen in ihrer „Agenda für Europa" angekündigt hatte, wurde nicht in den „ersten 100 Tagen" ihrer Amtszeit" vorgelegt. Die EU-Kommission hat jedoch eine Reihe von Dokumenten veröffentlicht, denen sich ein Paket von Prinzipien und Konzepten entnehmen lässt, die als Grundlage dienen können.

Am 19. Februar 2020 wurden zeitgleich die „Europäische Datenstrategie" und das „Weissbuch Zur künstlichen Intelligenz" publiziert, in denen die wichtigsten Prinzipien des strategischen Konzeptes der EU für KI dargelegt werden (inklusive ADM-Systeme, obwohl diese nicht explizit genannt werden). Zu diesen Prinzipien zählen: „Der Mensch zuerst" (Technologie, die für die Menschen arbeitet), technologische Neutralität (keine Technologie ist per se gut oder

12 https://www.politico.eu/article/eu-considers-temporary-ban-on-facial-recognition-in-public-spaces/

schlecht; dies wird ausschliesslich durch deren Nutzung bestimmt) und natürlich Unabhängigkeit und Optimismus. Wie von der Leyen es ausdrückt: „Wir wollen unsere Unternehmen, unsere Wissenschaftler·innen, Innovator·innen, Unternehmer·innen ermutigen, künstliche Intelligenz zu entwickeln. Und wir wollen unsere Bürger·innen ermutigen, auf diese zu vertrauen und sie zu nutzen. Wir müssen dieses Potenzial freisetzen."

All dem liegt die Auffassung zugrunde, dass neue Technologien keine neuen Werte mit sich bringen sollten. Die „neue digitale Welt", die von der Leyens Administration anstrebt, soll Bürger·innen- und Menschenrechte umfassend schützen. „Exzellenz" und „Vertrauen" – wie bereits im Titel des Weissbuchs hervorgehoben, werden als die tragenden Säulen angesehen, auf denen ein europäisches KI-Modell ruhen kann und sollte, wodurch es sich von den Strategien sowohl der USA wie auch Chinas unterscheidet.



In den Details des Weissbuchs sucht man diesen Anspruch allerdings vergeblich. So wird darin ein risikobasierter Ansatz für die Regulierung von KI dargelegt, bei dem der Grad an Regulierung den Auswirkungen von „KI"-Systemen auf das Leben der Bürger·innen entspricht. „Bei hohem Risiko wie etwa im Gesundheitswesen, bei der Polizei oder im Verkehrssektor ", heisst es in der Presseerklärung, „sollten KI-Systeme transparent und nachvollziehbar sein sowie menschlicher Aufsicht unterliegen". Zu den geforderten Schutzmechanismen gehören auch die Prüfung und Zertifizierung angewendeter Algorithmen; diese sollten ebenso weit Verbreitung finden wie bei „Kosmetikartikeln, Autos oder Spielzeug". Für „KI-Anwendungen ohne hohes Risiko" soll dagegen lediglich ein freiwilliges Kennzeichnungssystem gelten: „Die KI-Anwendungen der betreffenden Wirtschaftsakteure würden dann ein Gütesiegel erhalten."

Kritiker·innen haben jedoch angemerkt, dass bereits die in dem Papier vorgenommene Definition von „Risiko" nicht nur ein Zirkelschluss, sondern auch zu vage ist, was dazu führen könnte, dass diverse wirkungsvolle ADM-Systeme durch das Raster des vorgeschlagenen Rahmenwerks fallen[13].

Die im Verlauf der öffentlichen Konsultation zwischen Februar und Juni 2020 eingegangenen Stellungnahmen[14] machen deutlich, wie kontrovers dieses Konzept diskutiert wird. 42,5 % der Wortmeldungen stimmten darin überein, dass „verpflichtende Anforderungen" auf "KI-Anwendungen mit hohem Risiko" beschränkt werden sollten, während 30,6 % einer solchen Beschränkung kritisch gegenüberstanden.

Davon abgesehen gibt es keinerlei Beschreibung klarer Durchsetzungsmechanismen für solche Anforderungen. Und ebenso wenig wird ein Prozess beschrieben, wie man zu solchen gelangen könnte.

Für biometrische Technologien, insbesondere Gesichtserkennung, sind die daraus resultierenden Konsequenzen sofort ersichtlich. Hierzu schlägt das Weissbuch eine Unterscheidung vor zwischen „biometrischer Authentifizierung", die als nicht kontrovers betrachtet wird (z. B. Gesichtserkennung zur Entsperrung eines Smartphones), und „biometrischer Fernidentifikation" (wie etwa der Einsatz an öffentlichen Orten, um die Identität von Protestierenden festzustellen), die ernste Bedenken im Hinblick auf die Einhaltung von Menschenrechten und die Verletzung der Privatsphäre aufwerfen könnten.

13 „Um zwei Beispiele zu nennen: VioGén, ein ADM-System zur Vorhersage geschlechtsspezifischer Gewalt, und Ghostwriter, eine Anwendung zur Aufdeckung von Betrugsversuchen bei Prüfungen, würden mit hoher Wahrscheinlichkeit durchs Regulierungsraster fallen, obwohl sie mit extremen Risiken verbunden sind" (https://algorithmwatch.org/en/response-european-commission-ai-consultation/)

14 "Insgesamt gingen 1.215 Stellungnahmen ein, davon 352 von Unternehmen und Unternehmer-/Branchenverbänden, 406 von Bürger·innen (92 % EU-Bürger·innen), 152 von akademischen Einrichtungen/Forschungsinstituten und 73 von staatlichen Behörden. Die Zivilgesellschaft war durch 160 Wortmeldungen vertreten (darunter 9 von Verbraucherschutzorganisationen, 129 von Nichtregierungsorganisationen und 22 von Gewerkschaften). 72 Stellungnahmen fielen unter „Sonstige". Die Stellungnahmen kamen „aus aller Welt", darunter „Indien, China, Japan, Syrien, Irak, Brasilien, Mexiko, Kanada, die USA und das Vereinigte Königreich". (aus dem zusammenfassenden Bericht über die Konsultationen, siehe folgender Link: https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence)

Gemäss dem von der EU vorgeschlagenen System wären nur die Fälle der zweitgenannten Kategorie problematisch. In den FAQ zum Weissbuch heisst es: „Dies ist die Form der Gesichtserkennung, die den stärksten Eingriff in die Privatsphäre darstellt; sie ist in der EU grundsätzlich verboten", es sei denn, es besteht ein „wesentliches öffentliches Interesse" an ihrem Einsatz.

In dem erläuternden Dokument wird behauptet, die Genehmigung von Gesichtserkennung [ist] derzeit die Ausnahme. Die Erkenntnisse in diesem Bericht belegen jedoch unzweifelhaft das Gegenteil: Gesichtserkennung wird mit rasender Geschwindigkeit zur Norm. Eine geleakte Entwurfsfassung des Weissbuchs erkannte die Dringlichkeit des Problems offenbar an. Dort war der Vorschlag enthalten, ein drei bis fünf Jahre dauerndes Moratorium für den Einsatz von Gesichtserkennung an öffentlichen Orten zu verfügen, bis – und falls – ein Weg gefunden werden könne, sie mit demokratischen Kontrollmechanismen in Einklang zu bringen.

Unmittelbar vor der offiziellen Veröffentlichung des Weissbuchs forderte sogar EU-Kommissarin Margrethe Vestager, eine „Aussetzung" dieser Nutzungen.

Kurz nach Vestagers Forderung fügten offizielle Vertreter·innen der Kommission jedoch hinzu, dass diese „Aussetzung" nationale Regierungen nicht daran hindern würde, Gesichtserkennung gemäss den existierenden Regelungen einzusetzen. Letztendlich wurde in der finalen Fassung des Papiers jede Erwähnung eines Moratoriums gestrichen; stattdessen enthielt es die Ankündigung „einer breit angelegte[n] europäische[n] Debatte über die besonderen Umstände, die eine solche Nutzung rechtfertigen könnten". Dazu gehört, so steht es im Weissbuch, dass der Einsatz hinreichend begründet und verhältnismässig ist sowie demokratische Schutzmechanismen und die Achtung der Menschenrechte garantiert sind.

Das gesamte Dokument hindurch werden mit KI-basierten Technologien verbundene Risiken ganz allgemein als „potenziell" bezeichnet, die Vorteile dagegen als sehr real und unmittelbar dargestellt. Dies veranlasste viele[15] Menschenrechtsaktivist·innen zu der Aussage, das dem Weissbuch zugrundeliegende Narrativ spiegele eine besorgniserregende Umkehrung der Prioritätensetzung der EU wider; globale Wettbewerbsfähigkeit würde über den Schutz der Grundrechte gestellt.

Dennoch werden in den Dokumenten einige grundsätzliche Probleme angesprochen. So zum Beispiel die Interoperabilität solcher Lösungen und die Schaffung eines Netzwerks von Forschungszentren, die sich mit Anwendungsmöglichkeiten für KI beschäftigen und die auf „Exzellenz" und den Aufbau von Kompetenzen ausgerichtet sind.

Das Ziel ist, „in der EU in den nächsten zehn Jahren insgesamt mehr als 20 Mrd. EUR an KI-Investitionen pro Jahr zu mobilisieren".

Das Weissbuch scheint zudem von einem gewissen technologischen Determinismus geprägt zu sein. „Es ist äusserst wichtig" heisst es darin, „dass öffentliche Verwaltungen, Krankenhäuser, Versorgungsbetriebe und Verkehrsdienste, Finanzaufsichtsbehörden und andere Bereiche von öffentlichem Interesse rasch mit der Einführung KI-gestützter Produkte und Dienstleistungen beginnen. Ein besonderer Schwerpunkt wird auf den Bereichen Gesundheitsfürsorge und Verkehr liegen, in denen die Technologien so weit ausgereift sind, dass sie in grossem Massstab eingesetzt werden können.

Ob die Empfehlung zu einem raschen Einsatz von KI-Lösungen in allen Sphären menschlicher Aktivität mit den Anstrengungen der EU-Kommission vereinbar ist, die strukturellen Herausforderungen anzugehen, die ADM-Systemen im Hinblick auf Grundrechte und Fairness mit sich bringen, bleibt allerdings abzuwarten.

/ Entschliessung des EU-Parlaments zu ADM und Verbraucherschutz

Eine vom EU-Parlament im Februar 2020 verabschiedete Entschluss ging das Thema ADM-Systeme im Kontext des Verbraucherschutzes gezielter an. Die Entschliessung hob richtigerweise hervor, dass in den Bereichen „komplexe Algorithmensysteme und automatisierte Entscheidungsfindungsprozesse rasch technologische Fortschritte erzielt werden" und dass „diese Technologien zahlreiche Anwendungen, Chancen und Herausforderungen bieten und praktisch alle Bereiche des Binnenmarkts betreffen". Darüber hinaus betont der Text, dass „der derzeitige Rechtsrahmen der EU […] überprüft werden muss, um zu sehen, ob damit auf das Entstehen von KI und automatisierter Entscheidungsfindung reagiert […] werden kann".

15 Unter ihnen: Access Now (https://www.accessnow.org/cms/assets/uploads/2020/05/EU-white-paper-consultation_AccessNow_May2020.pdf), AI Now (https://ainowinstitute.org/ai-now-comments-to-eu-whitepaper-on-ai.pdf), EDRI (https://edri.org/can-the-eu-make-ai-trustworthy-no-but-they-can-make-it-just/) und AlgorithmWatch (https://algorithmwatch.org/en/response-european-commission-ai-consultation/).

DAS GESAMTE DOKUMENT HINDURCH WERDEN MIT KI-BASIERTEN TECHNOLOGIEN VERBUNDENE RISIKEN GANZ ALLGEMEIN ALS „POTENZIELL" BEZEICHNET, DIE VORTEILE DAGEGEN ALS SEHR REAL UND UNMITTELBAR DARGESTELLT

Die Entschliessung fordert einen „gemeinsamen Ansatz der EU für die Entwicklung automatisierter Entscheidungsfindungsprozesse" und trifft detaillierte Aussagen zu diversen Bedingungen, die solche Systeme erfüllen sollten, um im Einklang mit europäischen Werten zu stehen. Verbraucher·innen sollten „angemessen darüber informiert werden", auf welche Weise ihr Leben von Algorithmen beeinflusst wird, und sie sollten Zugang zu Personen mit Entscheidungsbefugnis haben, um Entscheidungen überprüfen und, falls nötig, korrigieren lassen zu können. Ausserdem sollten sie darüber informiert werden, „wenn die Preise von Waren oder Dienstleistungen auf der Grundlage automatisierter Entscheidungsfindung und der Erstellung von Profilen des Verbraucherverhaltens personalisiert wurden".

Die Entschliessung erinnert die EU-Kommission daran, dass ein sorgfältig ausgearbeiteter risikobasierter Regulierungsansatz erforderlich ist und betont, Produktsicherheitsvorschriften müssten in Betracht ziehen, dass ADM-Systeme sich „weiterentwickeln und in einer Art und Weise handeln können, die beim ersten Inverkehrbringen nicht vorgesehen ist" und „dass es problematisch ist, die Haftung in Fällen zu bestimmen, in denen die Schädigung der Verbraucher auf autonome Entscheidungsfindungsprozesse zurückzuführen ist".

Mit der Feststellung, dass „der Mensch letztlich für Entscheidungen verantwortlich und in der Lage sein muss, sich über Entscheidungen hinwegzusetzen, die im Zusammenhang mit freiberuflichen Dienstleistungen wie den medizinischen, juristischen und Buchhaltungsberufen sowie für den Bankensektor getroffen werden", wenn „berechtigte öffentliche Interessen auf dem Spiel stehen", nimmt die Entschliessung Bezug auf Art. 22 DSGVO. Insbesondere sollten „vor der Automatisierung professioneller Dienstleistungen" die Risiken „ordnungsgemäss" bewertet werden.

Am Ende listet die Entschliessung detaillierte Anforderungen an die Qualität und Transparenz beim Datenmanagement. Unter anderem wird betont, wie wichtig es sei, „nur hochwertige und tendenzfreie Datensätze zu verwenden, um die Leistung algorithmischer Systeme zu verbessern und das Vertrauen und die Akzeptanz der Verbraucher zu stärken"; „verständliche und tendenzfreie Algorithmen" zu verwenden; „Überprüfungsstrukturen" zu schaffen, die Verbraucher in die Lage zu versetzen, „eine Überprüfung endgültiger und dauerhafter automatisierter Entscheidungen durch Menschen sowie gegebenenfalls eine Entschädigung zu verlangen".

/ Das „Initiativrecht" des EU-Parlaments optimal nutzen

In ihrer Antrittsrede brachte Ursula von der Leyen sehr deutlich ihre Unterstützung eines „Initiativrechts" für das Europäische Parlament zum Ausdruck. „Wenn das Parlament mehrheitlich Entschliessungen annimmt, in denen die Kommission zur Vorlage von Legislativvorschlägen aufgefordert wird, sage ich zu, darauf – unter uneingeschränkter Wahrung der Grundsätze der Subsidiarität, Verhältnismässigkeit und besseren Rechtsetzung – in Form eines Rechtsakts zu reagieren."

Sollte es sich bei „KI" tatsächlich um eine Revolution handeln, die ein entsprechend auf sie abgestimmtes Paket

gesetzlicher Vorschriften erfordert, das mutmasslich im Laufe des 1. Quartals 2021 kommen wird, dann wollen die gewählten Repräsentant·innen mitreden. Dies, gepaart mit von der Leyens ausdrücklich erklärter Absicht, deren gesetzgeberische Kompetenzen zu stärken, könnte sogar in etwas gipfeln, das Politico als „parlamentarischen Moment" bezeichnet hat: in dessen Folge die Parlamentsausschüsse damit beginnen, viele verschiedene Berichte zu schreiben.

Jeder dieser Berichte untersucht spezifische Aspekte der Automatisierung in der öffentlichen Politik, die, ungeachtet der Tatsache, dass sie die kommende „KI"-Gesetzgebung prägen sollen, relevant für ADM sind.

So fordert etwa der Rechtsausschuss in seinem „Entwurf eines Berichts an die Kommission mit Empfehlungen zu einem Rahmen für die ethischen Aspekte von künstlicher Intelligenz, Robotik und damit zusammenhängenden Technologien" die Einrichtung einer „Europäischen Agentur für künstliche Intelligenz" und zugleich ein Netzwerk nationaler Aufsichtsbehörden in den einzelnen Mitgliedsstaaten, um sicherzustellen, dass Entscheidungen mit ethischen Implikationen, bei denen Automatisierung zum Einsatz kommt, ethisch einwandfrei sind und bleiben.

In seinem Berichtsentwurf über die „Rechte geistigen Eigentums bei der Entwicklung von KI-Technologien" legt derselbe Ausschuss seine Sichtweise in Bezug auf das künftige Verhältnis zwischen geistigem Eigentum und Automatisierung dar. Zunächst trifft der Berichtsentwurf die Aussage, dass „mathematische Methoden nicht patentierbar sind, es sei denn, es handelt sich um Erfindungen technischer Art", um zugleich mit Bezug auf algorithmische Transparenz zu behaupten, dass „Reverse Engineering eine Ausnahme von Geschäftsgeheimnissen darstellt".

Der Bericht geht sogar so weit, Betrachtungen darüber anzustellen, wie der Schutz „der durch KI geschaffenen technischen und künstlerischen Schöpfungen" sicherzustellen sei, „um diese Form des Schaffens zu fördern" und „ist der Ansicht, dass bestimmte durch KI erzeugte Werke mit geistigen Werken vergleichbar sind und daher urheberrechtlich geschützt werden könnten".

Und schliesslich legt der Ausschuss in einem dritten Dokument („Artificial Intelligence and Civil Liability") einen detaillierten „Risk Management Approach" für die Haftpflicht in Zusammenhang mit KI-Technologien vor. Demgemäss „haftet als zentraler Einstiegspunkt für Streitverfahren ausschliesslich die Partei, die am besten in der Lage ist, ein technologiebasiertes Risiko zu überwachen und zu managen".

Wichtige Prinzipien, die den Einsatz von ADM im Strafrecht betreffen, finden sich im „Berichtsentwurf über künstliche Intelligenz im Strafrecht und ihre Verwendung durch die Polizei und Justizbehörden in Strafsachen" des Ausschusses für bürgerliche Freiheiten, Justiz und Inneres. Im Anschluss an eine detaillierte Auflistung tatsächlicher, gegenwärtiger Nutzungen von „KI" – also eigentlich ADM-Systemen – durch die Polizei[16] hält es der Ausschuss „für notwendig, eine klare und faire Regelung für die Zuweisung der rechtlichen Verantwortung für die möglichen nachteiligen Folgen zu schaffen, die durch diese fortgeschrittenen digitalen Technologien verursacht werden".

Anschliessend werden einige Eigenschaften solcher Regelungen beschrieben: keine vollautomatischen Entscheidungen[17], algorithmische Erklärbarkeit, die „für die Nutzer verständlich" ist sowie „eine obligatorische Bewertung der Auswirkungen auf die Grundrechte [...] vor der Einführung oder dem Einsatz von KI-Systemen für die Strafverfolgung oder die Justiz", sowie, ergänzend, „eine periodische obligatorische Prüfung aller KI-Systeme, die von Strafverfolgungs- und Justizbehörden verwendet werden, um die algorithmischen Systeme zu prüfen und zu bewerten, sobald diese in Betrieb sind".

Darüber hinaus fordert der Berichtsentwurf „ein Moratorium für den Einsatz von Gesichtserkennungssystemen für die Strafverfolgung, bis die technischen Standards als vollständig grundrechtskonform angesehen werden können, die erzielten Ergebnisse nicht diskriminierend sind und die Öffentlichkeit Vertrauen in die Notwendigkeit und Verhältnismässigkeit des Einsatzes solcher Technologien hat".

16 Auf S. 5 stellt der Berichtsentwurf fest, dass „KI-Anwendungen, die von den Strafverfolgungsbehörden genutzt werden, Anwendungen wie Gesichtserkennungstechnologien, automatische Nummernschilderkennung, Sprecheridentifizierung, Spracherkennung, Lippenlesetechnologien, akustische Überwachung (d.h. Schusserkennungsalgorithmen), autonome Forschung und Analyse identifizierter Datenbanken, Vorhersage (präventive Polizeiarbeit und Kriminalitäts-Hotspot-Analyse), Verhaltenserkennungswerkzeuge, autonome Werkzeuge zur Erkennung von Finanzbetrug und Terrorismusfinanzierung, Überwachung sozialer Medien (Scraping und Data Harvesting zum Aufspüren von Zusammenhängen), IMSI-Catcher und automatisierte Überwachungssysteme mit unterschiedlichen Erkennungsfähigkeiten (wie Herzschlagerkennung und Wärmebildkameras)" umfassen.

17 "..., dass in Gerichts- und Strafverfolgungskontexten die endgültige Entscheidung immer von einem Menschen getroffen werden muss" (S. 6)

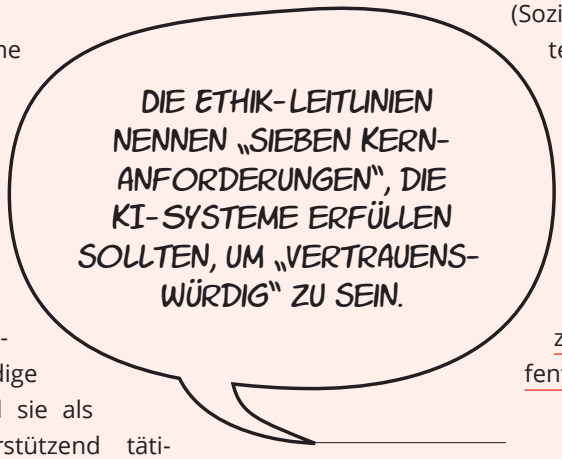
Ziel ist letztlich die Durchsetzung einer grösseren allgemeinen Transparenz solcher Systeme, verbunden mit einer Forderung an die Mitgliedsstaaten, für „ein umfassendes Verständnis“ der bei der Strafverfolgung und in den Justizbehörden zum Einsatz kommenden KI-Systemen zu sorgen sowie – nach Massgabe der Leitlinien eines „öffentlichen Registers” – eine detaillierte Beschreibung „der Art der verwendeten Instrumente, der Arten von Straftaten, auf die sie angewendet werden, und der Unternehmen, deren Instrumente eingesetzt werden“.

Während der vorliegende Bericht verfasst wird, arbeiten der Ausschuss für Kultur und Bildung sowie der Ausschuss für Industrie, Forschung und Energie jeweils an ihren eigenen Berichtsentwürfen.

Alle diese Initiativen mündeten in die Einsetzung eines „Sonderausschusses zu künstlicher Intelligenz im digitalen Zeitalter” (AIDA) am 18. Juni 2020. Mit 33 Mitgliedern und einem ersten Mandat für zwölf Monate wird er eine „Analyse der künftigen Auswirkungen“ der künstlichen Intelligenz auf die Wirtschaft in der EU sowie „insbesondere in den Bereichen Kompetenzen, Beschäftigung, Finanztechnologie, Bildung, Gesundheit, Verkehr, Tourismus, Landwirtschaft, Umwelt, Verteidigung, Industrie, Energie und E-Government” vornehmen.

/ Hochrangige Expertengruppe für KI & KI-Allianz

Die EU-Kommission setzte 2018 eine Hochrangige Expertengruppe für Künstliche Intelligenz (High-Level Expert Group on AI, HLEG) ein. Sie besteht aus 52 Expert·innen und soll die Implementierung der europäischen KI-Strategie unterstützen, indem sie Prinzipien identifiziert, die befolgt werden sollten, um „vertrauenswürdige KI“ zu erreichen. Ausserdem soll sie als Steuerungsausschuss der unterstützend tätigen KI-Allianz eine offene Plattform für verschiedene Interessenvertreter·innen schaffen (die zum Zeitpunkt der Verfassung dieses Berichts bereits mehr als 4.000 Mitglieder hat), um einen breiteren Input für die Arbeit der Hochrangigen Expertengruppe zu gewährleisten.



Nach Veröffentlichung des ersten Entwurfs zu den Ethik-Leitlinien für eine vertrauenswürdige KI im Dezember 2018 wurde nach Stellungnahme von mehr als 500 Beteiligten im April 2019 eine überarbeitete Fassung vorgelegt. Sie macht sich für einen „menschenzentrierten Ansatz“ stark, um während des gesamten Lebenszyklus des KI-Systems Gesetzeskonformität, Ethik und Robustheit zu erreichen. Dennoch bleibt dies ein freiwilliges Rahmenwerk ohne konkrete, anwendbare Empfehlungen für seine Operationalisierung, Implementierung und Durchsetzung.

Zivilgesellschaftliche sowie Verbraucherschutz- und Menschenrechtsorganisationen haben sich dazu geäussert und verlangt, diese Richtlinien in greifbare Menschenrechte zu überführen[18]. So forderte etwa Access Now, eine gemeinnützige Organisation für digitale Rechte und selbst Mitglied der HLEG, die Europäische Kommission dringend zu einer Klarstellung dahingehend auf, wie in einem nächsten Schritt verschiedene Interessengruppen „Vertrauenswürdige KI“ testen, anwenden, verbessern, befürworten und stärken können, und zugleich die Notwendigkeit anzuerkennen, Europas rote Linien festzulegen. In einem Gastbeitrag schrieben zwei andere Mitglieder der HLEG, die Gruppe hätte „anderthalb Jahre gearbeitet“, um jetzt zu erleben, wie „ihre detaillierten Vorschläge“ von der Europäischen Kommission im Weissbuch „zum grössten Teil ignoriert oder nur im Vorbeigehen erwähnt“ worden seien.[19] Ausserdem, so argumentierten sie, hätten Mitglieder der Gruppe, die ursprünglich die Aufgabe gehabt habe, Risiken und „rote Linien“ für KI zu identifizieren, auf autonome Waffensysteme, Citizen-Scoring (Sozialkredit-Systeme) und automatisierte Identifizierung von Einzelpersonen mit Hilfe von Gesichtserkennung hingewiesen und diese als Einsatzgebiete von KI benannt, die es zu vermeiden gelte. Allerdings ist es den Industrievertreter·innen, die den Ausschuss dominieren[20], gelungen, diese Prinzipien streichen zu lassen, bevor der Entwurf veröffentlicht wurde.

18 Z.B. der Europäische Verbraucherverband (BEUC): https://www.beuc.eu/publications/beuc-x-2020-049_response_to_the_ecs_white_paper_on_artificial_intelligence.pdf

19 Mark Coeckelbergh und Thomas Metzinger: Mehr Mut zur KI-Ethik, https://background.tagesspiegel.de/digitalisierung/mehr-mut-zur-ki-ethik

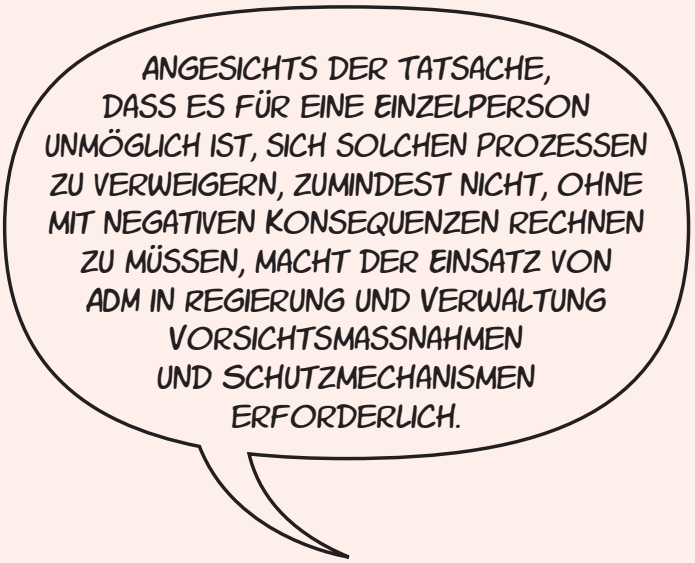
20 Die Gruppe setzte sich aus 24 Unternehmensvertreter·innen, 17 Akademiker·innen, 5 Organisationen der Zivilgesellschaft und 6 sonstigen Mitgliedern, u. a. der Agentur der Europäischen Union für Grundrechte, zusammen.

Diese Unausgewogenheit, die Potenziale von ADM, lässt sich auch durchgängig im zweiten Ergebnispapier der Expertengruppe feststellen. In ihrem Bericht „Policy and investment recommendations for trustworthy AI", veröffentlicht im Juni 2019, finden sich 33 Empfehlungen, die dazu gedacht sind, „Vertrauenswürdige KI in Richtung Nachhaltigkeit, Wachstum, Wettbewerbsfähigkeit und Inklusion zu steuern, während sie gleichzeitig Menschen ermächtigt, schützt und ihnen Nutzen bringt." Das Dokument ist im Wesentlichen ein Aufruf zur Ankurbelung der Einführung und Skalierung von KI im privaten und öffentlichen Sektor mit Hilfe von Investitionen in Tools und Anwendungen, um „vulnerable Bevölkerungsgruppen zu unterstützen" und „niemanden zurückzulassen".

Ungeachtet dessen und trotz aller legitimer Kritik bringen beide Richtlinien entscheidende Besorgnisse und Forderungen in Bezug auf Systeme zur automatisierten Entscheidungsfindung zum Ausdruck. So nennen etwa die Ethik-Leitlinien „sieben Kernanforderungen", die KI-Systeme erfüllen sollten, um „vertrauenswürdig" zu sein. Anschliessend werden die Voraussetzungen aufgelistet, denen die praktische Umsetzung jeder dieser Anforderungen genügen sollte: Vorrang menschlichen Handelns und menschliche Aufsicht; technische Robustheit und Sicherheit; Schutz der Privatsphäre und Datenqualitätsmanagement; Transparenz; Vielfalt, Nichtdiskriminierung und Fairness; gesellschaftliches und ökologisches Wohlergehen; Rechenschaftspflicht. Darüber hinaus ist die Pilotversion einer konkreten „Bewertungsliste für vertrauenswürdige KI" enthalten, die der praktischen Umsetzung dieser hochrangigen Grundsätze dienen soll. Ziel ist es, dass diese Bewertungsliste immer dann zur Anwendung kommt, „wenn KI-Systeme entwickelt, eingeführt und genutzt werden", und dass sie „zukünftig auf den spezifischen Anwendungsfall des KI-Systems zugeschnitten" wird.

Die Liste enthält zahlreiche Aspekte, die mit dem Risiko der Verletzung von Menschenrechten durch ADM-Systeme in Zusammenhang stehen, darunter fehlende menschliche Handlungsmacht und fehlende menschliche Aufsicht; mangelnde technische Robustheit und Sicherheit; die Unfähigkeit, unfaire Verzerrungen zu vermeiden oder einen gleichberechtigten und universellen Zugang zu solchen Systemen zu gewährleisten; sowie das Fehlen eines vernünftigen Zugangs zu den in sie eingespeisten Daten.

Im Zusammenhang bietet die Pilotversion der Bewertungsliste in den Leitlinien nützliche Fragen, die allen helfen können, die ADM-Systeme einsetzen. So fordert sie zum



Beispiel „eine Folgenabschätzung für Anwendungsfälle, bei denen die Möglichkeit einer Beeinträchtigung der Grundrechte besteht". Ebenso enthalten ist die Frage, ob „für den Fall eines selbstlernenden oder autonomen KI-Systems" „spezifischere Vorkehrungen zur Kontrolle und Aufsicht" vorhanden sind und „Prozesse zur Gewährleistung der Qualität und Integrität Ihrer Daten" eingeführt wurden.

Detaillierte Anmerkungen betreffen ebenfalls grundlegende Probleme im Zusammenhang mit ADM-Systemen wie deren Transparenz und Erklärbarkeit. Unter anderem wird gefragt, „inwieweit die Entscheidungen und damit das Ergebnis des KI-Systems nachvollziehbar sind" und „inwieweit die Entscheidung des Systems die Entscheidungsprozesse der Organisation beeinflusst?" Diese Fragen sind höchst relevant, um die Risiken, die mit der Einführung solcher Systeme einhergehen, beurteilen zu können.

Um Verzerrungen und diskriminierende Ergebnisse zu vermeiden, lenken die Leitlinien die Aufmerksamkeit auf „Aufsichtsverfahren, [...] mit deren Hilfe der Zweck, die Einschränkungen, Anforderungen und Entscheidungen des Systems klar und transparent analysiert und angegangen werden könnten" und fordern zugleich eine Beteiligung der Interessenträger während des gesamten Lebenszyklus von KI-Systemen.

Darüber hinaus beinhalten die Politik- und Investitionsempfehlungen die Festlegung roter Linien durch einen institutionalisierten „Dialogs über die Verwendung und die Auswirkungen von KI-Systemen" mit betroffenen Interessenvertretern, einschliesslich Fachleuten der Zivilgesellschaft. Weiterhin mahnen sie dringend ein „Verbot KI-gestützten massenhaften Scorings von Einzelpersonen gemäss der Definition der Ethik-Leitlinien" an und fordern „sehr klare und strenge Regeln für die Überwachung zu nationalen

Sicherheitszwecken und anderen Zwecken, die angeblich dem öffentlichen oder nationalen Interesse dienen". Dieses Verbot würde auch Technologien zur biometrischen Identifikation und zum biometrischen Profiling einschliessen.

Das Dokument sagt auch, dass „eine eindeutige Definition, ob, wann und wie KI zur automatisierten Personenerkennung verwendet werden darf [...] für die Schaffung einer vertrauenswürdigen KI in Zukunft entscheidend" ist und warnt: „Jede Form der Bürgerbewertung kann zum Verlust [von] Autonomie führen und den Grundsatz der Nichtdiskriminierung gefährden" und sollte daher „nur dann eingesetzt werden, wenn es eindeutig gerechtfertigt ist und die Massnahmen verhältnismässig und fair sind". Weiterhin wird betont, dass „Transparenz weder Diskriminierung verhindern noch Fairness gewährleisten" kann. Dies bedeutet, dass die Möglichkeit bestehen muss, sich einem solchen Bewertungsverfahren entziehen zu können, und zwar idealerweise so, dass den Betroffenen daraus keine Nachteile entstehen.

Auf der einen Seite heisst es im Dokument, dass es „anzuerkennen und zu berücksichtigen" gelte, „dass KI-Systeme dem Einzelnen und der Gesellschaft zwar einen erheblichen Nutzen bringen, gleichzeitig jedoch bestimmte Risiken bergen und möglicherweise negative, mitunter schwer absehbare, erkennbare oder messbare Auswirkungen (z.B. im Hinblick auf Demokratie, Rechtsstaatlichkeit, Verteilungsgerechtigkeit oder den menschlichen Geist als solchen) haben können". Auf der anderen Seite ist die Expertengruppe jedoch der Auffassung, dass „unnötig präskriptive Regulierung vermieden werden sollte".

Im Juli 2020 stellte die HLEG im Ergebnis eines Pilotprozesses, an dem sich 350 Interessenvertreter·innen beteiligten, ausserdem ihre abschliessende Bewertungsliste für vertrauenswürdige künstliche Intelligenz vor (Assessment List for Trustworthy Artificial Intelligence, ALTAI) vor.

Ziel der Checkliste, deren Nutzung vollständig freiwillig ist und die keinerlei regulatorische Konsequenzen hat, ist es, die sieben in den Ethik-Leitlinien zu KI der HLEG dargelegten Kernanforderungen in die Praxis umzusetzen. Im Kern geht es darum, allen, die KI-Lösungen implementieren möchten, die mit den Grundwerten der EU vereinbar sind – etwa Designer·innen und Entwickler·innen von KI, Datenwissenschaftler·innen, Beschaffungsbeamt·innen oder -spezialist·innen sowie juristische Berater·innen/ Compliance-Beauftragte – ein Instrumentarium zur Selbstbewertung zur Verfügung zu stellen.

/ Europarat: Zum Schutz von Menschenrechten bei ADM

Das Ministerkomitee des Europarates[21] hat ergänzend zur Einrichtung des Ad-hoc-Ausschusses für künstliche Intelligenz (CAHAI) im September 2019 ein substanzielles und überzeugendes Rahmenwerk veröffentlicht.

Als standardisierendes Instrument gedacht, beschreibt die „Recommendation to Member states on the human rights impacts of algorithmic systems"[22] „fundamentale Herausforderungen", die das Aufkommen und unsere „zunehmende Abhängigkeit" von solchen Systemen mit sich bringen und die relevant sind „für Demokratie und Rechtsstaatlichkeit".

Das Rahmenwerk, zu dem es eine Phase öffentlicher Konsultationen mit ausführlichen Beiträgen von Organisationen der Zivilgesellschaft gab, geht im Hinblick auf den Schutz von Grundwerten und Menschenrechten über das Weissbuch der EU-Kommission hinaus.

Die Empfehlung analysiert die Auswirkungen und sich entwickelnden Konfigurationen algorithmischer Systeme (Anhang A) und untersucht dazu alle Phasen des Prozesses, der zur Schaffung eines Algorithmus führt, das heisst Auftragsvergabe, Design, Entwicklung und laufender Einsatz.

Generell nimmt sie den in den Leitlinien der HLEG verfolgten „menschenzentrierten" KI-Ansatz auf, skizziert aber auch einklagbare „Verpflichtungen der Mitgliedsstaaten" (Anhang

21 Der Europarat ist zweierlei: „ein Regierungsorgan, in dem die nationalen Ansätze zu Problemen von europäischer Tragweite auf Augenhöhe diskutiert werden, und ein Forum zur Erarbeitung kollektiver Antworten auf diese Herausforderungen". Zu seinen Arbeitsfeldern gehören „die politischen Aspekte der europäischen Integration, der Schutz demokratischer Institutionen und der Rechtsstaatlichkeit sowie der Menschenrechte – mit anderen Worten, alle Probleme, die einer abgestimmten europaweiten Lösung bedürfen". Obgleich die Empfehlungen für die Regierungen der einzelnen Mitgliedsstaaten nicht bindend sind, kann das Ministerkomitee in bestimmten Fällen die Regierungen einzelner Mitgliedsstaaten ersuchen, ihm mitzuteilen, was sie auf diese Empfehlungen hin veranlasst haben. (Art. 15b der Satzung). Die Beziehungen zwischen dem Europarat und der Europäischen Union sind niedergelegt im (1) Compendium of Texts governing the relations between the Council of Europe and the European Union (Sammlung von massgebenden Texten zur Regelung der Beziehungen zwischen dem Europarat und der Europäischen Union) sowie (2) Memorandum of Understanding between the Council of Europe and the European Union (Gemeinsame Absichtserklärung zwischen dem Europarat und der Europäischen Union).

22 Unter der Aufsicht des Lenkungsausschusses für Medien und Informationsgesellschaft (CDMSI) und erstellt vom Expertenausschuss für die menschenrechtlichen Dimensionen automatisierter Datenverarbeitung und verschiedener Formen Künstlicher Intelligenz (MSI-AUT)

B) sowie Verantwortlichkeiten der Akteure des privatwirtschaftlichen Sektors (Anhang C). Hinzu kommen Prinzipien wie das der „informationellen Selbstbestimmung“[23] sowie eine Auflistung detaillierter Vorschläge für Rechenschaftslegungsmechanismen und effektive Rechtsbehelfe und die Forderung nach einer Folgenabschätzung für grundlegende Menschenrechte.

Zwar erkennt das Dokument das „signifikante Potenzial digitaler Technologien im Umgang mit gesellschaftlichen Herausforderungen und für gesellschaftlich vorteilhafte Innovationen und wirtschaftliche Entwicklungen“ an, mahnt jedoch gleichzeitig dringend zur Vorsicht. So soll sichergestellt werden, dass diese Systeme weder absichtlich noch zufällig „Ungerechtigkeiten aufgrund von ethnischer Herkunft und Geschlecht sowie andere gesellschaftliche und arbeitsmarktrelevante Ungerechtigkeiten“ fortschreiben, „die unsere Gesellschaften noch immer prägen“.

Im Gegenteil, der Einsatz algorithmischer Systeme sollte auf proaktive und sensible Art und Weise erfolgen, um diese Ungerechtigkeiten anzugehen und „Aufmerksamkeit auf die Bedürfnisse und Stimmen vulnerabler Gruppen zu lenken“.

Am bemerkenswertesten aber ist, dass die Empfehlung das potenziell höhere Risiko für die Verletzung von Menschenrechten aufzeigt, wenn algorithmische Systeme von Mitgliedsstaaten zur Bereitstellung öffentlicher Dienstleistungen und bei politischen Massnahmen eingesetzt werden. Angesichts der Tatsache, dass es für eine Einzelperson unmöglich ist, sich solchen Prozessen zu verweigern, zumindest nicht, ohne mit negativen Konsequenzen rechnen zu müssen, macht der Einsatz von ADM in Regierung und Verwaltung Vorsichtsmassnahmen und Schutzmechanismen erforderlich.

Ausserdem spricht die Empfehlung die Konflikte und Herausforderungen an, die aus Public-Private-Partnerships bei einer Vielzahl von Anwendungen erwachsen („weder klar staatlich noch klar privat“).

Zu den Empfehlungen für die Regierungen der Mitgliedsstaaten gehören unter anderem: Beendigung von Prozessen und Verweigerung des Einsatzes von ADM-Systemen, wenn „menschliche Kontrolle und Aufsicht sich als undurchführbar erweisen“ oder Menschenrechte bedroht sind; Einsatz von ADM-Systemen nur dann, wenn „auf allen Ebenen des Prozesses“ Transparenz, Rechenschaftspflicht, Rechtmässigkeit und der Schutz der Menschenrechte garantiert werden können. Darüber hinaus soll die Überwachung und Bewertung dieser Systeme „kontinuierlich“ erfolgen, „inklusiv und transparent“ sein und sowohl einen Dialog mit allen relevanten Interessengruppen als auch eine Analyse der ökologischen Auswirkungen sowie weiterer potenzieller externer Effekte auf „Bevölkerungen und Umgebungen“ beinhalten.

In Anhang A definiert der Europarat zudem Algorithmen mit „hohem Risiko“, die anderen Gremien als Anregung dienen können. Genauer ist dort ausgeführt, dass „der Begriff „hohes Risiko“ dann Anwendung findet, „wenn es um die Nutzung algorithmischer Systeme in Prozessen oder bei Entscheidungen geht, die ernsthafte Konsequenzen für Einzelpersonen haben können oder in Situationen, wo das Fehlen von Alternativen eine besonders hohe Wahrscheinlichkeit der Verletzung von Menschenrechten mit sich bringt, etwa indem sie Verteilungsungerechtigkeiten erzeugen oder verstärken“.

Das Dokument, zu dessen Annahme es keines einstimmigen Votums der Mitglieder bedurfte, ist nicht bindend.

/ Regulierung terroristischer Online-Inhalte

Nach einer langen Phase, in der es nur schleppend voranging, nahm die Erarbeitung der „Verordnung des Europäischen Parlaments und des Rates zur Verhinderung der Verbreitung terroristischer Online-Inhalte“ 2020 Fahrt auf. Sollte das verabschiedete Regelwerk noch immer automatisierte und proaktive Tools zur Erkennung und Entfernung von Online-Inhalten enthalten, würden diese mit hoher Wahrscheinlichkeit unter Art. 22 der DSGVO fallen.

Der Europäische Datenschutzbeauftragte (EDPS) drückt es so aus: „Da das von dem Vorschlag vorgesehene automatisierte Werkzeug nicht nur zur Entfernung und Vorratsdatenspeicherung von Inhalten (und zugehörigen Daten) im Zusammenhang mit dem Hochlader führen könnte, sondern letztendlich auch zu strafrechtlichen Untersuchungen dieses Hochladers, würden diese Werkzeuge bedeutende

23 „Staaten sollten sicherstellen, dass alle Prozesse im Zusammenhang mit Design, Entwicklung und kontinuierlicher Einführung algorithmischer Systeme den betroffenen Einzelpersonen Wege eröffnen, im Vorhinein Kenntnis über die damit verbundene vorgesehene Verarbeitung ihrer Daten zu erlangen (einschliesslich des Zwecks und der möglichen Ergebnisse) sowie ihre Daten, auch mit Hilfe von Interoperabilität, zu kontrollieren“, heisst es in Anhang B, Abschnitt 2.1.

Auswirkungen auf diese Person haben, sich erheblich nachteilig auf ihre Meinungsfreiheit auswirken und bedeutsame Risiken für ihre Rechte und Freiheiten darstellen" und somit unter Art. 22(2) DSGVO fallen.

Ebenfalls, und zwar unverzichtbar, würde es substanziellere Schutzmechanismen erfordern, als die von der Kommission derzeit vorgesehenen. Das Advocacy Netzwerk European Digital Rights (EDRi) erläutert dazu: „Die vorgeschlagene Verordnung zur Verhinderung der Verbreitung terroristischer Online-Inhalte bedarf einer grundlegenden Überarbeitung, um den Werten der Europäischen Union zu entsprechen und die fundamentalen Rechte und Freiheiten ihrer Bürger·innen zu schützen."

Ein frühzeitiger Sturm der Kritik von Seiten zivilgesellschaftlicher Gruppen und Ausschüssen des Europäischen Parlaments (EP), darunter Meinungen und Analysen der Europäischen Agentur für Menschenrechte (FRA) und von EDRi, sowie auch ein gemeinsamer kritischer Bericht dreier Sonderberichterstatter·innen der Vereinten Nationen als Reaktion auf den ursprünglichen Vorschlag hob die Gefahren für das Recht auf Meinungs- und Informationsfreiheit, die Freiheit und Pluralität der Medien, die unternehmerische Freiheit sowie des Rechtes auf Schutz der Privatsphäre und des Rechtes zum Schutz personenbezogener Daten hervor.

Kritische Punkte sind unter anderem eine unzureichende Definition terroristischer Inhalte, der Geltungsbereich der Verordnung (zum gegenwärtigen Zeitpunkt sind auch Inhalte für Bildungs- und journalistische Zwecke erfasst), die zuvor bereits genannte Forderung nach „proaktiven Massnahmen", das Fehlen effektiver richterlicher Aufsicht, unzureichende Berichtsverpflichtungen für Exekutivbehörden und fehlende Schutzmechanismen für „Fälle, in denen vernünftige Gründe für die Annahme vorliegen, dass grundlegende Rechte betroffen sind" (EDRi 2019).

Wie der Datenschutzbeauftragte (EDPS) betont, sollten zu solchen „angemessenen Schutzmechanismen" der Anspruch auf direktes Eingreifen einer Person und das Anrecht auf Erläuterung einer mit automatisierten Mitteln getroffenen Entscheidung gehören (EDRi 2019).

Vorgeschlagene bzw. geforderte Schutzmechanismen fanden zwar Eingang in den Berichtsentwurf des EP zu dem Vorschlag, dennoch bleibt abzuwarten, wer in der letzten Runde vor der finalen Abstimmung den längeren Atem hat. In Trialogen zwischen dem EU-Parlament, der neuen EU-Kommission und dem Europarat (die im Oktober 2019 begannen und hinter verschlossenen Türen stattfinden) sind einem geleakten Dokument zufolge nur noch geringfügige Änderungen möglich.

Aufsicht und Regulierung

/ Erste Entscheidungen zur Konformität von ADM-Systemen mit der DSGVO

„Obwohl es im Verlauf der Verhandlungen zur DSGVO und der Datenschutzrichtlinie für Strafverfolgungsbehörden keine grosse Debatte um das Thema Gesichtserkennung gab, wurden die gesetzlichen Regelungen so ausgestaltet, dass sie im Laufe der Zeit entsprechend den technologischen Entwicklungen angepasst werden können. [...] Jetzt, da die EU über die Ethik von KI und die Notwendigkeit von Regulierungen diskutiert, ist der Moment gekommen festzustellen, ob Gesichtserkennungstechnologien in einer demokratischen Gesellschaft überhaupt erlaubt werden können. Nur dann, wenn die Antwort auf diese Frage „Ja" lautet, wenden wir uns der Frage zu, auf welche Weise Schutzmechanismen und Rechenschaftspflichten gesetzlich zu verankern sind." – EDPS, Wojciech Wiewiórowski

„Gesichtserkennung ist ein besonders intrusiver biometrischer Mechanismus, der für die betroffenen Personen erhebliche Risiken der Verletzung ihrer Privatsphäre oder ihrer Bürgerrechte mit sich bringt." – (CNIL 2019)

Seit dem letzten „Automating Society"-Bericht hat es auf Grundlage der DSGVO die ersten Strafzahlungen und Entscheidungen gegeben, die in Zusammenhang mit einer Verletzung der von den nationalen Datenschutzbehörden (DPAs) erlassenen gesetzlichen Vorschriften standen. Die folgenden Fallstudien zeigen allerdings die praktischen Grenzen der DSGVO auf, wenn es um Art. 22 geht, der sich auf ADM-Systeme bezieht, und wie sie Datenschutzbehörden darauf zurückwirft, Fall-zu-Fall-Bewertungen vorzunehmen.

In Schweden kam man zu dem Schluss, dass ein für einen begrenzen Zeitraum in einer einzigen Schulklasse durchgeführtes Testprojekt zur Gesichtserkennung diverse Verpflichtungen gemäss der Datenschutzgrundverordnung

(insbes. Art. 2(14) und Art. 9 (2)) verletzte. (Europäischer Datenschutzausschuss 2019)

Ein ähnlicher Fall liegt auf Eis, nachdem die französische Commission Nationale de l'Informatique et des Libertés (CNIL) ihre Besorgnis geäussert hatte, als zwei Oberschulen planten, im Rahmen einer Partnerschaft mit dem US-amerikanischen Tech-Konzern Cisco Gesichtserkennungstechnologie einzusetzen. Die Stellungnahme hat keine rechtliche Wirkung, und die eingereichte Klage läuft noch.[24]

Da die Zustimmung der Nutzer·innen allgemein als ausreichend gilt, um biometrische Daten zu verarbeiten, ist eine Ex-ante-Genehmigung durch Datenschützer nicht erforderlich, um solche Testläufe durchzuführen. In Schweden war das allerdings anders. Der Grund: das bestehende Machtungleichgewicht zwischen dem Daten-Controller und den Datensubjekten. Stattdessen wurden eine angemessene Folgenabschätzung sowie Vorabkonsultationen mit der Datenschutzbehörde für notwendig erachtet.

Der Europäische Datenschutzbeauftragte (EDPS) bestätigte dies:

„Eine Zustimmung muss explizit und aus freien Stücken gegeben werden sowie informiert und bezogen auf den speziellen Anlass erfolgen. Allerdings steht es ausser Zweifel, dass Personen ihre Zustimmung nicht verweigern und noch viel weniger ihre Zustimmung erteilen können, wenn sie Zugang zu öffentlichen Orten haben müssen, die von Überwachungsmechanismen zur Gesichtserkennung abgedeckt sind. […] Und schliesslich ist höchst fraglich, ob die Technologie verpflichtenden Prinzipien wie Datenminimierung und Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen (data protection by design) folgt. Gesichtserkennungstechnologie hat bisher noch nie vollständig korrekt funktioniert, und dies hat ernsthafte Konsequenzen für alle Personen, die fälschlicherweise identifiziert werden, sei es als Kriminelle oder anderweitig. […] Allerdings wäre es ein Fehler, den Fokus ausschliesslich auf Fragen des Datenschutzes zu richten. Vielmehr handelt es sich hier um eine für eine demokratische Gesellschaft fundamentale ethische Frage." (EDPS 2019)

Access Now kommentierte:

„Im Zuge der zunehmenden Entwicklung von Projekten zur Gesichtserkennung sehen wir bereits, dass die DSGVO nützliche Vorschriften zum Schutz von Menschenrechten enthält, die gerichtlich gegen die unrechtmässige Sammlung und Nutzung sensibler Daten wie etwa biometrischer Daten durchgesetzt werden können. Allerdings könnten der unverantwortliche und häufig unbegründete Hype um die Effizienz solcher Technologien und die dahinterstehenden ökonomischen Interessen dazu führen, dass Zentral- und Lokalregierungen sowie private Unternehmen versuchen, das Gesetz zu umgehen."

/ Von der Polizei in Südwales eingesetzte automatisierte Gesichtserkennung für rechtswidrig erklärt

Im Laufe des Jahres 2020 erlebte das Vereinigte Königreich den ersten prominenten Anwendungsfall der Law Enforcement Directive[25] (Durchsetzungsrichtlinie) zum Einsatz von Gesichtserkennungstechnologie an öffentlichen Orten durch die Polizei. Die Entscheidung gilt als bedeutsamer Präzedenzfall zu einem heiss diskutierten Thema, und das Urteil fand grosse Aufmerksamkeit sowohl bei Akteur·innen der Zivilgesellschaft wie auch bei Rechtsgelehrten in ganz Europa und darüber hinaus.[26]

Der Kläger war Ed Bridges, ein 37 Jahre alter Mann aus Cardiff. Inhalt seiner Klage war der Vorwurf, sein Gesicht sei

24 Siehe Kapitel France in der Gesamtausgabe des Automating Society Reports unter https://automatingsociety.algorithmwatch.org/report2020/france/ sowie (Kayalki 2019)

25 Die Durchsetzungsrichtlinie, in Kraft seit Mai 2018, „regelt den Umgang mit der Verarbeitung persönlicher Daten durch Datenverantwortliche zum ‚Zwecke der Durchsetzung von Recht und Gesetz – welche/was nicht unter den Geltungsbereich der DSGVO fallen/fällt". https://www.dataprotection.ie/en/organisations/law-enforcement-directive

26 Die Entscheidung wurde am 4. September 2019 vom Obersten Gericht in Cardiff in der Sache Bridges ./. Die Polizei von Südwales getroffen. (Hight Court of Justice 2019)

DIE EU BETEILIGT SICH MIT 8.199.387,75 EURO AN DEM PROJEKT ZUR ENTWICKLUNG „VERBESSERTER METHODEN ZUR GRENZÜBERWACHUNG" ZUR BEKÄMPFUNG IRREGULÄRER MIGRATION.

ohne seine Einwilligung sowohl während der Weihnachtseinkäufe 2017 als auch bei einem friedlichen Protest gegen den Einsatz von Waffen ein Jahr später gescannt worden.

In seinem ersten Urteil hatte das Gericht den Einsatz von Technologie zur automatischen Gesichtserkennung (Automated Facial Recognition Technology AFR) durch die Polizei von Südwales noch für rechtmässig und angemessen erklärt. Doch die Bürgerrechtsorganisation „Liberty" legte Berufung gegen das Urteil ein, und das Berufungsgericht für England und Wales entschied gegen die Abweisung der Klage durch das Oberste Gericht und erklärte die Technologie am 11. August 2020 für illegal.

In der Urteilsbegründung gegen die Polizei von Südwales in drei von fünf Anklagepunkten führte das Berufungsgericht aus, das bestehende Rahmenwerk zum Einsatz von AFR weise „fundamentale Defizite" auf und sein Einsatz entspreche nicht dem Grundsatz der „Verhältnismässigkeit"; zudem war keine angemessene Datenschutzfolgeabschätzung (Data Protection Impact Assessment, DPIA) durchgeführt worden, da mehrere entscheidende Schritte hierfür fehlten.

Das Gericht urteilte allerdings nicht, dass das System diskriminierende Ergebnisse aufgrund von ethnischer Zugehörigkeit oder Geschlecht erbringe, da die Polizei von Südwales keine ausreichenden Belege darüber gesammelt habe, um darüber gerichtlich befinden zu können.[27] Unabhängig davon war das Gericht der Auffassung, die folgende beachtenswerte Anmerkung hinzufügen zu müssen: „Da es sich bei AFR um eine neuartige und kontroverse Technologie handelt, hoffen wir, dass alle Polizeibehörden, die planen, sie in der Zukunft einzusetzen, den Wunsch haben mögen, sich dahingehend rückzuversichern, dass alles vernünftigerweise Mögliche unternommen wurde, um sicherzustellen, dass die verwendete Software keine geschlechtsspezifischen oder ethnischen Verzerrungen aufweist."

Nach dem Urteil forderte Liberty die Polizei von Südwales und andere Polizeibehörden auf, die Nutzung von Gesichtserkennungstechnologie zu beenden.

27 Die Polizeibehörde behauptete, keinen Zugang zur demografischen Zusammensetzung des für den Test verwendeten Datenpakets für den angewendeten Algorithmus „Neoface" gehabt zu haben. Das Gericht stellte fest, es bleibe „dennoch die Tatsache, dass sich die SWP zu keinem Zeitpunkt ausreichend versichert hat, und zwar weder selbst noch mit Hilfe einer unabhängigen Überprüfung, dass das Softwareprogramm in diesem Fall keinen inakzeptablen geschlechtsspezifischen oder rassistischen Bias" habe.

ADM in der Praxis: Grenzverwaltung und Überwachung

Während die EU-Kommission und ihre Interessenvertreter·innen darüber debattierten, ob Gesichtserkennungstechnologien reguliert oder verboten werden sollten, liefen in ganz Europa bereits Pilotprojekte zum testweisen Einsatz solcher Systeme.

Dieser Abschnitt beleuchtet eine fundamentale und häufig übersehene Verbindung, nämlich die zwischen Biometrie und den Grenzmanagementsystemen der EU. Hier zeigt sich deutlich, wie Technologien, die diskriminierende Ergebnisse erbringen können, auf Einzelindividuen – z. B. Migrant·innen – angewandt werden könnten, die bereits am stärksten von Diskriminierung betroffen sind.

/ Gesichtserkennung und die Nutzung biometrischer Daten in Politik und Praxis der EU

Im Laufe des Jahres 2019 zogen Gesichtserkennung und andere biometrische Identifikationstechnologien eine Menge Aufmerksamkeit von Regierungen, der EU, der Zivilgesellschaft sowie Bürger·innen – und Menschenrechtsorganisationen auf sich, insbesondere in den Bereichen Strafverfolgung und Grenzsicherung.

Die Europäische Kommission beauftragte ein Konsortium staatlicher Agenturen, „den aktuellen Stand des Einsatzes von Gesichtserkennung bei Ermittlungsverfahren in allen EU-Mitgliedsstaaten festzustellen" um auf einen „möglichen Austausch von Gesichtsdaten" hinzuarbeiten. Sie beauftragten die Beraterfirma Deloitte mit der Durchführung einer Machbarkeitsstudie zur Ausweitung des Prüm-Systems zur Abfrage von Gesichtsbildern. Das EU-weit eingesetzte System ermöglicht den wechselseitigen automatisierten Zugriff auf DNA-Profile, daktyloskopische Daten und Daten aus nationalen Fahrzeugregistern. Es besteht die Sorge, dass eine europaweite Datenbank mit Gesichtsbildern allgegenwärtige, ungerechtfertigte oder illegale Überwachung zur Folge haben könnte.

/ Grenzenlose Grenzverwaltungssysteme

Wie schon in der vorherigen Ausgabe von Automating Society berichtet, ist die Implementierung von übergreifenden, interoperablen Grenzmanagementsystemen, die auf einen Vorschlag der EU-Kommission von 2013 zurückgehen, bereits europaweit im Gange. Obwohl die angekündigten neuen Systeme (EES, ETIAS[28], ECRIS-TCN[29]) erst 2022 in Betrieb gehen sollen, hat die Verordnung über das Einreise-/Ausreisesystem (EES) Gesichtsbilder als biometrische Identifikatoren eingeführt und die Nutzung von Gesichtserkennungstechnologie zu Verifizierungszwecken erstmals in einem EU-Gesetz verankert.[30]

Die Änderungen wurden von der Europäischen Agentur für Menschenrechte (FRA) bestätigt: „Es ist zu erwarten, dass in den auf EU-Ebene eingesetzten umfänglichen IT-Systemen, die für Asylverfahren, Migration und Sicherheitszwecke genutzt werden, eine systematischere Einführung der Verarbeitung von Gesichtsbildern stattfinden wird […], sobald die dafür notwendigen juristischen und technischen Schritte vollzogen sind".

Ana Maria Ruginis Andrei von der Europäischen Agentur für das Betriebsmanagement von IT-Grosssystemen im Raum der Freiheit, der Sicherheit und des Rechts (eu-LISA) zufolge wurde diese erweiterte Interoperabilitätsarchitektur „aufgebaut, um den perfekten Motor für den erfolgreichen Kampf gegen die Bedrohungen der inneren Sicherheit und eine effektive Kontrolle von Migration zu schaffen und blinde Flecken beim Identitätsmanagement zu tilgen". In der Praxis bedeutet dies die „Vorhaltung der Fingerabdrücke, Gesichtsbilder und anderer persönlicher Daten von bis zu 300 Millionen Nicht-EU-Bürgern durch Zusammenführung

28 ETIAS (EU Travel Information and Authorisation System) ist das von eu-LISA entwickelte neue „Visa Waiver"-System für das EU-Grenzmanagement. „Die bei Antragstellung eingegebenen Informationen werden automatisch mit bereits existierenden EU-Datenbanken (Eurodac, SIS und VIS), künftig auch den Systemen EES und ECRIS-TCN sowie mit den relevanten Interpol-Datenbanken abgeglichen. So wird eine verbesserte Überprüfung im Hinblick auf Sicherheitsrisiken, Risiken im Zusammenhang mit illegaler Einwanderung und Gesundheitsrisiken erreicht." (ETIAS 2019)

29 Das Europäische Strafregisterinformationssystem – Drittstaatsangehörige (ECRIS-TCN), das von eu-LISA entwickelt werden soll, wird eine Erweiterung des bereits bestehenden EU-Strafregisterinformationssystems (ECRIS) in Form eines zentralisierten „hit/no-hit"-Systems sein, das Informationen über Nicht-EU-Bürger abrufbar macht, sofern sie in einem EU-Staat vorbestraft sind.

30 EES wird den Betrieb im 1. Quartal 2022 aufnehmen, ETIAS wird bis Ende 2022 folgen. Beide Systeme sollen „im Zuständigkeitsbereich des Rates für Justiz und Inneres (JI-Rat) als Game Changer" fungieren.

der Daten fünf voneinander getrennter Systeme.“ (Campbell 2020)

/ ETIAS: Automatisierte Screenings bei Grenzübertritt

Das Europäische Reiseinformations- und Genehmigungssystem (ETIAS), das zum Zeitpunkt der Erstellung dieses Berichts noch immer nicht in Betrieb ist, wird verschiedene Datenbanken nutzen, um Reisende aus Nicht-EU-Staaten (die kein Visum bzw. keinen Visa-Waiver benötigen) vor ihrer Einreise nach Europa einer automatisierten digitalen Sicherheitskontrolle zu unterziehen.

Das System wird künftig Daten für die verbesserte „Identifizierung potenzieller Sicherheitsrisiken oder Risiken irregulärer Einwanderung“ sammeln und analysieren (ETIAS 2020). Ziel ist „die Stärkung von Grenzkontrollen; die Vermeidung bürokratischer Hürden und Verspätungen für Reisende, die an den Grenzen vorstellig werden; und die Sicherstellung einer koordinierten und harmonisierten Risikobewertung von Einreisenden aus Drittstaaten“. (ETIAS 2020)

Ann-Charlotte Nygård, Leiterin der Abteilung Technical Assistance and Capacity Building bei der FRA, sieht bei ETIAS zwei spezielle Risiken: „erstens die Nutzung von Daten, die zu einer unbeabsichtigten Diskriminierung bestimmter Gruppen führen könnte, etwa dann, wenn Antragsteller·innen einer bestimmten ethnischen Gruppe mit einem hohen Immigrationsrisiko angehören; zweitens in Zusammenhang mit einem Sicherheitsrisiko, dass auf der Basis im Herkunftsland erhaltener Vorstrafen bewertet wird. Manche dieser früheren Verurteilungen könnten von Europäern als unangemessen angesehen werden, so zum Beispiel Urteile gegen Angehörige der LGBT-Community in bestimmten Ländern. Um dies zu verhindern, [...] müssen Algorithmen auditiert werden, damit sichergestellt ist, dass sie nicht diskriminierend wirken, und diese Art des Auditings muss interdisziplinäre Expert·innen einbeziehen.“ (Nygård 2019)

/ iBorderCtrl: Gesichtserkennung und Risikoeinstufung an den Grenzen

iBorderCtrl war ein Projekt, an dem Sicherheitsbehörden aus Ungarn, Lettland und Griechenland beteiligt waren und das zum Ziel hatte „schnellere und genauere Grenzkontrollen für Drittstaatenangehörige beim Grenzeintritt in EU-Mitgliedsstaaten an Landübergängen zu ermöglichen”. iBorderCtrl nutzte Gesichtserkennungstechnologie, einen Lügendetektor sowie ein Scoring-System. Auf Basis der gewonnenen Daten teilte es menschlichen Grenzbeamt·innen mit, ob es jemanden für gefährlich hielt oder ob es die Berechtigung einer Person zum Grenzübertritt in Zweifel zog.

Das Projekt lief Ende August 2019 aus, und die Ergebnisse – für jedwede potenzielle EU-weite Implementierung des Systems – waren widersprüchlich.

Obwohl „noch definiert werden muss, inwieweit das System oder Teile davon zum Einsatz kommen werden“, sieht die Projekt-Website unter „Ergebnisse“ „die Möglichkeit, die ähnlichen Funktionalitäten des neuen ETIAS-Systems zu integrieren“ und die „Fähigkeiten“ dahingehend zu erweitern, dass „das Grenzübertrittsverfahren dorthin gebracht werden kann, wo die Reisenden sind (Bus, Auto, Zug etc.)“.

Allerdings wurden weder die Module näher spezifiziert, auf die sich dies bezieht, noch die ADM-bezogenen Tools einer öffentlichen Bewertung unterzogen.

Gleichzeitig findet sich in den FAQ zum Projekt die Bestätigung, dass das getestete System „derzeit für einen Einsatz an der Grenze als nicht geeignet” angesehen wird, und zwar, „zum einen, weil es sich um einen Prototyp handelt, zum anderen wegen der technologischen Infrastruktur auf EU-Ebene“. Dies bedeutet, dass „für eine Nutzung durch die Grenzbehörden weitere Entwicklungsschritte und eine Integration in die bestehenden EU-Systeme erforderlich“ wären.

Insbesondere – und zwar ungeachtet der Tatsache, dass das iBorderCtrl-Konsortium in der Lage war, die Funktionsfähigkeit einer solchen Technologie für Grenzkontrollen im Prinzip aufzuzeigen – ist ebenso klar, dass vor jedem tatsächlichen Einsatz ethischen, rechtlichen und gesellschaftlichen Auflagen Rechnung getragen werden muss.

/ Vergleichbare Horizon2020-Projekte

Im Rahmen des Programms Horizon2020 nahmen diverse Anschlussprojekte die Testung und Entwicklung neuer Systeme und Technologien für Grenzmanagement und -kontrolle in den Fokus. Diese sind auf der Webseite des Forschungs- und Entwicklungsinformationsdienstes der Europäischen Gemeinschaft (CORDIS) aufgelistet, wo Informationen zu allen EU-finanzierten Forschungsaktivitäten in diesem Zusammenhang vorgehalten werden.

Die Seite verzeichnet derzeit unter dem Programm/Thema „H2020-EU.3.7.3. – Strengthen security through border management" (Verstärkte Sicherheit durch Grenzmanagement) der Europäischen Union 38 laufende Projekte. Das übergeordnete Programm „Sichere Gesellschaften – Schutz der Freiheit und Sicherheit Europas und seiner Bürger", verfügt über ein Gesamtbudget von knapp 1,7 Milliarden Euro und finanziert 350 Projekte. Es bekämpft nach eigener Aussage „Unsicherheit, egal, ob sie nun aus Verbrechen, Gewalt, Terrorismus, Natur- oder menschengemachten Katastrophen, Cyber-Attacken, Verletzungen der Privatsphäre oder anderen Formen gesellschaftlicher und ökonomischer Verwerfungen erwächst, von denen Bürgerinnen und Bürger zunehmend betroffen sind", und zwar mit Hilfe von Projekten, die vorrangig der Entwicklung neuer, auf KI und ADM basierender technologischer Systeme dienen.

Einige Projekte sind bereits beendet worden und/oder ihre Anwendungen sind schon im Einsatz, so zum Beispiel FastPass, ABC4EU, MOBILEPASS und EFFISEC – sie alle dienten der Untersuchung von Anforderungen an „integrierte, interoperable, automatisierte Grenzkontrollen (ABC)", Identifikationssysteme und „smarte" Schleusen an diversen Grenzübergängen.

TRESSPASS ist ein laufendes Projekt, das im Juni 2018 gestartet wurde und im November 2021 beendet sein wird. Die EU ist mit einem Beitrag in Höhe von knapp zehn Millionen Euro an dem Projekt beteiligt. Ziel der Koordinator·innen von iBorderCRL (wie auch von FLYSEC und XP-DITE) ist es, die von iBorderCRL „implementierten und getesteten Ergebnisse und Konzepte wirksam einzusetzen" und „sie innerhalb eines starken gesetzlichen und ethischen Rahmenwerks" zu einer „multimodalen, risikobasierten Grenzübertrittslösung [zu] erweitern ."(Horizon2020 2019)

Das Projekt ist nach eigener Aussage darauf ausgerichtet, die alte, unmoderne „regelbasierte" Strategie für Sicherheitskontrollen an Grenzübergangspunkten in eine neue „risikobasierte" umzuwandeln. Dies schliesst die Anwendung biometrischer und sensorischer Technologien sowie ein risikobasiertes Managementsystem und sachbezogene Modelle für die Bewertung von Identität, Hab und Gut, Fähigkeiten und Absichten ein. Es zielt darauf ab, mit Hilfe von „Links zu Altsystemen und externen Datenbanken wie VIS/SIS/PNR" Kontrollen zu ermöglichen und sammelt zu Sicherheitszwecken Daten aus all den vorgenannten Datenquellen.

Ein anderes Pilotprojekt ist FOLDOUT. Es startete im September 2018 und wird im Februar 2022 beendet sein. Der Beitrag der EU zu dem Projekt, dass sich mit der Entwicklung „verbesserter Methoden zur Grenzüberwachung" befasst, um illegaler Einwanderung entgegenzuwirken, und dabei einen speziellen Fokus auf „das Aufspüren von Menschen in dichter Vegetation in extremen Klimaverhältnissen" legt, beläuft sich auf 8.199.387,75 Euro. Durch Kombination „verschiedener Sensoren und Technologien sowie einer intelligenten Einbindung derselben in eine effektive und robuste Detektionsplattform" sollen Reaktionsszenarien vorgeschlagen werden. Pilotversuche dazu laufen in Bulgarien, flankiert von Demonstrationsmodellen in Griechenland, Finnland und Französisch-Guayana.

MIRROR (Migrationsrisiken aufgrund von Missverständnissen im Hinblick auf Chancen und Anforderungen) startete Anfang Juni 2019 und wird Ende Mai 2022 beendet sein. Der Beitrag der EU zu dem Projekt, das darauf abzielt zu „verstehen, wie Europa im Ausland wahrgenommen wird; Diskrepanzen zwischen Vorstellung und Realität aufzudecken; Fälle von gezielter Manipulation durch Medien aufzuspüren; und Fähigkeiten zu entwickeln, solchen Missverständnissen und den daraus erwachsenden Sicherheitsbedrohungen entgegenzuwirken", beläuft sich auf rund 5,2 Millionen Euro. Basierend auf „wahrnehmungsspezifischer Bedrohungsanalyse, wird das MIRROR-Projekt Methoden der automatisierten Text-, Multimedia- und soziale Netzwerk-Analyse für verschiedene Arten von Medien (einschliesslich sozialer Medien) mit empirischen Studien verknüpfen", um „Technologien" zu entwickeln und „handlungsorientierte Erkenntnisse" zu gewinnen, „die [...] von Grenzbehörden und politischen Entscheidungsträgern sorgfältig validiert sind, z. B. über Pilotprojekte".

Zu weiteren, bereits abgeschlossenen Projekten, die jedoch Erwähnung finden sollen, gehört Trusted Biometrics under Spoofing Attacks (TABULA RASA), das im November 2010 startete und im April 2014 beendet wurde. Es analysierte „die Schwächen von Software für biometrische Identifikationsprozesse in Bezug auf ihre Anfälligkeit für Spoofing, die zu einer verminderten Effizienz biometrischer Geräte führt". Ein weiteres Projekt, Bodega, das im Juni 2015 startete und im Oktober 2018 beendet wurde, untersuchte, auf welche Weise sich „menschlicher Sachverstand" nutzen lässt, wenn es um die „Einführung smarterer Grenzkontrollsysteme wie etwa biometriebasierte automatisierte Schleusen und Selbstbedienungssysteme" geht.

Quellen:

Access Now (2019): Comments on the draft recommendation of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems https://www.accessnow.org/cms/assets/uploads/2019/10/Submission-on-CoE-recommendation-on-the-human-rights-impacts-of-algorithmic-systems-21.pdf

AlgorithmWatch (2020): Our response to the European Commission's consultation on AI https://algorithmwatch.org/en/response-european-commission-ai-consultation/

Campbell, Zach/Jones, Chris (2020): Leaked Reports Show EU Police Are Planning a Pan-European Network of Facial Recognition Databases https://theintercept.com/2020/02/21/eu-facial-recognition-database/

CNIL (2019): French privacy regulator finds facial recognition gates in schools illegal https://www.biometricupdate.com/201910/french-privacy-regulator-finds-facial-recognition-gates-in-schools-illegal

Coeckelbergh, Mark / Metzinger, Thomas(2020): Europe needs more guts when it comes to AI ethics https://background.tagesspiegel.de/digitalisierung/europe-needs-more-guts-when-it-comes-to-ai-ethics

Committee of Ministers (2020): Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154

Commissioner for Human Rights (2020): Unboxing artificial intelligence: 10 steps to protect human rights https://www.coe.int/en/web/commissioner/-/unboxing-artificial-intelligence-10-steps-to-protect-human-rights

Committee on Legal Affairs (2020): Draft Report: With recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies https://www.europarl.europa.eu/doceo/document/JURI-PR-650508_EN.pdf

Committee on Legal Affairs (2020): Artificial Intelligence and Civil Liability https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf

Committee on Legal Affairs (2020): Draft Report: On intellectual property rights for the development of artificial intelligence technologies https://www.europarl.europa.eu/doceo/document/JURI-PR-650527_EN.pdf

Committee on Civil Liberties, Justice and Home Affairs (2020): Draft Report: On artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters https://www.europarl.europa.eu/doceo/document/LIBE-PR-652625_EN.pdf

Delcker, Janosch(2020): Decoded: Drawing the battle lines – Ghost work – Parliament's moment https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-drawing-the-battle-lines-ghost-work-parliaments-moment/?utm_source=POLITICO.EU&utm_campaign=5a7d137f82-EMAILCAMPAIGN_2020_09_09_08_59&utm_medium=email&utm_term=0_10959edeb55a7d137f82-190607820

Data Protection Commission(2020): Law enforcement directive https://www.dataprotection.ie/en/organisations/law-enforcement-directive

EDRi (2019): FRA and EDPS: Terrorist Content Regulation requires improvement for fundamental rights https://edri.org/our-work/fra-edps-terrorist-content-regulation-fundamental-rights-terreg/

GDPR (Art 22): Automated individual decision-making, including profiling https://gdpr-info.eu/art-22-gdpr/

European Commission (2018): White paper: On Artificial Intelligence – A European approach to excellence and trust https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission (2020): A European data strategy https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

European Commission (2020): Shaping Europe's digital future – Questions and Answers https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_264

European Commission (2020): White Paper on Artificial Intelligence: Public consultation towards a European approach for excellence and trust https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence

European Commission (2018): Security Union: A European Travel Information and Authorisation System – Questions & Answers https://ec.europa.eu/commission/presscorner/detail/en/MEMO_18_4362

European Data Protection Board (2019): Facial recognition in school renders Sweden's first GDPR fine https://edpb.europa.eu/news/national-news/2019/facial-recognition-school-renders-swedens-first-gdpr-fine_en

European Parliament (2020): Artificial intelligence: EU must ensure a fair and safe use for consumers https://www.europarl.europa.eu/news/en/press-room/20200120IPR70622/artificial-intelligence-eu-must-ensure-a-fair-and-safe-use-for-consumers

European Parliament (2020): On automated decision-making processes: ensuring consumer protection and free movement of goods and services https://www.europarl.europa.eu/doceo/document/B-9-2020-0094_EN.pdf

European Data Protection Supervisor (2019): Facial recognition: A solution in search of a problem? https://edps.europa.eu/press-publications/press-news/blog/facial-recognition-solution-search-problem_de ETIAS (2020): European Travel Information and Authorisation System (ETIAS) https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders/etias_en

ETIAS (2019): European Travel Information and Authorisation System (ETIAS) https://www.eulisa.europa.eu/Publications/Information%20Material/Leaflet%20ETIAS.pdf

Horizon2020 (2019): robusT Risk basEd Screening and alert System for PASSengers and luggage https://cordis.europa.eu/project/id/787120/reporting

High Court of Justice (2019): Bridges v. the South Wales Police https://www.judiciary.uk/wp-content/uploads/2019/09/bridges-swp-judgment-Final03-09-19-1.pdf

High-Level Expert Group on Artificial Intelligence (2020): Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Hunton Andrew Kurth (2020): UK Court of Appeal Finds Automated Facial Recognition Technology Unlawful in Bridges v South Wales Police https://www.huntonprivacyblog.com/2020/08/12/uk-court-of-appeal-finds-automated-facial-recognition-technology-unlawful-in-bridges-v-south-wales-police/

Kayalki, Laura (2019): French privacy watchdog says facial recognition trial in high schools is illegal https://www.politico.eu/article/french-privacy-watchdog-says-facial-recognition-trial-in-high-schools-is-illegal-privacy/

Kayser-Bril, Nicolas (2020): EU Commission publishes white paper on AI regulation 20 days before schedule, forgets regulation https://algorithmwatch.org/en/story/ai-white-paper/

Leyen, Ursula von der (2019): A Union that strives for more – My agenda for Europe https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf

Leyen, Ursula von der (2020): Paving the road to a technologically sovereign Europe https://delano.lu/d/detail/news/paving-road-technologically-sovereign-europe/209497

Leyen, Ursula von der (2020): Shaping Europe's digital future https://twitter.com/eu_commission/status/1230216379002970112?s=11

Leyen, Ursula von der (2019): Opening Statement in the European Parliament Plenary Session by Ursula von der Leyen, Candidate for President of the European Commission https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_19_4230

Nygård, (2019): The New Information Architecture as a Driver for Efficiency and Effectiveness in Internal Security https://www.eulisa.europa.eu/Publications/Reports/eu-LISA%20Annual%20Conference%20Report%202019.pdf

Sabbagh, Dan (2020): This article is more than 1 month old South Wales police lose landmark facial recognition case https://www.theguardian.com/technology/2020/aug/11/south-wales-police-lose-landmark-facial-recognition-case

South Wales Police (2020): Automated Facial Recognition https://afr.south-wales.police.uk/

Valero, Jorge (2020): Vestager: Facial recognition tech breaches EU data protection rules https://www.euractiv.com/section/digital/news/vestager-facial-recognition-tech-breaches-eu-data-protection-rules/

SCHWEIZ

STORY

SEITE 38

FORSCHUNG

SEITE 42

Erfahren Sie mehr dazu im Forschungskapitel unter „Krebsdiagnosen und -behandlungen

Schweizer Polizei automatisiert die Vorhersage von Verbrechen, hat aber wenig vorzuweisen

Eine Überprüfung von drei automatisierten Systemen, die von der Schweizer Polizei und Justiz eingesetzt werden, offenbart ernsthafte Probleme. Die Auswirkungen können aufgrund mangelnder Transparenz nicht beurteilt werden.

Von Nicolas Kayser-Bril

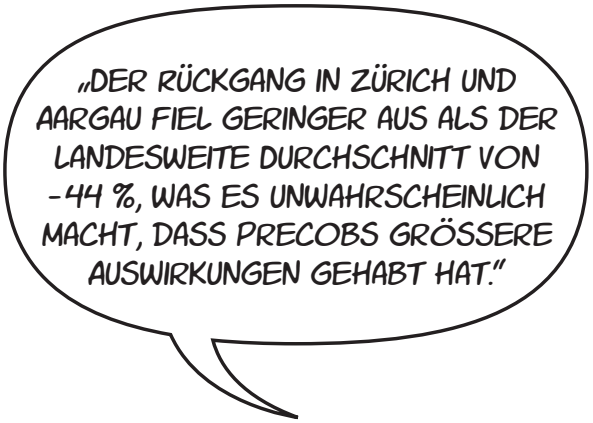
In der Schweiz ist seit 2013 Precobs im Einsatz. Das Tool wird von einem deutschen Hersteller vertrieben, der aus seiner Verwandtschaft mit „Minority Report", einer Science-Fiction-Geschichte, in der „Precogs" manche Verbrechen vorhersagen, bevor sie begangen werden, keinen Hehl macht. (Der Plot dreht sich um die häufig von den Precogs ausgelösten Fehlalarme und deren anschliessende Vertuschung durch die Polizei.)

/ Vorhersage von Einbrüchen

Das System versucht anhand von Daten aus der Vergangenheit Einbruchsdiebstähle vorauszusagen, basierend auf der Annahme, dass Einbrecher häufig in kleinen Gebieten operieren. Wird in einem Quartier ein Cluster von Einbrüchen entdeckt, sollte die Polizei in dieser Gegend häufiger Streife fahren, um diese Delikte zu unterbinden, so die Theorie.

Precobs wird in drei Kantonen genutzt: Zürich, Aargau und Basel-Landschaft. Damit ist knapp ein Drittel der Schweizer Bevölkerung abgedeckt. Seit Mitte der 2010er Jahre sind Einbrüche dramatisch zurückgegangen. Die Polizei des Kantons Aargau beklagte sich im April 2020 sogar, es gebe inzwischen zu wenige Einbrüche, um Precobs überhaupt nutzen zu können.

Doch der Rückgang fand in allen Schweizer Kantonen statt, und die drei, die Precobs nutzen, gehören keinesfalls zu denen, in denen er am stärksten war. Zwischen 2012-2014 (als Einbruchsdiebstähle ihren Höchststand erreichten) und 2017-2019 (als Precobs in den drei Kantonen im Einsatz war) ging die Zahl der Einbrüche in allen Kantonen zurück, nicht nur in den dreien, die die Software verwendeten. Der Rückgang in Zürich und Aargau fiel geringer aus als der landesweite Durchschnitt von -44 %, was es unwahrscheinlich macht, dass Precobs grössere Auswirkungen gehabt hat.



Eine von der Universität Hamburg 2019 vorgestellte Studie fand keinerlei Belege für die Effizienz von Predictive-Policing-Lösungen, inklusive Precobs. Es gibt keine öffentlich einsehbaren Dokumente, die belegen, wie viel die Schweizer Behörden für das System ausgegeben haben, aber in München wurden 100.000 Euro für die Installation von Precobs (exklusive Betriebskosten) bezahlt.

/ Prognose von Gewalt gegen Frauen

Sechs Kantone (Glarus, Luzern, Schaffhausen, Solothurn, Thurgau und Zürich) nutzen das System Dyrias-Intimpartner, um die Wahrscheinlichkeit zu prognostizieren, mit der jemand einen gewalttätigen Übergriff auf seine Intimpartnerin verübt. Dyrias, „Dynamic System for the Analysis of Risk", wird ebenfalls von einer deutschen Firma hergestellt und vertrieben.

Einem Bericht des öffentlich-rechtlichen Schweizer Fernsehens SRF zufolge verlangt Dyrias von Polizeibeamt·innen die Beantwortung von 39 Entscheidungsfragen zu einer verdächtigen Person. Anschliessend spuckt das Tool auf einer Skala von 1 (harmlos) bis 5 (gefährlich) eine Punktewertung aus. Die Gesamtzahl der mit Hilfe des Tools überprüften Personen ist unbekannt, allerdings zeigte eine Zählung des SRF 2018, dass 3.000 Einzelpersonen als „gefährlich" eingestuft waren (wobei diese Einstufung nicht nur unter Verwendung von Dyrias zustande kam).

Laut Herstellerangaben identifiziert die Software acht von zehn potenziell gefährlichen Personen. Allerdings betrachtete eine andere Studie die falsch positiven Fälle, also Personen, die als gefährlich eingestuft wurden, die aber tatsächlich harmlos waren. Sie kam zum Ergebnis, dass sechs von zehn durch die Software als gefährlich markierte Personen als „harmlos" einzustufen gewesen wären. Mit anderen Worten, Dyrias rühmt sich nur deshalb guter Ergebnisse, weil es keinerlei Risiken eingeht und die Einstufung „gefährlich" äusserst grosszügig verteilt. (Die Herstellerfirma zweifelt die Ergebnisse an).

Selbst dann, wenn die Performance des Systems verbessert würde, wäre eine Beurteilung seiner Wirkung immer noch unmöglich. Justyna Gospodinov, Co-Direktorin der BIF-Frauenberatung, einer Organisation, die Opfer häuslicher Gewalt unterstützt, sagte gegenüber AlgorithmWatch, die Kooperation mit der Polizei würde sich zwar verbessern und die systematische Risikoeinschätzung sei eine gute Sache, konnte aber zu Dyrias keinerlei Aussage treffen. „Wenn

wir eine Frau neu aufnehmen, wissen wir nicht, ob die Software zum Einsatz kam oder nicht."

/ Rückfallprognose

Seit 2018 nutzen die Justizbehörden aller Kantone der Deutschschweiz ROS (Risikoorientierter Sanktionenvollzug). Das Tool stuft Gefangene in die Kategorie „A" ein, wenn sie kein Rückfallrisiko haben, in die Kategorie „B" , wenn sie ein neues Delikt begehen könnten, oder in die Kategorie „C", wenn sie erneut ein Gewaltverbrechen begehen könnten. Gefangene können mehrfach getestet werden, doch können sie durch Folgetests lediglich von Kategorie A nach Kategorie B oder C wandern, nicht umgekehrt.

Ein Bericht des SRF enthüllte, basierend auf einer Studie von 2013 der Universität Zürich, dass lediglich ein Viertel der in Kategorie C eingestuften Gefangenen nach ihrer Entlassung weitere Delikte begingen (ein falsch positiver Anteil von 75 %), und lediglich einer von fünf, die weitere Delikte verübten, in Kategorie C eingestuft war (ein falsch negativer Anteil von 80 %), . Eine neue Version des Tools kam 2017 auf den Markt, ist jedoch bisher noch nicht geprüft worden.

Die Kantone der französisch- und italienischsprachigen Schweiz arbeiten an einer Alternative zu ROS, die ab 2022 zum Einsatz kommen dürfte. Das Tool behält zwar dieselben drei Kategorien bei, wird aber nur in Verbindung mit Gefangeneninterviews funktionieren, die ebenfalls einem Rating unterzogen werden.

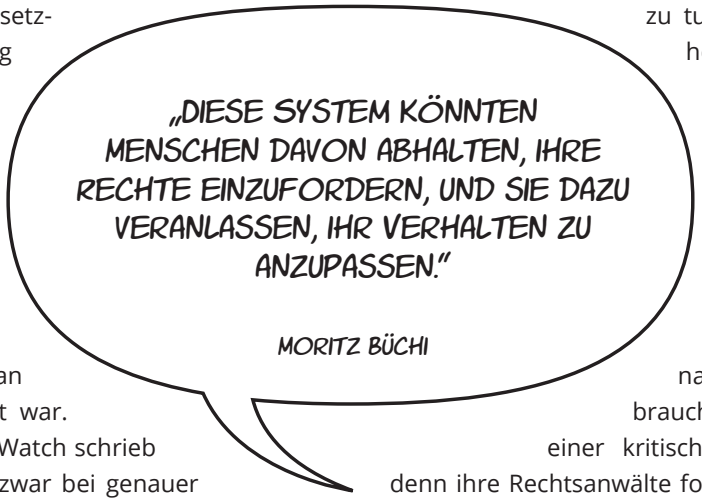
/ Mission: Impossible

Sozialwissenschaftler·innen sind ab und an bei der Prognose allgemeiner Ergebnisse sehr erfolgreich. Im Jahr 2010 sagte das Bundesamt für Statistik voraus, dass die Einwohnerzahl des Landes bis 2020 die 8,5-Millionen-Marke erreichen werde (aktuell leben in der Schweiz 8,6 Millionen Menschen). Doch würden Wissenschaftler·innen niemals versuchen, das Datum vorherzusagen, an dem eine beliebige Person sterben wird: Das Leben ist einfach zu kompliziert.

In dieser Hinsicht unterscheidet sich die Demografie nicht von der Kriminologie. Ungeachtet der Tatsache, dass kommerzielle Anbieter das Gegenteil behaupten, ist die Vorhersage individuellen Verhaltens nahezu unmöglich. 2017 unternahm eine Gruppe von Wissenschaftler·innen den Versuch, das Thema abschliessend zu klären. Sie baten 160 Forscherteams, auf der Basis präziser Daten, die seit ihrer Geburt über sie gesammelt worden waren, Prognosen zur schulischen Leistung von Tausenden Jugendlichen zu treffen, Vorhersagen zu ihrer Wahrscheinlichkeit, zu Hause

„LEDIGLICH EIN VIERTEL DER IN KATEGORIE C EINGESTUFTEN GEFANGENEN BEGINGEN NACH IHRER ENTLASSUNG WEITERE DELIKTE (EIN FALSCH POSITIVER ANTEIL VON 75 %)"

rausgeworfen zu werden, sowie zu vier weiteren Szenarien. Für jedes der Kinder standen Tausende von Datenpunkten zur Verfügung. Die Ergebnisse, die im April 2020 veröffentlicht wurden, sind ernüchternd. Nicht nur gelang es keinem einzigen Team, auch nur ein einziges Ergebnis korrekt vorherzusagen. Auch lieferten diejenigen, die Künstliche Intelligenz einsetzten, keine bessere Leistung ab als jene, die nur einige wenige Variablen und grundlegende statistische Modelle nutzten.



Moritz Büchi, leitender Forscher an der Universität Zürich, ist der einzige Schweizer Akademiker, der an diesem Experiment beteiligt war. In einer E-Mail an AlgorithmWatch schrieb er, kriminelle Delikte seien zwar bei genauer Überprüfung nicht unter den Ergebnissen gewesen, dennoch könnten die aus dem Experiment gewonnenen Erkenntnisse wahrscheinlich auch auf Prognosen über kriminelles Verhalten angewendet werden. Was allerdings nicht bedeute, so Dr. Büchi, dass der Versuch der Erstellung von Prognosen nicht unternommen werden sollte. Doch aus Simulationen einsatzbereite Tools zu machen, lege ihnen den „Mantel der Objektivität" an, der kritisches Denken verhindere, was potenziell vernichtende Konsequenzen für Menschen zur Folge habe, deren Zukunft vorhergesagt werde.

Precobs, so fügte er hinzu, falle nicht in diese Kategorie, da es nicht versuche, das Verhalten bestimmter Individuen vorherzusagen. Mehr Überwachung könne durchaus eine abschreckende Wirkung auf Kriminelle haben. Dennoch verlasse man sich bei der Identifizierung von Hotspots auf historische Daten. Dies könne in einer sich selbst verstärkenden Feedback-Schleife zu einer übermässigen polizeilichen Überwachung von Gemeinden führen, für die in der Vergangenheit Delikte berichtet worden seien.

/ Abschreckwirkungen

Trotz deren sehr durchwachsener Bilanz und der Belege dafür, dass eine Vorhersage individuellen Verhaltens so gut wie unmöglich ist, nutzen die Schweizer Strafvollzugsbehörden nach wie vor Tools, die genau das zu tun behaupten. Deren Beliebtheit verdankt sich zum Teil ihrer Intransparenz. Über Precobs, Dyrias und ROS existieren nur äusserst dürftige öffentlich zugängliche Informationen. Die betroffenen Menschen, die in ihrer überwältigenden Mehrzahl arm sind, haben nur selten die finanziellen Ressourcen, die man braucht, um automatisierte Systeme einer kritischen Prüfung zu unterziehen, denn ihre Rechtsanwälte fokussieren sich für gewöhnlich darauf, die grundlegenden Fakten zu verifizieren, die von der Staatsanwaltschaft vorgetragen werden.

Der Journalist Timo Grossenbacher, der 2018 für den SRF Recherchen über ROS und Dyrias angestellt hatte, sagte AlgorithmWatch, es sei „beinahe unmöglich", Menschen zu finden, die von diesen Systemen betroffen seien Nicht, dass es keine Fälle gebe: Allein ROS werde Jahr für Jahr auf Tausende von Häftlingen angewandt. Vielmehr sei es so, dass die Intransparenz dieser System Watchdogs daran hindere, algorithmische Polizeiarbeit näher zu beleuchten.

Ohne mehr Transparenz könnten diese Systeme Herrn Büchi von der Universität Zürich zufolge eine „Abschreckwirkung" auf die Schweizer Gesellschaft entfalten. „Diese System könnten Menschen davon abhalten, ihre Rechte einzufordern, und sie dazu veranlassen, ihr Verhalten zu modifizieren", schreibt er. „Dies ist eine Form vorauseilenden Gehorsams. Das Bewusstsein, möglicherweise (zu Unrecht) von diesen Algorithmen erwischt zu werden, könnte Menschen dazu bringen, sich stärker konform mit den von ihnen wahrgenommenen gesellschaftlichen Normen zu verhalten. Selbstausdruck und alternative Lebensstile könnten unterdrückt werden."

Forschung

VON PROF. DR. JUR. NADJA BRAUN BINDER UND CATHERINE EGLI

Kontextualisierung

Die Schweiz ist ein zutiefst föderalistisches Land mit einer ausgeprägten Gewaltenteilung. Daher werden technische Innovationen im öffentlichen Sektor häufig zuerst in den Kantonen entwickelt.

Ein Beispiel dafür ist die Einführung einer elektronischen Identität (eID). Auf Bundesebene ist der zur Einführung der eID erforderliche Gesetzgebungsprozess noch nicht abgeschlossen, dagegen in einem Kanton eine offiziell genehmigte elektronische Identität bereits in Betrieb. Schaffhausen führte 2017 im Rahmen der kantonalen Schweizer eGovernment-Strategie als erster Kanton eine digitale Identität für seine Einwohner·innen ein. Mit Hilfe dieser eID können Bürger·innen online unter anderem einen Antrag auf ein Angelfischer-Patent stellen, die Steuerverbindlichkeiten für Gewinne aus Immobilienvermögen oder Kapitalerträgen berechnen oder eine Fristverlängerung für die Abgabe ihrer Steuererklärung beantragen.

Darüber hinaus kann bei der Kindes- und Erwachsenenschutzbehörde ein Erwachsenenkonto eingerichtet werden, und Ärzt·innen können eine Kostengutsprache für Patient·innen beantragen, die in ein Spital ausserhalb ihres Kantons eingewiesen werden. Ein anderes Beispiel, das als Teil eines Pilotprojektes im September 2019 in demselben Kanton startete, bietet Einwohner·innen die Möglichkeit, via Smartphone Auszüge aus dem Betreibungsregister zu bestellen. Diese Dienstleistungen werden laufend ausgebaut (Schaffhauser 2020). Die eID selbst sich ist zwar kein ADM-Prozess, stellt aber dennoch eine unabdingbare Voraussetzung für den Zugang zu digitalen staatlichen Dienstleistungen dar und ist daher auch geeignet, den Zugang zu automatisierten Abläufen zu erleichtern, etwa im Bereich Steuern. Die Tatsache, dass ein einzelner Kanton auf diesem Weg bereits ein Stück weiter vorangeschritten ist als die gesamte Schweiz auf Bundesebene, ist typisch für die Schweiz.

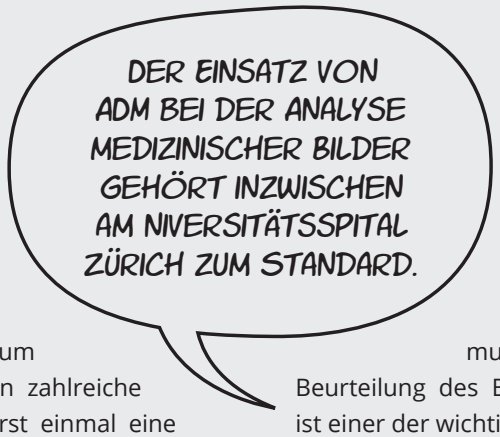
Eine weitere Charakteristik der Schweiz ist die direkte Demokratie. So ist etwa das Gesetzgebungsverfahren für eine nationale eID noch nicht abgeschlossen, weil es über den entsprechenden Gesetzentwurf des Parlaments demnächst ein Referendum geben wird (eID-Referendum 2020). Diejenigen, die das Referendum angestrengt haben, stehen einer offiziellen eID nicht grundsätzlich ablehnend gegenüber, möchten aber private Unternehmen daran hindern, die eID zu vergeben und sensible private Daten zu verwalten.

Ein weiteres Element, das ebenfalls in Betracht gezogen werden muss, ist die gute Wirtschaftslage des Landes. Sie ermöglicht grossartige Fortschritte auf einzelnen Gebieten, so zum Beispiel automatisierte Entscheidungen im medizinischen Bereich und zahlreichen Forschungsfeldern. Obwohl es in der Schweiz aufgrund der ausgeprägt föderalen Struktur und der ressortbezogenen Grenzen von Verantwortlichkeiten auf Bundesebene keine zentrale KI- oder ADM-Strategie gibt, findet eine global wettbewerbsfähige sektorspezifische Forschung statt.

Fallbeispiele für Algorithmische Entscheidungsfindung (ADM)

/ Diagnose und Behandlung von Krebserkrankungen

Momentan führt die Schweiz Untersuchungen zum Einsatz automatisierter Entscheidungsfindung in der Medizin durch, weshalb ADM im Gesundheitssektor bereits weiter entwickelt wurde als in anderen Bereichen. Derzeit sind mehr als 200 verschiedene Arten von Krebs bekannt, und



es gibt knapp 120 Medikamente, mit denen sie behandelt werden können. Jahr für Jahr werden unzählige Krebsdiagnosen gestellt, und da jeder Tumor sein eigenes spezifisches Profil mit bestimmten Genmutationen hat, die sein Wachstum fördern, tun sich für Onkolog·innen zahlreiche Probleme auf. Haben sie jedoch erst einmal eine Diagnose gestellt und die mögliche Genmutation definiert, müssen sie sich durch einen ständig wachsenden Berg medizinischer Forschungsliteratur arbeiten, um die effektivste Behandlungsmethode zu finden.

Aus diesem Grund sind die Spitäler der Universität Genf europaweit die ersten, in denen Watson for Genomics® zum Einsatz kommt. Das Tool von IBM Watson Health erleichtert das Auffinden therapeutischer Optionen und macht Vorschläge zur Behandlung von Krebspatient·innen. Nach wie vor untersuchen Ärzt·innen die Genmutationen und beschreiben, wo und in welcher Zahl sie auftreten, doch Watson for Genomics® kann diese Informationen nutzen, um eine Datenbank mit etwa drei Millionen Publikationen zu durchsuchen. Anschliessend erstellt das Programm einen Bericht, in dem die im Tumor des Patienten/der Patientin vorgefundenen genetischen Abweichungen klassifiziert werden, und schlägt geeignete Therapien und klinische Tests vor.

Bisher mussten Onkolog·innen diese Arbeit selbst machen – mit dem Risiko, eine eventuell mögliche Behandlungsmethode zu übersehen. Diese Recherchen übernimmt nun ein Computerprogramm. Allerdings müssen Onkolog·innen die Literaturliste, die das Programm erstellt, nach wie vor sorgfältig prüfen, bevor sie sich für eine Behandlungsmethode entscheiden. Watson for Genomics® spart also eine Menge Zeit bei der Analyse und bietet wichtige zusätzliche Informationen. In Genf fliesst der von diesem ADM-Tool erstellte Bericht in die Vorbereitungen zur Tumorkonferenz ein, bei der die Ärzt·innen die von Watson for Genomics® vorgeschlagenen Behandlungsmethoden zur Kenntnis nehmen und sie im Plenum diskutieren, um gemeinsam eine Behandlungsstrategie für jede/n einzelne/n Patient·in zu entwickeln (Schwerzmann/Arroyo 2019).

Im Universitätsspital Zürich kommt ebenfalls ADM zum Einsatz. Da sie sich speziell für Aufgaben eignet, die sich ständig wiederholen, vor allem in der Radiologie und Pathologie, wird sie genutzt, um die Brustdichte zu berechnen. Während der Mammografie analysiert ein Computeralgorithmus automatisch die Röntgenbilder und stuft das Brustgewebe in die Kategorie A, B, C oder D ein (ein international anerkanntes Raster für Risikoanalyse). Somit ist der Algorithmus Ärzt·innen eine grosse Hilfe bei der Beurteilung des Brustkrebsrisikos, denn die Brustdichte ist einer der wichtigsten Risikofaktoren bei Brustkrebs. Der Einsatz von ADM bei der Analyse medizinischer Bilder gehört inzwischen am Universitätsspital Zürich zum Standard. Auch die Forschung zu hochentwickelten Algorithmen für die Interpretation von Ultraschallbildern schreitet weiter voran (Lindner 2019).

Ausserdem werden bei Mammografie-Screenings mehr als ein Drittel der Brustkrebserkrankungen übersehen. Aus diesem Grund finden derzeit wissenschaftliche Untersuchungen statt, die klären sollen, wie ADM die Interpretation von Ultraschallbildern (US) unterstützen kann. Die Interpretation von US-Brustbildern steht in scharfem Kontrast zur standardmässig eingesetzten digitalen Mammografie, die im Wesentlichen beobachterabhängig ist und sehr gut ausgebildete, erfahrene Radiolog·innen braucht. Daher hat ein Spin-off des Unispitals Zürich erforscht, auf welche Weise ADM die Bildgebung mittels Ultraschall unterstützen und standardisieren kann. Dazu wurde anhand der Brustbilder, des Reportings und des Datensystems der menschliche Entscheidungsfindungsprozess simuliert. Die Technologie ist äusserst präzise, und so könnte dieser Algorithmus künftig zum Einsatz kommen, um menschliche Entscheidungsfindungsprozesse nachzubilden, und zum Standard bei der Entdeckung, Markierung und Klassifizierung von Brustläsionen mittels Ultraschall werden (Ciritisis a.o. 2019 S. 5458–5468).

/ Chatbot bei der Sozialversicherungsbehörde

Um die administrative Kommunikation zu vereinfachen und zu unterstützen, nutzen bestimmte Kantone sogenannte Chatbots. Insbesondere wurde 2018 ein Chatbot bei der „Sozialversicherungsanstalt des Kantons St. Gallens" (SVA St. Gallen) getestet. Die SVA St. Gallen ist ein Kompetenzzentrum für alle Arten von Sozialversicherung, einschliesslich Prämienverbilligung für Krankenversicherungen. Der Abschluss einer Krankenversicherung ist in der Schweiz obligatorisch. Sie schützt die Einwohner·innen im Falle von Krankheit, Schwangerschaft und Unfällen und bietet allen dasselbe Leistungsspektrum. Finanziert wird

sie über die Krankenversicherungsbeiträge (Prämien) der Bürger·innen. Die Prämien variieren je nach Versicherer und sind abhängig vom Wohnort einer Person und der Art der erforderlichen Versicherung. Sie richten sich nicht nach dem Einkommen. Dank Zuschüssen durch die Kantone (Prämienverbilligung) zahlen Bürgerinnen mit niedrigem Einkommen, Kinder und junge Erwachsene in Vollzeit- oder Berufsausbildung häufig reduzierte Beiträge. Die Entscheidung, wer zu einer Beitragsreduzierung berechtigt ist, fällt der jeweilige Kanton (FOPH 2020).

Am Ende eines jeden Jahres gehen bei der SVA St. Gallen ungefähr 80.000 Anträge auf Prämienverbilligung ein. Um die mit dieser Antragsflut verbundene Arbeitsbelastung zu reduzieren, testete die Behörde einen Chatbot via Facebook Messenger. Das Ziel dieses Pilotprojekts bestand darin, Kund·innen eine alternative Kommunikationsmethode anzubieten. Der erste digitale Verwaltungsassistent war so gestaltet, dass er Antragsteller·innen automatische Antworten auf die wichtigsten Fragen im Zusammenhang mit Prämienverbilligungen anbot. Zum Beispiel: Was ist eine Prämienverbilligung und wie kann sie beantragt werden? Kann ich eine Prämienverbilligung beantragen? Gibt es Sonderfälle, und wie sollte ich weiter vorgehen? Wie wird eine Prämienverbilligung berechnet und ausgezahlt? Darüber hinaus konnte der Chatbot, sofern eine entsprechende Eingabe erfolgte, Kund·innen auf weitere Dienstleistungen der SVA St. Gallen weiterleiten, unter anderem den Prämienverbilligungsrechner und das interaktive Registrierungsformular. Der Chatbot trifft zwar nicht die abschliessende Entscheidung über die Gewährung einer Prämienverbilligung, reduziert jedoch die Anzahl der Anfragen, da er nichtberechtigte Bürger·innen über die wahrscheinliche Ablehnung ihres Antrags informieren kann. Zudem spielt er eine wesentliche Rolle bei der Verbreitung von Informationen (Ringeisen/Bertolosi-Lehr/Demaj 2018 S. 51-65).

Aufgrund des positiven Feedbacks zu seinem ersten Testlauf integrierte die SVA St. Gallen den Chatbot 2019 in ihre Webseite. Ausserdem ist geplant, ihn schrittweise zu erweitern, um andere von der SVA St. Gallen abgedeckte Versicherungsprodukte einzubinden. Möglicherweise wird der Chatbot auch für Dienstleistungen im Zusammenhang mit Alters- und Hinterlassenenversicherung (AHV), Invalidenversicherung (IV) und Erwerbsausfallversicherung zum Einsatz kommen (IPV-Chatbot 2020).

/ Strafvollzugssystem

Das Schweizer Strafvollzugssystem basiert auf einem Stufenmodell. Entsprechend dieses Systems wird Häftlingen während der Verbüssung ihrer Gefängnisstrafe schrittweise mehr Freiheit gewährt. Dies macht das Ganze zu einem kollaborativen Prozess zwischen Strafvollzugsbehörden, Haftanstalten, Therapieanbieter·innen und Bewährungshelfer·innen. Natürlich sind das Fluchtrisiko und die Rückfallgefahr entscheidende Faktoren, wenn es darum geht, diese immer grösseren Freiheiten zu gewähren.

ROS TEILT DIE ARBEIT MIT DEN DELINQUENTEN IN VIER PROZESSSCHRITTE EIN: TRIAGE, ABKLÄRUNG, PLANUNG UND VERLAUF.

In den vergangenen Jahren wurde als Reaktion auf die Tatsache, dass verurteilte Schwerverbrecher diverse tragische Gewaltakte und Sexualdelikte verübten, die ROS (Risikoorientierte Sanktionierung) eingeführt. Das vorrangige Ziel der ROS besteht darin, durch einen einheitlichen, über verschiedene Vollzugsstufen und Vollzugseinrichtungen hinweg konsequent auf Rückfallprävention und Reintegration ausgerichteten Sanktionenvollzug Wiederholungstaten zu verhindern. ROS teilt die Arbeit mit den Delinquenten in vier Prozessschritte ein: Triage, Abklärung, Planung und Verlauf. Im Rahmen der Triagierung wird für jeden Einzelfall die Notwendigkeit von Risiko- und Bedarfsabklärung eingeschätzt. Basierend auf dieser Klassifizierung, wird während der Abklärungsphase eine differenzierte individuelle Fallanalyse durchgeführt. Während der Planungsphase werden diese Ergebnisse in eine individuelle Interventionsplanung (Fallführung) für den betreffenden Gefangenen überführt, die im weiteren Verlauf regelmässig überprüft wird (ROSNET 2020).

Am Beginn dieses Prozesses spielt die Triagierung eine ganz entscheidende Rolle – sowohl für den Straftäter/die Straftäterin selbst als auch im Hinblick auf ADM, denn diese Sichtung wird von einem ADM-Tool vorgenommen, dem sogenannten Fall-Screening-Tool (FaST). FaST stuft alle Fälle automatisch in die Kategorien A, B oder C ein. Falltyp A bedeutet, dass es keinen erhöhten Abklärungsbedarf gibt; Falltyp B entspricht einem allgemeinen Delinquenzrisiko; Falltyp C entspricht dem Risiko von Gewalt- und/oder Sexualdelikten.

Diese Klassifizierung wird unter Verwendung von Informationen aus dem Strafregisterauszug festgelegt und basiert auf allgemeinen statistischen Risikofaktoren wie Alter, verübten Gewaltdelikten bis zum 18. Lebensjahr, jugendanwaltschaftlichen Einträgen, Anzahl der Vorstrafen, Deliktkategorie, Strafmass, polymorphe Kriminalität, deliktfreie Zeit nach Entlassung und Delikte im Zusammenhang mit häuslicher Gewalt. Treffen Risikofaktoren zu, die wissenschaftlichen Erkenntnissen zufolge eine spezifische Verbindung mit Gewalt- oder Sexualdelikten haben, kommt Falltyp C zur Anwendung. Treffen Risikofaktoren zu, die eine spezifische Verbindung zu allgemeiner Delinquenz haben, wird Falltyp B zugeordnet. Werden keine oder nur vereinzelte Risikofaktoren gefunden, erfolgt eine Einstufung in Falltyp A. Die Zuordnung der einzelnen Merkmale (Risikofaktoren) erfolgt in Form von Entscheidungsfragen, deren Antworten unterschiedliche Gewichtungen haben (Punktwerte). Ist ein Risikofaktor gegeben, wird dessen Punktwert zur Punktwertsumme hinzugerechnet. Für das Endergebnis werden die gewichteten und bestätigten Posten zu einem Punktwert aufaddiert, der entweder zu einer Einstufung in Falltyp A, B oder C führt, die wiederum als Basis für die Entscheidung darüber dient, ob eine weitere Abklärung nötig ist (Prozessschritt 2).

Diese Einstufung wird vollautomatisch von der ADM-Anwendung vorgenommen. Allerdings ist zu betonen, dass es sich hierbei nicht um eine Risikoanalyse handelt, sondern vielmehr um eine Methode, mit deren Hilfe die Fälle mit erhöhtem Abklärungsbedarf herausgefiltert werden (Treuhardt/Kröger 2018 S. 24-32).

Ungeachtet dessen hat die Triagierung Auswirkungen darauf, wie Verantwortliche einer bestimmten Institution Entscheidungen treffen und welche Abklärungen vorgenommen werden. Hieraus leitet sich auch das sogenannte „Problemprofil" von Delinquenten ab, Ausgangspunkt für die Planung der Strafvollzugsmassnahmen/Fallführung (Prozessschritt 3). Insbesondere definiert diese Planung mögliche Formen der Vollzugslockerung wie etwa offener Vollzug, Freigang oder externe Unterbringung. Weiterhin ist zu sagen, dass sich in allen anderen Phasen des ROS offenkundig keine ADM-Anwendungen finden. FaST wird daher ausschliesslich während des Prozessschritts 1 (Triagierung) eingesetzt.

/ Predictive Policing

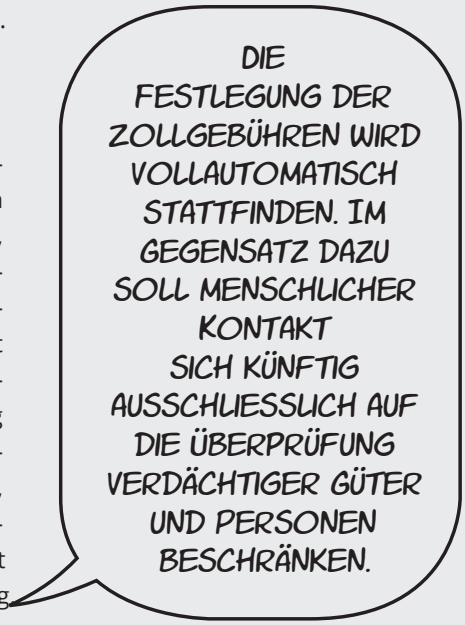
In einigen Kantonen, insbesondere in Basel-Landschaft, Aargau und Zürich, nutzt die Polizei Software zur Verhinderung von Straftaten. Dort vertraut man auf das kommerzielle Software-Paket „PRECOBS" (Pre-Crime Observation System), das ausschliesslich für die Prognose von Wohnungseinbrüchen verwendet wird. Diese relativ weit verbreitete Straftat ist wissenschaftlich sehr gut untersucht, und die Polizeibehörden verfügen in der Regel über eine solide Datenbasis in Bezug auf die räumliche und zeitliche Verteilung von Einbruchsdiebstählen sowie die speziellen Wesensmerkmale der Straftaten. Zudem deuten diese Delikte auf professionelles Vorgehen hin und weisen daher eine überdurchschnittliche Wahrscheinlichkeit für Folgedelikte auf. Ausserdem können mit Hilfe relativ weniger Datenpunkte entsprechende Prognosemodelle erstellt werden. PRECOBS basiert daher auf der Annahme, dass Einbrecher innerhalb kurzer Zeit mehrmals zuschlagen, wenn sie in einer bestimmten Gegend schon einmal erfolgreich waren.

Die Software wird genutzt, um in den Polizeiberichten zu Einbrüchen nach bestimmten Mustern zu suchen, wie etwa

Vorgehensweise der Täter und wann und wo sie zuschlagen. Anschliessend erstellt PRECOBS eine Prognose für Gebiete, in denen innerhalb der nächsten 72 Stunden ein erhöhtes Risiko für Wohnungseinbrüche besteht. Daraufhin schickt die Polizei gezielt Streifen in diese Bereiche. PRECOBS generiert also Prognosen auf der Basis zuvor eingespeister Entscheidungen und nutzt keine Methoden maschinellen Lernens. Es gibt zwar Pläne, PRECOBS künftig auch auf andere Delikte auszuweiten (wie zum Beispiel Auto- oder Taschendiebstahl) und folglich neue Funktionalitäten aufzubauen. Dennoch ist festzuhalten, dass die Nutzung von Predictive Policing in der Schweiz derzeit auf einen relativ kleinen und klar definierten Bereich der präventiven Polizeiarbeit beschränkt ist (Blur 2017, Leese 2018 S. 57-72).

/ Zollabfertigung

Auf Bundesebene dürfte ADM insbesondere bei der Eidgenössischen Zollverwaltung (EZV) Einzug halten, denn dieses Departement ist bereits hochautomatisiert. Die Beurteilung von Zollerklärungen erfolgt schon heute grösstenteils elektronisch. Der Beurteilungsvorgang lässt sich in vier Schritte unterteilen: summarisches Prüfverfahren, Annahme der Zollerklärung, Verifikation und Überprüfung, gefolgt von einer Bewertungsentscheidung.



Bei elektronischen Zollerklärungen findet das summarische Prüfverfahren in Form einer Plausibilitätsprüfung statt und wird direkt von dem eingesetzten System vorgenommen. Nach Abschluss der Plausibilitätsprüfung fügt das Datenverarbeitungssystem automatisch Datum und Uhrzeit der Annahme der elektronischen Zollerklärung hinzu, was bedeutet, dass die Zollerklärung akzeptiert wurde. Bis zu diesem Zeitpunkt läuft das Verfahren ohne jeglichen menschlichen Eingriff durch die Behörden ab.

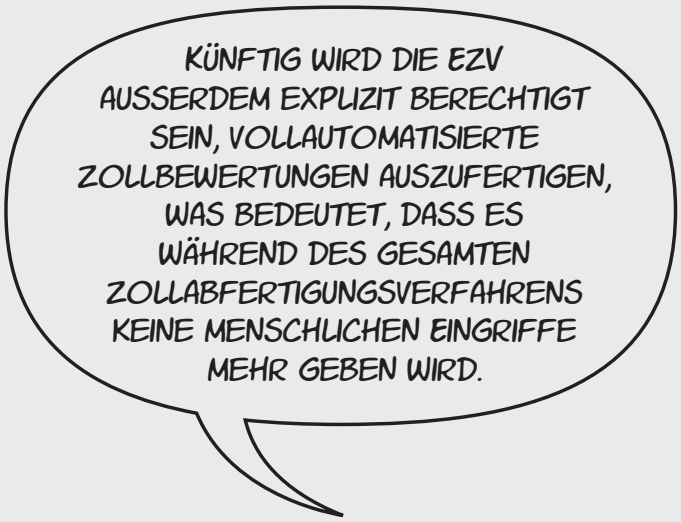
Das Zollbüro kann allerdings im Anschluss eine vollständige oder stichprobenartige Überprüfung und Verifizierung der deklarierten Güter vornehmen. Zu diesem Zweck nimmt das computerbasierte System auf Basis einer Risikoanalyse eine Auswahl vor. Die letzte Phase des Verfahrens bildet die Ausfertigung der Bewertungsentscheidung. Es ist nicht bekannt, ob diese Bewertungsentscheidung ebenfalls schon ohne jeglichen menschlichen Eingriff ausgefertigt werden kann. Allerdings wird das DaziT-Programm diese Ungewissheit ausräumen.

Das DaziT-Programm ist eine Massnahme des Bundes zur Digitalisierung sämtlicher Zollverfahren bis 2026, um Grenzübertritte zu vereinfachen und zu beschleunigen. Die für Kundenbeziehungen zuständigen Abteilungen der Grenzkontrollbehörden, die für die Bewegung von Gütern und Personen zuständig sind, werden grundlegend neu aufgestellt. Kund·innen, die sich korrekt verhalten, soll ermöglicht werden, ihre Formalitäten digital und unabhängig von Zeit und Ort abzuwickeln. Die exakte Implementierung des DaziT-Programms ist zwar noch in der Planungsphase, aber die Überarbeitung des DaziT betreffenden Zollgesetzes ist Teil der Revision des Bundesgesetzes über den Datenschutz (DSG).

Dies wird weiter unten detaillierter erläutert, und es sollte dazu führen, dass die oben genannte Ungewissheit in Bezug auf das automatische Zollbewertungsverfahren ausgeräumt wird: Künftig wird die EZV ausserdem explizit berechtigt sein, vollautomatisierte Zollbewertungen auszufertigen, was bedeutet, dass es während des gesamten Zollabfertigungsverfahrens keine menschlichen Eingriffe mehr geben wird. Damit wird die Entscheidung über die Festlegung der Zollgebühren vollautomatisch stattfinden. Im Gegensatz dazu soll menschlicher Kontakt sich künftig ausschliesslich auf die Überprüfung verdächtiger Güter und Personen beschränken. (EZV 2020).

/ Unfall- und Militärversicherung

Im Prozess der Überarbeitung des Datenschutzgesetzes (weiter unten im Detail erläutert) wurde entschieden, dass die Unfall- und Militärversicherungsunternehmen berechtigt sein werden, eine automatisierte Verarbeitung persönlicher Daten vorzunehmen. Es ist nicht klar, welche Tätigkeiten die Versicherungsunternehmen in Zukunft automatisieren werden. Allerdings könnten sie zum Beispiel Algorithmen nutzen, um die von Versicherungsnehmer·innen vorgelegten medizinischen Gutachten zu evaluieren. Mit Hilfe dieses vollautomatisierten Systems könnten Versicherungsprämien berechnet sowie Entscheidungen im



Zusammenhang mit der Geltendmachung von Ansprüchen getroffen und mit anderen Sozialleistungen abgestimmt werden. Es ist geplant, diese Körperschaften zu autorisieren, automatisierte Entscheidungen herauszugeben.

/ Automatische Fahrzeugerkennung

In den vergangenen Jahren macht sich angesichts der Nutzung automatisierter Systeme wie etwa Kameras, die Nummernschilder von Fahrzeugen einfangen, sie mit Hilfe optischer Buchstabenerkennung lesen und anschliessend mit einer Datenbank abgleichen, sowohl unter Politiker·innen als auch in der Öffentlichkeit zunehmend Sorge breit. Diese Technologie ist für verschiedene Zwecke einsetzbar, doch die Schweiz nutzt sie derzeit nur in beschränktem Umfang. Auf Bundesebene kommt das System zur automatischen Fahrzeugerkennung und Verkehrsüberwachung lediglich als taktisches Tool in Abhängigkeit von Lage- und Risikobeurteilungen, wirtschaftlichen Erwägungen und nur an Staatsgrenzen zum Einsatz (parlament.ch 2020). Der Halbkanton Basel-Landschaft hat eine gesetzliche Grundlage für die automatische Aufzeichnung von Nummernschildern und ihren Abgleich mit entsprechenden Datenbanken in Kraft gesetzt. (EJPD 2019).

/ Zuteilung von Primarschüler·innen

Ein anderer Algorithmus, der bisher aber noch nicht im Einsatz ist, wurde entwickelt, um Primarschüler·innen zuzuteilen. Internationale Studien deuten darauf hin, dass die soziale und ethnische Entmischung in städtischen Schulen zunimmt. Dies ist problematisch, da die soziale und ethnische Zusammensetzung der Schülerschaft ungeachtet ihres familiären Hintergrunds nachweisliche Auswirkungen auf die Leistungen der Kinder hat. In keinem anderen OECD-Land sind diese sogenannten „Struktureffekte“ so auffällig wie in der Schweiz. Die unterschiedliche Zusammensetzung von Schülerschaften ist im Wesentlichen der Segregation von Wohngebieten und den entsprechenden Schuleinzugsgebieten geschuldet. Das Zentrum für Demokratie Aarau hat deshalb vorgeschlagen, Schüler·innen nicht nur anhand ihrer sozialen und sprachlichen Herkunft zu mischen, sondern auch bei der Festlegung von Einzugsgebieten, so dass eine höchstmögliche Durchmischung der Schülerschaften erreicht werden kann. Um diesen Prozess zu optimieren, wurde ein völlig neuer, detaillierter Algorithmus entwickelt, der künftig zum Einsatz kommen könnte, um Schulzuteilung und Schulraumplanung zu unterstützen. Der Algorithmus wurde mit Hilfe der Zensusdaten von 1. bis 3.-Klässler·innen im Kanton Zürich so trainiert, dass er die Einzugsgebiete von Schulen rekonstruieren und die soziale Zusammensetzung einzelner Schulen untersuchen kann. Einbezogen wurden ebenfalls Daten zum Verkehrsaufkommen sowie dem vorhandenen Netzwerk von Trottoirs und Fusswegen sowie Unter- und Überführungen. Diese Daten könnten künftig genutzt werden, um zu berechnen, welche Schüler·innen welcher Schule zugeteilt werden müssen, um eine bessere Durchmischung der Klassen zu erreichen. Zugleich wird eine Überlastung der Kapazitäten von Schulgebäuden vermieden, während die für den Schulweg benötigte Zeit angemessen bleibt (ZDA 2019).

Politik, staatliche Aufsicht und öffentliche Debatte

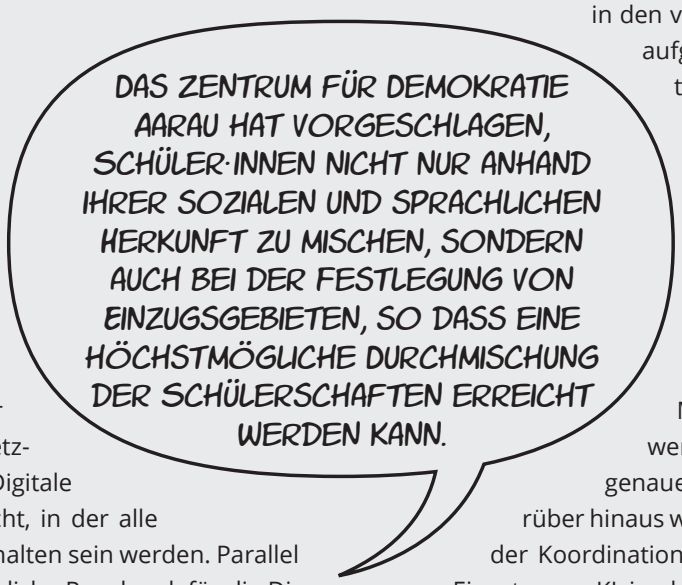
/ Die föderale Struktur der Schweiz als dominierender Aspekt

Bei Berichten über die Schweizer Politik ist vor allem die föderale Struktur des Landes zu betonen. Dieser Umstand wurde in den vorgenannten Beispielen für ADM bereits angesprochen. Die Schweiz ist ein Bundesstaat. Sie besteht aus 26 höchst autonomen Mitgliedsstaaten (Kantonen), die wiederum ihren Gemeinden weitgehende Gestaltungsspielräume gewähren. Infolgedessen hängt die politische und öffentliche Debatte um ADM grösstenteils von der jeweiligen Regierung ab, was im vorliegenden Bericht nicht erschöpfend beschrieben werden kann. Zudem birgt diese

Fragmentation in Politik, Regulierung und Forschung das Risiko, parallel an einander überlagernden Themen zu arbeiten, weshalb die Eidgenossenschaft auch, wie weiter unten beschrieben, eine stärkere Koordination anstrebt. Allerdings trägt die Bundesregierung die volle Verantwortung für bestimmte relevante Rechtsbereiche und politische Führungsprinzipien, welche für alle Schweizer Regierungen bindend sind und sich daher auch auf die gesamte Bevölkerung auswirken. Daher werden im Folgenden diese Bereiche der aktuellen politischen Debatte auf Bundesebene dargestellt.

/ Regierung

Derzeit wird die Rolle von ADM in der Gesellschaft, im Allgemeinen als KI bezeichnet, vor allem als Teil einer breiteren Diskussion um Digitalisierung behandelt. Es gibt zwar keine spezielle Strategie der Bundesregierung im Hinblick auf KI oder ADM, jedoch hat sie in den letzten Jahren eine „Strategie Digitale Schweiz" auf den Weg gebracht, in der alle Aspekte im Hinblick auf KI enthalten sein werden. Parallel dazu wird das nationale gesetzliche Regelwerk für die Digitalisierung durch die Revision des Bundesgesetzes zum Datenschutz (DSG) ganz generell angepasst.



/ Digitale Schweiz

Vor dem Hintergrund der zunehmenden Digitalisierung staatlicher Dienstleistungen rief die Eidgenossenschaft 2018 die Strategie „Digitale Schweiz" ins Leben. Unter anderem nimmt sie dabei die aktuellen Entwicklungen bei der KI in den Fokus (BAKOM 2020). Verantwortlich für die Strategie, insbesondere für Koordination und Implementierung, ist die „Interdepartementale Koordinationsgruppe Digitale Schweiz" mit ihrer Verwaltungseinheit „Geschäftsstelle Informationsgesellschaft Schweiz" (Digital Switzerland 2020).

Als Teil der Strategie „Digitale Schweiz" setzte der Bundesrat eine Arbeitsgruppe zum Thema „Künstliche Intelligenz" ein und beauftragte diese, einen Bericht über die Herausforderungen im Zusammenhang mit KI vorzulegen. Der Bericht wurde vom Bundesrat im Dezember 2019 zur Kenntnis genommen (SBFI 2020). Neben der Diskussion über die wesentlichsten Herausforderungen der KI – nämlich Nachvollziehbarkeit und systematische Fehler in Daten oder Algorithmen – listet der Bericht konkrete Handlungsbedarfe auf. Es wird anerkannt, dass alle Herausforderungen, einschliesslich dieser Handlungsbedarfe, stark von dem jeweils betrachteten Themenfeld abhängen, weshalb der Bericht 17 Themenfelder tiefgründiger untersuchte, etwa KI im Gesundheitswesen, in der Verwaltung und in der Justiz (SBDI 2020).

Im Prinzip, so der Bericht, seien die Herausforderungen der KI in der Schweiz bereits weitgehend erkannt und in den verschiedenen Politikbereichen aufgenommen worden. Ungeachtet dessen stellt der interdepartementale Bericht einen gewissen Handlungsbedarf fest, weshalb der Bundesrat vier Massnahmen beschlossen hat: Auf dem Gebiet des internationalen Rechts sowie in Bezug auf den Einsatz von KI bei der öffentlichen Meinungs- und Willensbildung werden zusätzliche Berichte zur genaueren Abklärung beauftragt. Darüber hinaus werden Wege zur Verbesserung der Koordination im Zusammenhang mit dem Einsatz von KI in der Bundesverwaltung geprüft. Insbesondere wird geprüft, ob ein Kompetenznetzwerk mit speziellem Fokus auf technische Aspekte der Anwendung von KI in der Bundesverwaltung geschaffen werden kann. Und schliesslich werden als wesentlicher Bestandteil der Strategie „Digitale Schweiz" Leitlinien für die KI-relevante Politik berücksichtigt werden. In diesem Zusammenhang hat der Bundesrat beschlossen, die interdepartementale Arbeit fortzuführen und bis zum Frühjahr 2020 strategische Leitlinien für die Eidgenossenschaft entwickeln zu lassen. (SBFI 2020)

Ausserdem hat der Bundesrat in seiner Sitzung am 13. Mai 2020 beschlossen, ein nationales Kompetenzzentrum für Datenwissenschaft zu gründen. Das Bundesamt für Statistik (BFS) wird dieses interdisziplinäre Zentrum per 1. Januar 2021 einrichten. Es wird die Bundesverwaltung darin unterstützen, Projekte im Bereich Datenwissenschaft umzusetzen. Zu diesem Zweck sollen der Wissenstransfer innerhalb der Bundesverwaltung sowie der Austausch mit wissenschaftlichen Kreisen, Forschungsinstituten und den für die praktische Anwendung zuständigen Stellen begünstigt werden. Das Kompetenzzentrum wird insbesondere

dazu beitragen, unter Berücksichtigung des Datenschutzes transparente Informationen zu produzieren. Die Argumentation für das neue Zentrum wird von einer Stellungnahme des Bundesrats gestützt, in der es heisst, die Datenwissenschaft werde zunehmend wichtiger, nicht zuletzt in der öffentlichen Verwaltung. Dem Bundesrat zufolge umfasst Datenwissenschaft „intelligente" Berechnungen (Algorithmen), mit denen bestimmte komplexe Aufgaben automatisiert werden können (Bundesrat 2020).

Überblick

Da das Bundesgesetz über den Datenschutz (DSG) aufgrund der rasanten technologischen Entwicklungen nicht mehr aktuell ist, beabsichtigt der Bundesrat, es diesen veränderten technologischen und gesellschaftlichen Bedingungen anzupassen und dabei insbesondere die Transparenz der Datenverarbeitung zu verbessern sowie das informationelle Selbstbestimmungsrecht betroffener Personen zu stärken. Gleichzeitig sollte diese Totalrevision die Schweiz in die Lage versetzen, die überarbeitete Datenschutzkonvention des Europarates ETS 108 zu ratifizieren und die Richtlinie (EU) 2016/680 des Europäischen Parlamentes und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten zum Zwecke der Verhütung, Ermittlung, Aufdeckung oder Verfolgung von Straftaten oder der Strafvollstreckung zu übernehmen, wozu sie nach dem Schengen-Abkommen verpflichtet ist. Darüber hinaus dürfte die Revision die Schweizer Datenschutzgesetzgebung insgesamt näher an die Anforderungen der Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung) heranrücken. Die Revision wird derzeit im Parlament diskutiert (EJPD 2020).

Während die Totalrevision der existierenden Bundesdatenschutzgesetzgebung nachweislich in eine Überarbeitung des gesamten Gesetzes mit all seinen diversen Aspekten münden wird, ist im Hinblick auf ADM vor allem eine neue Regelung von besonderem Interesse. Im Falle sogenannter „automatisierter Einzelentscheidungen" soll es eine Verpflichtung geben, die betroffene Person zu informieren, „wenn diese [Entscheidung] für sie mit einer Rechtsfolge verbunden ist oder sie erheblich beeinträchtigt". Die betroffene Person kann zudem verlangen, dass die Entscheidung von einer natürlichen Person überprüft wird oder dass ihr die Logik mitgeteilt wird, auf der die Entscheidung beruht. Damit wird ein differenziertes Regelwerk für Entscheidungen durch bundesstaatliche Stellen gegeben sein. Entsprechend kann, selbst wenn die bundesstaatlichen Stellen die automatisierte Einzelentscheidung als solche kennzeichnen müssen, die Möglichkeit einer betroffenen Person, eine Überprüfung durch einen menschlichen Sachbearbeiter zu verlangen, durch andere Bundesgesetze eingeschränkt sein. Im Gegensatz zur DSGVO der EU gibt es weder ein Verbot bestimmter automatisierter Entscheidungen noch besteht ein Anspruch darauf, nicht Gegenstand einer solchen Entscheidung zu sein. (SBFI, Staatssekretariat für Bildung, Forschung und Innovation (2019): Herausforderungen der künstlichen Intelligenz – Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, in: *admin.ch*, [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].)

/ Zivilgesellschaft, akademischer Bereich und sonstige Organisationen

Darüber hinaus forschen, diskutieren und arbeiten eine Anzahl von Foren in der Schweiz zum Thema digitale Transformation und deren Chancen, Herausforderungen, Anforderungen und ethischen Aspekten. Die meisten von ihnen widmen sich dem Thema auf allgemeiner Ebene, wobei einige speziell ADM oder KI ansprechen.

/ Forschungsinstitute

Die Schweiz verfügt über eine ganze Reihe bekannter, seit langem existierender Forschungszentren zur KI-Technologie. Dazu gehören das Dalle-Molle-Forschungsinstitut für Künstliche Intelligenz (IDSIA) in Lugano (SUPSI 2020) und das IDIAP Research Institute in Martigny (Idiap 2020) sowie die Forschungszentren der Eidgenössischen Technischen Hochschule in Lausanne (EPFL) (EPFL 2020) und Zürich (ETH) (ETH 2020). Ergänzt werden die akademischen Initiativen von privaten Initiativen wie etwa der Swiss Group for Artificial Intelligence and Cognitive Science (SGAICO), indem sie Forscher·innen und Nutzer·innen zusammenbringen sowie Wissenstransfer, Vertrauensbildung und Interdisziplinarität befördern. (SGAICO 2020)

/ Staatliche Forschungsförderung

Die Eidgenossenschaft geht das Thema KI auch durch gezielte finanzielle Unterstützung an. So investiert zum Bei-

METHODEN DES MASCHINELLEN LERNENS KOMMEN, SOWEIT ZU SEHEN IST, BEI STAATLICHEN AKTIVITÄTEN IM ENGEREN SINNE WIE ETWA DER POLIZEIARBEIT ODER IM STRAFVERFOLGUNGS- UND -VOLLZUGSSYSTEM NICHT ZUM EINSATZ.

spiel die Bundesregierung über die Swiss National Science Foundation (SNSF) in zwei nationale Forschungsprogramme (SNF 2020): das Nationale Forschungsprogramm 77 „Digitale Transformation" (NRP 77) (NRP77 2020) und das Nationale Forschungsprogramm 75 „Big Data" (NRP 75) (NRP 75 2020). Ersteres untersucht die Wechselbeziehungen und konkreten Auswirkungen der digitalen Transformation in der Schweiz und konzentriert sich dabei auf Bildung und Lernen, ethische Fragen, Vertrauenswürdigkeit, Governance, die Wirtschaft und den Arbeitsmarkt (NFP 77 2020). Letzteres hat zum Ziel, die wissenschaftliche Basis für eine effektive und angemessene Nutzung grosser Datenmengen zu schaffen. Dementsprechend untersuchen die beiden Forschungsprojekte Fragen im Zusammenhang mit den gesellschaftlichen Auswirkungen der Informationstechnologie und befassen sich mit konkreten Anwendungen (SNF 2020).

Ein weiteres Institut, das auf diesem Gebiet aktiv ist, ist die Stiftung für Technologiefolgen-Abschätzung (TA-Swiss). Der Auftrag des Kompetenzzentrums der Akademien der Wissenschaften Schweiz ist im Bundesgesetz über die Forschung festgeschrieben. Das staatlich finanzierte Beratungsgremium hat diverse Studien zur KI beauftragt. Die wichtigste wurde am 15. April 2020 veröffentlicht und befasst sich mit dem Einsatz von KI in verschiedenen Bereichen (Konsum, Arbeitswelt, Bildung, Forschung, Medien, öffentliche Verwaltung und Gerichtsbarkeit). Sie kommt zu der Einschätzung, dass ein eigenständiges Gesetz zur Nutzung von KI nicht wirksam wäre. Dennoch sollten Bürger·innen, Konsument·innen und Arbeitnehmer·innen bei all ihren Interaktionen mit dem Staat, Unternehmen oder Arbeitgeber·innen so transparent wie möglich über den Einsatz von KI informiert werden. Nutzen staatliche Institutionen oder Unternehmen KI, dann sollten sie es auf der Grundlage klarer Regelungen sowie in einer verständlichen und transparenten Art und Weise tun. (Christen, M. et al. 2020).

/ Digital Society Initiative

Die „UZH Digital Society Initiative" wurde 2016 gestartet. Das an der Universität Zürich angesiedelte Kompetenzzentrum widmet sich einer kritischen Beschäftigung mit allen Aspekten der digitalen Gesellschaft. Ihr Ziel ist die kritische Begleitung und Mitgestaltung der Digitalisierung in Gesellschaft, Demokratie, Wissenschaft, Kommunikation und Wirtschaft. Darüber hinaus will sie auf eine zukunftsorientierte Art und Weise das aktuelle, durch die Digitalisierung ausgelöste Umdenken kritisch begleiten und mitgestalten und die Universität Zürich sowohl national wie auch international als Kompetenzzentrum für die kritische Reflexion über alle Aspekte der digitalisierten Gesellschaft positionieren (UZH 2020).

/ Digitale Gesellschaft

Die Digitale Gesellschaft ist ein gemeinnütziger und breit abgestützter Verein für Bürger- und Konsumentenschutz im digitalen Zeitalter. Die zivilgesellschaftliche Organisation arbeitet seit 2011 für eine nachhaltige, demokratische und freie Öffentlichkeit und hat das Ziel, Grundrechte in einer

digital vernetzten Welt zu verteidigen. Digitale Gesellschaft (o. J.): Über uns, in: Digitale Gesellschaft, [online] https://www.digitale-gesellschaft.ch/uber-uns/ [30.01.2020].)

/ Sonstige Organisationen

Auch einige andere Schweizer Organisationen sollten Erwähnung finden. Sie konzentrieren sich auf die Digitalisierung im Allgemeinen, insbesondere in einem ökonomischen Kontext, z. B. der Schweizerische Verband der Telekommunikation (asut 2020), digitalswitzerland (Castle 2020), die Swiss Data Alliance und Swiss Fintech Innovations (SFTI).

Wichtigste Resultate

ADM kommt in der Schweiz in diversen Teilbereichen des öffentlichen Sektors zum Einsatz, tendenziell allerdings nicht in einer zentralisierten oder umfassenden Art und Weise. So nutzen zum Beispiel nur einige wenige Kantone ADM in der Polizeiarbeit, und die verwendeten Systeme sind verschieden. Der Vorteil einer solchen Herangehensweise liegt darin, dass die Kantone oder der Bund von den Erfahrungen anderer Kantone profitieren können. Der Nachteil sind eventuelle Effizienzverluste.

Punktuell existieren gesetzliche Grundlagen, jedoch gibt es kein einheitliches ADM-Gesetz oder e-Government-Gesetz oder Ähnliches. Es gibt auch keine spezielle KI- oder ADM-Strategie, wobei in jüngerer Zeit das Augenmerk auf eine bessere Koordination sowohl zwischen den Departementen auf Kantonsebene als auch zwischen Bund und Kantonen gelegt wird. Methoden des maschinellen Lernens kommen, soweit zu sehen ist, bei staatlichen Aktivitäten im engeren Sinne wie etwa der Polizeiarbeit oder im Strafverfolgungs- und -vollzugssystem nicht zum Einsatz.

Auf dieser Ebene wird ebenfalls über ADM diskutiert oder ADM in ausgewählten Bereichen eingesetzt, allerdings nicht vollumfänglich. Im breiteren öffentlichen Sektor kommt ADM häufiger und flächendeckender zum Einsatz. Ein gutes Beispiel ist ihre Nutzung im Schweizer Gesundheitswesen. Das Universitätsklinikum Genf ist das erste Krankenhaus Europas, in dem es zum Einsatz kommt, um Behandlungen für Krebspatient·innen vorzuschlagen.

Quellen:

Asut (o. J.): in: asut.ch, [online] https://asut.ch/asut/de/page/index.xhtml [30.01.2020]

Bundesamt für Kommunikation BAKOM (o. J.): Digitale Schweiz, in: admin.ch, [online] https://www.bakom.admin.ch/bakom/de/home/digital-und-internet/strategie-digitale-schweiz.html [30.01.2020].

Der Bundesrat (o.J.): Der Bundesrat schafft ein Kompetenzzentrum für Datenwissenschaft, In: admin.ch, [online] https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-79101.html [15.05.2020].)

Christen, M. et al. (2020): Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz, in: TA-Swiss, [online] https://www.ta-swiss.ch/themen-projekte-publikationen/informationsgesellschaft/kuenstliche-intelligenz/ [15.05.2020].

Ciritsis, Alexander / Cristina Rossi / Matthias Eberhard / Magda Marcon / Anton S. Becker / Andreas Boss (2019): Automatic classification of ultrasound breast lesions using a deep convolutional neural network mimicking human decision-making, in: European Radiology, Jg. 29, Nr. 10, S. 5458–5468, doi: 10.1007/s00330-019-06118-7.

digitalswitzerland (Castle, Danièle Digitalswitzerland (2019): Digitalswitzerland - Making Switzerland a Leading Digital Innovation Hub, in: digitalswitzerland, [online] https://digitalswitzerland.com [30.01.2020])

Digital Switzerland (2020): (Ofcom, Federal Office Of Communications (o. J.): Digital Switzerland Business Office, in: admin.ch, [online] https://www.bakom.admin.ch/bakom/en/homepage/ofcom/organisation/organisation-chart/information-society-business-office.html [30.01.2020].)

EPFL (o. J.): in: epfl, [online] https://www.epfl.ch/en/ [30.01.2020]

EJPD (o. J.): Stärkung des Datenschutzes, in: admin.ch, [online] https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/datenschutzstaerkung.html [30.01.2020c].

E-ID Referendum (o.J.): in: e-id-referendum.ch/, [online] https://www.e-id-referendum.ch [31.1.2020].

EJPD (o. J.): Stärkung des Datenschutzes, in: admin.ch, [online] https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/datenschutzstaerkung.html [30.01.2020c].

EJPD Eidgenössisches Justiz- und Polizeidepartement (2019): Änderung der Geschwindigkeitsmessmittel-Verordnung (SR 941.261) Automatische Erkennung von Kontrollschildern, in: admin.ch, [online] https://www.admin.ch/ch/d/gg/pc/documents/3059/Erl_Bericht_de.

EZV (2020): EZV, Eidgenössische Zollverwaltung (o. J.): Transformationsprogramm DaziT, in: admin.ch, [online] https://www.ezv.admin.ch/ezv/de/home/themen/projekte/dazit.html [30.01.2020].

ETH Zurich - Homepage (o. J.): in: ETH Zurich - Homepage | ETH Zurich, [online] https://ethz.ch/en.html [30.01.2020].

Federal office of public health FOPH (2020): (Health insurance: The Essentials in Brief (o. J.): in: admin.ch, [online] https://www.bag.admin.ch/bag/en/home/versicherungen/krankenversicherung/krankenversicherung-das-wichtigste-in-kuerze.html [13.02.2020].)

Geschäft Ansehen (o. J.): in: *parlament.ch*, [online] https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20143747 [30.01.2020].

Heinhold, Florian (2019): Hoffnung für Patienten?: Künstliche Intelligenz in der Medizin, in: *br.ch*, [online] https://www.br.de/br-fernsehen/sendungen/gesundheit/kuenstliche-intelligenz-ki-medizin-102.html [30.01.2020].

Idiap Research Institute (o. J.): in: *Idiap Research Institute, Artificial Intelligence for Society,* [online] https://www.idiap.ch/en [30.01.2020]

Der IPV-Chatbot – SVA St.Gallen (o. J.): in: svasg.ch, [online] https://www.svasg.ch/news/meldungen/ipv-chatbot.php [30.01.2020].

Leese, Matthias (2018): Predictive Policing in der Schweiz: Chancen, Herausforderungen Risiken, in: Bulletin zur Schweizerischen Sicherheitspolitik, Jg. 2018, S. 57–72.

Lindner, Martin (2019): KI in der Medizin: Hilfe bei einfachen und repetitiven Aufgaben, in: Neue Zürcher Zeitung, [online] https://www.nzz.ch/wissenschaft/ki-in-der-medizin-hilfe-bei-einfachen-und-repetitiven-aufgaben-ld.1497525?reduced=true [30.01.2020

Medinside (o. J.): in: *Medinside,* [online] https://www.medinside.ch/de/post/in-genf-schlaegt-der-computer-die-krebsbehandlung-vor [14.02.2020].

NRP 75 Big Data (o. J.): in: SNF, [online] http://www.snf.ch/en/researchinFocus/nrp/nfp-75/Pages/default.aspx [30.01.2020].

NFP [Nr.] (o. J.): in: nfp77.ch, [online] http://www.nfp77.ch/en/Pages/Home.aspx [30.01.2020]

NRP 75 Big Data (o. J.): in: *SNF,* [online] http://www.snf.ch/en/researchinFocus/nrp/nfp-75/Pages/default.aspx [30.01.2020].

NFP [Nr.] (o. J.): in: nfp77.ch, [online] http://www.nfp77.ch/en/Pages/Home.aspx [30.01.2020]

Ringeisen, Peter / Andrea Bertolosi-Lehr / Labinot Demaj (2018): Automatisierung und Digitalisierung in der öffentlichen Verwaltung: digitale Verwaltungsassistenten als neue Schnittstelle zwischen Bevölkerung und Gemeinwesen, in: Yearbook of Swiss Administrative Sciences, Jg. 9, Nr. 1, S. 51–65, doi: 10.5334/ssas.123.

ROSNET > ROS allgemein (o. J.): in: *ROSNET,* [online] https://www.rosnet.ch/de-ch/ros-allgemein [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (o. J.): Künstliche Intelligenz, in: *admin.ch*, [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (2019): Herausforderungen der künstlichen Intelligenz - Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, in: *admin.ch*, [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (o. J.): Künstliche Intelligenz, in: *admin.ch*, [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (2019): Herausforderungen der künstlichen Intelligenz - Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, in: *admin.ch*, [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

Schaffhauser eID+ - Kanton Schaffhausen (o. J.): in: sh.ch, [online] https://sh.ch/CMS/Webseite/Kanton-Schaffhausen/Beh-rde/Services/Schaffhauser-eID--2077281-DE.html [30.01.2020].

SGAICO - Swiss Group for Artificial Intelligence and Cognitive Science (2017): in: *SI Hauptseite,* [online] https://swissinformatics.org/de/gruppierungen/fg/sgaico/ [30.01.2020]

SNF, [online] http://www.snf.ch/en/Pages/default.aspx [30.01.2020]

Srf/Blur;Hesa (2017): Wie «Precobs» funktioniert - Die wichtigsten Fragen zur «Software gegen Einbrecher», in: Schweizer Radio und Fernsehen (SRF), [online] https://www.srf.ch/news/schweiz/wie-precobs-funktioniert-die-wichtigsten-fragen-zur-software-gegen-einbrecher

SUPSI - Dalle Molle Institute for Artificial Intelligence - Homepage (o. J.): in: *idsia,* [online] http://www.idsia.ch [30.01.2020].

Swissdataaliance (o. J.): in: swissdataalliance, [online] https://www.swissdataalliance.ch [30.01.2020].

Swiss Fintech Innovations (SFTI introduces Swiss API information platform (2019): in: *Swiss Fintech Innovations – Future of Financial Services,* [online] https://swissfintechinnovations.ch [30.01.2020]).

Schwerzmann, Jacqueline Amanda Arroyo (2019): Dr. Supercomputer – Mit künstlicher Intelligenz gegen den Krebs, in: Schweizer Radio und Fernsehen (SRF), [online] https://www.srf.ch/news/schweiz/dr-supercomputer-mit-kuenstlicher-intelligenz-gegen-den-krebs

Treuthardt, Daniel / Melanie Kröger (2019): Der Risikoorientierte Sanktionenvollzug (ROS) – empirische Überprüfung des Fall-Screening-Tools (FaST), in: Schweizerische Zeitschrift für Kriminologie, Jg. 2019, Nr. 1–2, S. 76–85.); (Treuthardt, Daniel / Melanie Kröger / Mirjam Loewe-Baur (2018): Der Risikoorientierte Sanktionenvollzug (ROS) – aktuelle Entwicklungen, in: Schweizerische Zeitschrift für Kriminologie, Jg. 2018, Nr. 2, S. 24–32.

ZDA (2019): Durchmischung in städtischen Schulen, in: *zdaarau.ch,* [online] https://www.zdaarau.ch/dokumente/SB-17-Durchmischung-Schulen-ZDA.pdf [30.01.2020].

Team

/ Beate Autering

Gestaltung und Layout

Beate Autering ist freiberufliche Diplomdesignerin und ist Mitbetreiberin der Ateliergemeinschaft beworx. Neben der Erstellung von Designs, Grafiken und Illustrationen bietet sie Bildbearbeitung und Postproduktion an.

/ Nadja Braun Binder

Co-Autorin des **Forschungskapitels**

Nadja Braun Binder studierte Rechtwissenschaften an der Universität Bremen, wo sie auch promovierte. Ihre akademische Karriere führte sie 2011 an das Deutsche Forschungsinstitut für öffentliche Verwaltung in Speyer, wo sie unter anderem zur Automatisierung von Verwaltungsabläufen forschte. 2017 folgte sie nach ihrer Habilitation an der Deutschen Universität für Verwaltungswissenschaften in Speyer einem Ruf an die rechtswissenschaftliche Fakultät der Universität Zürich. Dort war sie bis 2019 als Assistenzprofessorin tätig. Seit 2019 ist sie Professorin für Öffentliches Recht an der Universität Basel. Im Mittelpunkt ihrer Forschungsarbeit stehen Rechtsfragen im Zusammenhang mit der Digitalisierung von Regierungsbehörden und Verwaltung. Aktuell führt sie eine Studie zum Einsatz künstlicher Intelligenz in der öffentlichen Verwaltung des Kantons Zürich durch.

/ Fabio Chiusi

Projektmanager und Mitherausgeber des **Reports** sowie **Autor** der Einleitung und des **Europa-Kapitels**

Foto: Julia Bornkessel

Fabio ist Projektmanager des Automating Society Reports 2020. Nach zehn Jahren im Bereich Tech Reporting arbeitete er als Berater und wissenschaftlicher Mitarbeiter in den Bereichen Daten und Politik (Tactical Tech) und KI im Journalismus (Polis LSE). Er koordinierte den Bericht „Persuasori Social“ über die Regulierung politischer Kampagnen in den sozialen Medien für das Projekt PuntoZero und arbeitete während der laufenden Legislaturperiode als technikpolitischer Mitarbeiter in der Abgeordnetenkammer des italienischen Parlaments. Als Fellow am Nexa Center for Internet & Society in Turin ist er Lehrbeauftragter an der Universität San Marino und unterrichtet die Fächer „Journalismus und Neue Medien” sowie “Verlagswesen und Digitale Medien”. Er ist Autor mehrerer Essays über Technologie und Gesellschaft, zuletzt ‘Io non sono qui. Visioni e inquietudini da un futuro presente’ (DeA Planeta, 2018), das derzeit ins Polnische und Chinesische übersetzt wird. Er schreibt als Tech-Policy-Reporter beim Kollektiv-Blog ValigiaBlu.

/ Samuel Daveti

Co-Autor des **Comics**

Samuel Daveti ist ein Gründungsmitglied des der Kulturvereinigung „Double Shot“. Er ist der Autor des französischsprachigen Graphic Novels Akron Le guerrier (Soleil, 2009) und er ist Kurator der Anthologie Fascia Protetta (Double Shot, 2009). Im Jahr 2011 wurde er Gründungsmitglied des Comic-Kollektivs Mammaiuto. Samuel schrieb auch Un Lungo Cammino (Mammaiuto, 2014; Shockdom, 2017), als Film für das Medienunternehmen Brandon Box veröffentlicht wird. 2018 schrieb er The Three Dogs, mit Zeichnungen von Laura Camelli, das den Micheluzzi-Preis auf der Napoli Comicon 2018 und den Boscarato Preis für den besten Webcomic auf dem Treviso Comic Book Festival.

/ Catherine Egli

Co-Autorin des Forschungskapitels

Catherine Egli hat kürzlich ihren zweisprachigen Doppelabschluss (Double Master's Degree) in Rechtswissenschaften der Universitäten BASEL und Genf erhalten. Im Mittelpunkt ihrer Masterarbeit stand das Thema automatisierte individuelle Entscheidungsfindung und Regelungsbedarf im Schweizer Verwaltungsverfahrensgesetz. Neben ihrem Studium führte sie für den Lehrstuhl von Prof. Dr. Nadja Braun Binder Recherchen zu juristischen Fragen im Zusammenhang mit automatisierter Entscheidungsfindung durch. Ihre bevorzugten Forschungsgebiete liegen im Bereich Gewaltentrennung, Digitalisierung der öffentlichen Verwaltung und digitale Demokratie.

/ Sarah Fischer

Mitherausgeberin

Sarah Fischer ist als Projektmanagerin für das Projekt „Ethik der Algorithmen" der Bertelsmann Stiftung tätig, wo sie vor allem für wissenschaftliche Studien verantwortlich ist. Zuvor arbeitete sie als Postdoctoral Fellow am Graduiertenkolleg „Vertrauen und Kommunikation in einer digitalisierten Welt" an der Universität Münster, wo sie sich insbesondere mit dem Thema „Vertrauen in Suchmaschinen" beschäftigte. Ebenfalls im Rahmen dieses Graduiertenkollegs erwarb sie mit einer Promotion über Vertrauen in Gesundheitsdienstleistungen im Internet ihren Doktortitel. Sie studierte Kommunikationswissenschaften an der Friedrich-Schiller-Universität Jena und ist Co-Autorin der Aufsätze „Wo Maschinen irren können. Fehlerquellen und Verantwortlichkeiten in Prozessen algorithmischer Entscheidungsfindung" und „Was Deutschland über Algorithmen weiss und denkt".

/ Leonard Haas

Redaktionsassistent

Leonard Haas arbeitet als wissenschaftlicher Mitarbeiter bei AlgorithmWatch und war unter anderem für die Konzeption, Weiterentwicklung und Pflege des Projekts AI Ethics Guidelines Global Inventory mitverantwortlich. Er studiert im Master Sozialwissenschaften an der Humboldt Universität Berlin und hält zwei Bachelorabschlüsse von der Universität Leipzig in Digital Humanities und Politikwissenschaften. Sein Forschungsschwerpunkt bildet die Automatisierung der Arbeit und des Regierens. Ausserdem interessiert er sich für Allgemeinwohl-orientierte Datenpolitik und Arbeitskämpfe in der Tech-Branche.

/ Graham Holliday

Lektor der englischsprachigen Gesamtausgabe

Photo: Josh White

Graham Holliday ist freiberuflicher Redakteur, Autor und Journalismus-Trainer. Er war fast zwanzig Jahre lang in diversen Positionen für die BBC tätig und außerdem Korrespondent für Reuters in Ruanda. Derzeit ist er Redakteur für die CNN-Sendungen „Parts Unknown" und „Roads & Kingdoms". Die ersten beiden von ihm verfassten Bücher wurden von dem inzwischen verstorbenen Anthony Bourdain herausgebracht und erhielten unter anderem Rezensionen in der New York Times, der Los Angeles Times, dem Wall Street Journal, Publisher's Weekly, dem Library Journal und auf National Public Radio.

/ Nikolas Kayser-Bril

Mitherausgeber und Autor der Story

Photo: Julia Bornkessel

Nicolas ist Datenjournalist und arbeitet für AlgorithmWatch als Reporter. Er war Wegbereiter für neue Formen des Journalismus in Frankreich und Europa und ist einer der führenden Experten für Datenjournalismus. Er hält regelmässig Vorträge auf internationalen Konferenzen, unterrichtet Journalismus an französischen Journalismusschulen und gibt Schulungen in Redaktionen. Als autodidaktischer Journalist und Entwickler (und Absolvent der Wirtschaftswissenschaften) entwickelte er 2009 zunächst kleine interaktive, datengesteuerte Anwendungen für Le Monde in Paris. Anschliessend baute er 2010 das Datenjournalismus-Team bei OWNI auf, bevor er von 2011 bis 2017 Journalism++ mitbegründete und leitete. Nicolas ist auch einer der Hauptverfasser des Datajournalism Handbook, dem Nachschlagewerk für Datenjournalismus.

/ Anna Mätzener

Mitherausgeberin

Anna Mätzener ist Leiterin von AlgorithmWatch Schweiz. Sie ist promovierte Mathematikerin mit den Nebenfächern Philosophie und italienische Sprachwissenschaften und hat an der Universität Zürich studiert. Vor ihrer Tätigkeit bei AlgorithmWatch Schweiz war sie Programmplanerin für Mathematik und Wissenschaftsgeschichte in einem internationalen Wissenschaftsverlag und zuletzt Mathematik-Lehrerin an einem Gymnasium in Zürich.

/ Lorenzo Palloni

Co-Autor des Comics

Lorenzo Palloni, mehrfach ausgezeichneter Cartoonist, ist Autor mehrerer Graphic Novels und Webcomics und Mitbegründer des Comic-Künstlerkollektivs Mammaiuto. Derzeit arbeitet er an diversen Titeln für den französischen und italienischen Buchmarkt. Er ist ausserdem Lehrer für Scriptwriting und Storytelling an der Scuola Internazionale di Comics di Reggio Emilia (Internationale Comic-Schule in Reggio Emiliana).

/ Kristina Penner

Co-Autorin des Europa-Kapitels

Photo: Julia Bornkessel

Kristina Penner ist Referentin der Geschäftsführung bei AlgorithmWatch.Ihre Forschungsinteressen umfassen ADM in Sozialsystemen, Social Scoring und die gesellschaftlichen Auswirkungen von ADM, sowie die Nachhaltigkeit neuer Technologien aus einer ganzheitlichen Perspektive. Ihre Analyse des EU-Grenzverwaltungssystems baut auf ihrer bisherigen Erfahrung in der Forschung und Beratung zum Asylrecht auf. Ihre bisherigen Erfahrungen umfassen Projekte zu konfliktsensitivem Journalismus und zur Nutzung von Medien in der Zivilgesellschaft und sowie zur Beteiligung von Interessengruppen an Friedensprozessen auf den Philippinen. Sie hat einen Master-Abschluss in International Studies / Peace and Conflict Research von der Goethe-Universität in Frankfurt.

/ Alessio Ravazzani

Co-Autor des Comics

Alessio Ravazzani ist redaktioneller Grafikdesigner, Cartoonist und Illustrator. Er arbeitet für die renommiertesten Comic- und Graphic-Novel-Verlage Italiens und gehört zu den Autoren und Gründungsmitgliedern des Kollektivs Mammaiuto.

/ Friederike Reinhold

Mitarbeit an der Einleitung, einschliesslich der Handlungsempfehlungen

Friederike Reinhold ist als Senior Policy Advisor von AlgorithmWatch verantwortlich für die Weiterentwicklung des Policy- und Advocacy-Bereichs. Vor ihrer Tätigkeit bei AlgorithmWatch arbeitete Friederike als Referentin im Auswärtigen Amt, für den Norwegian Refugee Council (NRC) im Iran, für die Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) in Afghanistan sowie am Wissenschaftszentrum Berlin für Sozialforschung (WZB).Afghanistan, and at the WZB Berlin Social Science Center.

/ Matthias Spielkamp

Mitherausgeber

Foto: Julia Bornkessel

Matthias Spielkamp ist Mitgründer und Geschäftsführer der Organisation AlgorithmWatch, die mit der Theodor-Heuss-Medaille ausgezeichnet und für einen Grimme Online Award nominiert wurde. Er war Sachverständiger in Anhörungen des Europarats, der EU-Parlaments, des Bundestags und ist Mitglied der Global Partnership on Artificial Intelligence (GPAI). Matthias ist Vorstandsmitglied bei Reporter ohne Grenzen, Mitglied des Kuratoriums der Stiftung Warentest und im Beirat des Whistleblower Netzwerks. Er ist Autor und Herausgeber von Büchern zu Algorithmen, KI und Automatisierung, Internet Governance, der Zukunft des Journalismus und des Urheberrechts. Seine journalistischen Beiträge sind in MIT Technology Review, Die Zeit, brand eins und vielen anderen Publikationen erschienen. MA Philosophie, FU Berlin, MA Journalismus, University of Colorado at Boulder.

/ Beate Stangl

Layout

Beate Stangl arbeitet als Diplomdesignerin in Berlin und gestaltet mit Schwerpunkt Editorial Design u.a. für beworx, die Friedrich-Ebert-Stiftung, Buske Verlag, UNESCO Welterbe Deutschland e.V., Agentur Sehstern, iRights Lab, Landesspracheninstitut Bochum.

/ Marc Thümmler

Koordination

Foto: Julia Bornkessel

Marc Thümmler is in charge of public relations and outreach at AlgorithmWatch. He has a master's degree in media studies, has worked as a producer and editor in a film company, and managed projects for the Deutsche Kinemathek and the civil society organization Gesicht Zeigen. In addition to his core tasks at AlgorithmWatch, Marc has been involved in the crowdfunding and crowdsourcing campaign OpenSCHUFA, and he coordinated the first issue of the Automating Society report, published in 2019.

ORGANISATIONEN

/ AlgorithmWatch Schweiz

AlgorithmWatch ist eine gemeinnützige Forschungs- und Advocacy-Organisation mit dem Ziel, Systeme algorithmischer / automatisierter Entscheidungsfindung (ADM) und deren Auswirkungen auf die Gesellschaft zu beobachten und zu analysieren. Um menschliche Selbstbestimmung und Grundrechte zu schützen sowie das Gemeinwohl zu maximieren, halten wir es für entscheidend, ADM-Systeme zur Rechenschaft zu ziehen und sie einer demokratischen Kontrolle zu unterwerfen. Der Einsatz von ADM-Systemen, die wesentlich individuelle und kollektive Rechte beeinträchtigen, muss nicht nur nicht nur auf klare und zugängliche Weise öffentlich gemacht werden, Einzelpersonen müssen auch in der Lage sein zu verstehen, wie Entscheidungen zu zustandekommen und sie gegebenenfalls angefochten werden können. Deshalb befähigen wir die Bürger·innen, ADM-Systeme besser zu verstehen und Wege zu entwickeln, diese Prozesse demokratisch zu kontrollieren - mit einer Mischung aus Technologien, Regulierung und geeigneten Aufsichtsinstitutionen. Damit streben wir an, zu einer fairen und inklusiven Gesellschaft beizutragen und den Nutzen von ADM-Systemen für die gesamte Gesellschaft zu maximieren.

https://algorithmwatch.ch/de/

ALGORITHM WATCH / CH

/ Bertelsmann Stiftung

Die Bertelsmann Stiftung setzt sich für eine gerechte Teilhabe aller am gesellschaftlichen Leben ein. Sie engagiert sich in den Bereichen Bildung, Demokratie, Gesellschaft, Gesundheit, Kultur und Wirtschaft. Durch ihr Engagement will sie alle Bürgerinnen und Bürger ermutigen, sich für das Gemeinwohl einzusetzen. Die 1977 von Reinhard Mohn gegründete, gemeinnützige Einrichtung hält die Mehrheit der Kapitalanteile der Bertelsmann SE & Co. KGaA. Die Bertelsmann Stiftung arbeitet operativ und ist unabhängig vom Unternehmen sowie parteipolitisch neutral. Die Bertelsmann Stiftung setzt sich im Projekt „Ethik der Algorithmen“ mit den gesellschaftlichen Folgen algorithmischer Entscheidungsfindung auseinander. Das Ziel des Projekts ist es, zu einer Gestaltung algorithmischer Systeme beizutragen, die zu mehr Teilhabe für alle führt. Nicht das technisch Mögliche, sondern das gesellschaftlich Sinnvolle muss Leitbild sein – damit maschinelle Entscheidungen den Menschen dienen.

https://www.bertelsmann-stiftung.de

BertelsmannStiftung

/ Engagement Migros

Der Förderfonds Engagement Migros ermöglicht Pionierprojekte im gesellschaftlichen Wandel, die neue Wege beschreiten und zukunftsgerichtete Lösungen erproben. Der wirkungsorientierte Förderansatz verbindet finanzielle Unterstützung mit coachingartigen Leistungen im Pionierlab. Engagement Migros wird von den Unternehmen der Migros- Gruppe mit jährlich rund 10 Millionen Franken ermöglicht und ergänzt seit 2012 das Migros- Kulturprozent.

https://www.engagement-migros.ch

ENGAGEMENT
EIN FÖRDERFONDS DER MIGROS-GRUPPE

Vivre dans une société automatisée. Comment les systèmes de prise de décision automatisée se sont-ils généralisés, et que peut-on y faire ?

Par Fabio Chiusi

La date de bouclage de ce rapport était le 30 septembre 2020.
Les développements ultérieurs n'ont pas pu être inclus.

Par une journée nuageuse d'août, à Londres, les étudiant·es étaient en colère. Ils et elles affluaient par centaines sur Parliament Square pour manifester leur colère, leurs pancartes affichant leur soutien pour des alliés inhabituels : leurs professeurs, et une cible plus insolite encore – un algorithme.

Les écoles du Royaume-Uni avaient fermé leurs portes en mars, en raison de la pandémie de COVID-19. Alors que le virus continuait à sévir à travers l'Europe à l'été 2020, les étudiant·es savaient que leurs examens de fin d'année allaient être annulés et leurs évaluations, d'une manière ou d'une autre, modifiées. Ce qu'ils et elles n'auraient pas pu imaginer, cependant, c'est que des milliers d'entre eux allaient se retrouver avec des notes plus basses que prévu.

Les étudiant·es qui manifestaient savaient quel était le responsable, comme le laissaient entendre leurs chants et leurs pancartes : le système de prise de décision automatisée (ADM, pour *automated decision-making*) mis en place par l'Office of Qualifications and Examinations Regulation (Ofqual). Celui-ci prévoyait de produire la meilleure évaluation possible en se basant sur les données disponibles pour les résultats des deux certificats de fin d'études secondaires, le GCSE et le *A level*, de sorte que « la distribution des notes suive un modèle similaire à celui des autres années, afin que les étudiants de cette année ne se retrouvent pas pénalisés par les circonstances ».

Le gouvernement souhaitait éviter l'excès d'optimisme[1] qui aurait résulté du seul jugement humain, d'après ses propres estimations : par rapport aux séries des années précédentes, les notes auraient été trop élevées. Mais à vouloir être « juste, autant que possible, envers les étudiant·es qui n'ont pas pu passer leurs examens cet été », le gouvernement a essuyé un échec spectaculaire, et en cette grise journée d'août, les étudiant·es continuaient d'affluer, de scander des chants et de brandir des pancartes pour exprimer leur besoin urgent de justice sociale. Certain·es étaient désespérés, d'autres s'effondraient en pleurant.

« Arrêtez de nous voler notre avenir », pouvait-on lire sur une pancarte, faisant écho aux manifestations des « vendredis pour l'avenir » des militant·es écologistes. D'autres, en revanche, ciblaient plus spécifiquement les failles du système de notation ADM : « notez mon travail, pas mon code postal », ou encore « nous sommes des étudiant·es, pas des statistiques », dénonçant les résultats discriminatoires du système[2].

Enfin, un chant jaillit de la foule, un chant qui est devenu le symbole de la contestation : « Fuck the algorithm ». Craignant que le gouvernement n'automatise leur avenir de manière opaque et désinvolte, sans tenir compte de leurs compétences et de leurs efforts, les étudiant·es se mirent à crier pour ne pas voir leurs chances être indûment affectées par un code mal conçu. Ils et elles voulaient avoir leur mot à dire, et nous ferions bien de les écouter.

Les algorithmes ne sont ni « neutres », ni « objectifs » ; pourtant, nous avons tendance à penser qu'ils le sont. En vérité, ils ne font que reproduire les préjugés et les croyances de ceux et celles qui les programment et les déploient. Ces individus sont donc, ou du moins devraient être, responsables des bons et des mauvais choix algorithmiques, non pas les « algorithmes » ou les systèmes ADM eux-mêmes. La machine est effrayante, mais le fantôme en son sein est toujours humain. Et les êtres humains, bien plus encore que les algorithmes, sont des machines complexes.

Dans tous les cas, les étudiant·es contestataires n'étaient pas naïfs au point de croire que leurs malheurs étaient exclusivement le fait d'un algorithme. D'ailleurs, ils et elles n'attaquaient pas l'« algorithme » dans un élan de déterminisme technologique : ils et elles étaient motivés par un désir de protection et de promotion de la justice sociale. À cet égard, leur protestation ressemble davantage à celle des luddites. Tout comme ce mouvement ouvrier qui détruisait les métiers à tisser et à tricoter mécaniques au XIX^{e} siècle, ils et elles savent que les systèmes ADM sont purement une histoire de pouvoir, et ne doivent pas être considérés comme une technologie prétendument objective. Alors, ils se mirent à scander « justice pour la classe ouvrière » et à demander la démission du ministre de la Santé, décrivant le système ADM comme étant une preuve flagrante de « classisme ».

Pour finir, les étudiant·es parvinrent à abolir le système qui mettait en danger leur parcours éducatif et leurs chances dans la vie : dans un revirement spectaculaire, le gouvernement britannique abandonna le système ADM, propice aux erreurs, et utilisa les notes prédites par les enseignant·es.

1 « Les travaux de recherche suggèrent que, en estimant les notes que les élèves sont susceptibles d'obtenir, les enseignants ont tendance à être optimistes (quoique pas dans tous les cas) », écrit l'Ofqual, cf. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/909035/6656-2_-_Executive_summary.pdf

2 Cf. le chapitre sur le Royaume-Uni pour plus de détails.

Mais cette histoire ne se résume pas au fait que les manifestant·es ont finalement eu gain de cause. Cet exemple illustre parfaitement comment des systèmes mal conçus, mis en œuvre et supervisés, qui reproduisent les préjugés et la discrimination de leurs créateurs, ne tirent pas parti du potentiel des systèmes ADM, par exemple en matière de comparabilité et d'équité.

Cette contestation révèle, plus clairement que bien d'autres luttes du passé, que nous ne nous contentons plus d'automatiser la société. Nous l'avons déjà automatisée – et quelqu'un l'a enfin remarqué.

/ De l'automatisation de la société à la société automatisée

Lorsque nous avons publié la première édition de ce rapport, nous avions décidé de l'appeler « *L'automatisation de la société* », car les systèmes ADM étaient pour l'essentiel nouveaux, expérimentaux et inexplorés – et surtout, ils étaient l'exception plutôt que la norme.

Cette situation a rapidement évolué. Comme le révèlent clairement les nombreux cas réunis dans ce rapport grâce à notre réseau exceptionnel de chercheur·ses, le déploiement des systèmes ADM a considérablement augmenté en à peine plus d'un an. Ceux-ci touchent désormais toutes sortes d'activités humaines, et plus particulièrement la distribution des services à des millions de citoyen·nes européens – ainsi que leur accès à leurs droits.

L'opacité tenace qui entoure l'utilisation toujours croissante des systèmes ADM nous oblige d'autant plus urgemment à redoubler d'efforts. C'est pourquoi nous avons ajouté quatre pays (l'Estonie, la Grèce, le Portugal et la Suisse) aux 12 que nous avions déjà analysés dans l'édition précédente de ce rapport, portant le total à 16 pays. Bien que cette liste soit loin d'être exhaustive, cela nous permet de brosser un tableau plus large des scénarios d'ADM à travers l'Europe. Compte tenu de l'impact que ces systèmes peuvent avoir sur notre vie quotidienne, et de la profondeur avec laquelle ils remettent en question nos intuitions – voir nos normes et nos règles – sur la relation entre la gouvernance démocratique et l'automatisation, nous pensons qu'il s'agit là d'une initiative essentielle.

Cela s'avère particulièrement vrai dans le contexte de la pandémie de COVID-19, une période au cours de laquelle nous avons assisté à l'adoption (souvent précipitée) d'une multitude de systèmes ADM visant à contribuer à la sécurisation de la santé publique grâce à des outils basés sur des données et à l'automatisation. Nous considérons que cette évolution est si importante que nous avons décidé d'y consacrer un « rapport préliminaire », publié en août 2020 dans le cadre du projet *L'automatisation de la société*.

Même en Europe, les exemples de systèmes ADM déployés sont légion. Songez simplement à certains des cas présentés dans ce rapport, qui viennent s'ajouter aux nombreux cas – de la sécurité sociale à l'éducation, en passant par le système de santé et la justice – que nous avions déjà abordés dans l'édition précédente. Dans les pages qui suivent, et pour la première fois, nous faisons le point sur le développement de ces cas de trois manières. Tout d'abord, par le biais d'articles journalistiques, puis par des sections de recherche cataloguant différents exemples, et enfin, par des bandes dessinées. Ces systèmes ADM sont devenus tellement omniprésents dans nos vies que nous voulions communiquer leur fonctionnement et ce qu'ils *font réellement pour nous*, de manière à la fois rigoureuse et innovante, afin de toucher toutes sortes de publics. Après tout, les systèmes ADM ont un impact sur nous tous·tes.

Ou du moins, ils le devraient. Nous avons pu voir, par exemple, comment un nouveau service automatisé et proactif permet de distribuer les allocations familiales en Estonie. Les parents n'ont même plus besoin de demander ces allocations : dès la naissance, l'État recueille toutes les informations sur chaque nouveau-né et ses parents et les rassemble dans des bases de données. Ainsi, les parents reçoivent automatiquement les prestations auxquelles ils ont droit.

En Finlande, l'identification des facteurs de risque individuels liés à l'exclusion sociale chez les jeunes adultes est automatisée grâce à un outil développé par le géant japonais Fujitsu. En France, les données des réseaux sociaux peuvent être extraites pour alimenter des algorithmes à apprentissage automatique qui sont employés pour détecter la fraude fiscale.

L'Italie teste actuellement la « jurisprudence prédictive ». Cette méthode a recours à l'automatisation pour aider les juges à comprendre les tendances de décisions de justice précédentes sur un sujet particulier. Et au Danemark, le gouvernement a voulu surveiller le clavier et la souris de chaque étudiant·e pendant les examens, ce qui a entraîné – une fois n'est pas coutume – des manifestations d'étudiant·es massives qui ont conduit au retrait du système, du moins pour l'instant.

/ Redressons les torts de l'ADM

En principe, les systèmes ADM sont susceptibles d'améliorer la vie des gens en traitant d'énormes quantités de données, en aidant les personnes impliquées dans des processus décisionnels et en fournissant des applications sur mesure.

En pratique, cependant, nous avons trouvé très peu de cas qui démontraient de manière convaincante un tel impact positif.

Parmi ceux-ci, le système VioGén, déployé en Espagne depuis 2007 pour évaluer les risques dans les affaires de violence domestique, bien qu'il soit loin d'être parfait, affiche des « indices de performance raisonnables » et a contribué à protéger de nombreuses femmes contre la maltraitance à leur encontre.

Au Portugal, un système automatisé centralisé déployé pour dissuader la fraude aux prescriptions médicales, a apparemment réduit la fraude de 80 % en une seule année. En Slovénie, un système similaire utilisé pour lutter contre la fraude fiscale s'est révélé utile pour les inspecteurs, selon les autorités fiscales[3].

La reconnaissance faciale, qui était pratiquement absente de l'édition 2019, est testée et déployée à un rythme alarmant dans toute l'Europe.

Lorsque l'on regarde l'état actuel des systèmes ADM en Europe, les exemples positifs présentant des avantages évidents se font rares. Tout au long de ce rapport, nous décrivons comment la grande majorité des utilisations a plutôt tendance à nuire aux gens qu'à les aider. Mais pour juger véritablement de l'impact positif et négatif de ces systèmes, nous avons besoin de plus de transparence sur la finalité et de plus de données sur le fonctionnement des systèmes ADM qui sont testés et déployés.

Le message destiné aux responsables politiques ne pourrait être plus clair. Si nous souhaitons vraiment tirer le meilleur parti de leur potentiel tout en respectant les droits fondamentaux et la démocratie, le moment est venu de passer à l'action, de rendre ces systèmes transparents et de remédier aux injustices de l'ADM.

/ La reconnaissance faciale à chaque coin de rue

Différents pays adoptent différents outils. Il y a toutefois une technologie qui est commune dans la plupart d'entre eux : la reconnaissance faciale. Il s'agit sans doute du développement le plus récent, le plus rapide et le plus préoccupant présenté dans ce rapport. La reconnaissance faciale, qui était pratiquement absente de l'édition 2019, est testée et déployée à un rythme alarmant dans toute l'Europe. En à peine plus d'un an depuis notre dernier rapport, la reconnaissance faciale a fait son apparition dans les écoles, les stades, les aéroports et même les casinos. Elle est également utilisée dans des applications de police prédictive, pour appréhender les criminels, pour lutter contre le racisme, et, dans le cadre de la pandémie de COVID-19, pour faire respecter la distanciation sociale, à la fois par le biais d'applications et de systèmes de vidéosurveillance « intelligente ».

Les nouveaux déploiements de l'ADM se poursuivent, alors même que les preuves de leur manque de précision s'accumulent. Et lorsque des difficultés apparaissent, les partisans de ces systèmes essaient simplement de les dissimuler d'une manière ou d'une autre. En Belgique, un système de reconnaissance faciale utilisé par la police est toujours « partiellement actif », bien qu'une interdiction temporaire ait été prononcée par l'Organe de contrôle de l'information

3 Cf. le chapitre sur la Slovénie pour plus de détails.

policière. Et en Slovénie, l'utilisation de la technologie de reconnaissance faciale par la police a été légalisée cinq ans après avoir fait ses débuts.

Cette tendance, si elle n'est pas remise en cause, risque de normaliser l'idée que l'on peut être surveillé en permanence et de manière opaque, cristallisant ainsi un nouveau statu quo de surveillance de masse généralisée. C'est la raison pour laquelle de nombreux membres de la communauté des libertés civiles auraient souhaité voir une réponse politique beaucoup plus forte de la part des institutions européennes face à cette situation[4].

Même le simple fait de sourire fait désormais partie d'un système ADM employé dans certaines banques en Pologne : plus un·e employé·e sourit, plus sa prime est importante. Et ce ne sont pas seulement les visages qui sont surveillés : en Italie, un système de surveillance sonore a été proposé pour lutter contre le racisme dans les stades de football.

/ Les boîtes noires sont toujours des boîtes noires

En 2015, Frank Pasquale, professeur à la faculté de droit de Brooklyn, disait qu'une société interconnectée basée sur des systèmes algorithmiques opaques était une sorte de « boîte noire ». Cinq ans plus tard, malheureusement, la métaphore reste valable – et elle s'applique à tous les pays que nous avons étudiés dans le cadre de ce rapport, dans tous les domaines : la transparence des systèmes ADM est insuffisante, que ce soit dans le secteur public ou privé. En Pologne, cette opacité est même imposée, comme en témoigne la loi qui a instauré son système automatisé de détection des comptes bancaires utilisés à des fins illégales (« STIR »). En effet, la loi dispose que la divulgation des algorithmes et des indicateurs de risque employés peut entraîner jusqu'à 5 ans de prison.

Bien que nous rejetions catégoriquement l'idée que tous ces systèmes sont intrinsèquement mauvais (nous adoptons plutôt une approche objective et factuelle), il est incontestablement dommageable de ne pas pouvoir évaluer leur fonctionnement et leur impact sur la base de connaissances précises et factuelles. Ne serait-ce que parce que l'opacité entrave sérieusement la collecte des preuves nécessaires pour pouvoir porter un jugement éclairé sur le déploiement d'un système ADM.

Ajoutez à cela la difficulté que nos chercheurs et nos journalistes ont rencontrée pour accéder à des données sur ces systèmes, et vous obtenez un scénario préoccupant pour quiconque souhaite les contrôler et s'assurer que leur déploiement soit compatible avec les droits fondamentaux, l'État de droit et la démocratie.

/ Remettre en question le status quo algorithmique

Que fait l'Union européenne à ce sujet ? Si les documents stratégiques produits par la Commission européenne, sous l'égide d'Ursula Von der Leyen, parlent d'« intelligence artificielle » plutôt que de faire directement référence aux systèmes ADM, ils expriment toutefois des intentions louables : promouvoir et développer une « IA digne de confiance », qui accorde la « priorité aux individus »[5].

Cependant, comme nous le décrivons dans le chapitre consacré à l'Europe, l'approche globale de l'UE privilégie l'impératif commercial et géopolitique visant à mener la « révolution de l'IA » plutôt que de s'assurer que ses produits soient conformes aux mécanismes de protection démocratiques, une fois adoptés comme outils politiques.

Cette absence de courage politique, qui est particulièrement évidente dans la décision d'abandonner toute suggestion de moratoire sur les technologies de reconnaissance faciale dans les lieux publics dans les règlements européens sur l'IA, est surprenante. Surtout à une époque où de nombreux États membres se voient confrontés à un nombre croissant de difficultés et de revers juridiques en raison de systèmes ADM déployés à la va-vite qui ont eu un impact négatif sur les droits des citoyens.

Une affaire historique nous vient des Pays-Bas, où des défenseurs des droits civils ont porté devant les tribunaux un système automatisé invasif et opaque censé détecter la fraude aux aides sociales (SyRI), et ont obtenu gain de cause. En effet, ce système a été jugé contraire à la Convention européenne des droits de l'homme par un tribunal de la Haye en février, et a donc été abandonné. Mais l'affaire a également fait jurisprudence : selon l'arrêt de la cour, les gouvernements ont la « responsabilité particulière » de protéger les droits fondamentaux lorsqu'ils mettent en œuvre de tels systèmes ADM. Assurer cette transparence

4 Comme détaillé dans le chapitre sur l'Europe

5 Cf. le chapitre sur l'Europe, et en particulier la section sur le « livre blanc sur l'IA » de la Commission européenne.

Face au déploiement continu de systèmes d'ADM à travers l'Europe, on est en droit de se demander : le niveau de supervision actuel est-il suffisant ?

indispensable est considéré comme un élément crucial de cette responsabilité.

Depuis notre premier rapport, les médias et les militant·es de la société civile se sont imposés comme une véritable force motrice de la responsabilisation vis-à-vis des systèmes ADM. En Suède, par exemple, des journalistes sont parvenus à obtenir la publication du code du système Trelleborg, qui permet de prendre des décisions entièrement automatisées concernant les demandes de prestations sociales. À Berlin, le projet pilote de reconnaissance faciale de la gare de Südkreuz n'a pas réussi à déboucher sur l'application du système dans toute l'Allemagne. Cette issue n'a été possible que grâce à la bruyante opposition des militant·es, si bruyante qu'ils et elles sont parvenus à influencer la position des partis et, en fin de compte, le programme politique des gouvernements.

Les militants grecs d'Homo Digitalis ont pu démontrer qu'aucun vrai voyageur n'avait participé aux essais du système nommé « iBorderCtrl », un projet financé par l'UE qui visait à utiliser l'ADM pour les contrôles aux frontières, révélant ainsi que les capacités de beaucoup de ces systèmes sont souvent surestimées. Dans le même temps, au Danemark, un système de profilage pour la détection précoce des risques associés aux familles et aux enfants vulnérables (le « modèle de Gladsaxe ») a été mis en suspens grâce au travail d'universitaires, de journalistes et de l'Autorité de protection des données (APD) nationale.

Les APD elles-mêmes ont également joué un rôle majeur dans d'autres pays. En France, la CNIL, autorité nationale de protection de la vie privée, a statué qu'un projet de surveillance sonore et un projet de reconnaissance faciale dans des lycées étaient tous deux illégaux. Au Portugal, l'APD a refusé d'approuver la mise en place de systèmes de vidéosurveillance par la police dans les municipalités de Leiria et Portimão, car ceux-ci ont été jugés disproportionnés et auraient constitué « une surveillance et un suivi à grande échelle des personnes, de leurs habitudes et de leur comportement, ainsi qu'une identification des personnes à partir de données relatives à leurs caractéristiques physiques ». Parallèlement, aux Pays-Bas, l'APD néerlandaise a demandé plus de transparence dans les algorithmes prédictifs utilisés par les agences gouvernementales.

Enfin, certains pays ont eu recours à un médiateur pour se faire conseiller. Au Danemark, ces conseils ont permis d'élaborer des stratégies et des orientations éthiques concernant l'utilisation des systèmes ADM dans le secteur public. En Finlande, le médiateur parlementaire adjoint a estimé que les évaluations fiscales automatisées étaient illégales.

Et pourtant, devant le déploiement continu de ces systèmes à travers l'Europe, on est en droit de se demander : ce niveau de supervision est-il suffisant ? Lorsque le médiateur polonais a mis en doute la légalité du système de détection des sourires utilisé dans une banque (et mentionné ci-dessus), sa décision n'a pas empêché un projet pilote ultérieur d'être entrepris dans la ville de Sopot, ni dissuadé plusieurs entreprises de manifester leur intérêt pour l'adoption de ce système.

/ L'inadéquation des audits, de la mise en application, des compétences et des explications

Le militantisme est principalement une démarche réactive. La plupart du temps, les militant·es ne peuvent réagir que si un système ADM est en train d'être testé ou s'il a déjà été déployé. Et le temps que les citoyen·nes mettent sur pied une réponse, il se peut que leurs droits aient déjà

été indûment piétinés, même avec les protections censées être accordées par le droit européen et la législation des États membres. C'est pourquoi il est si important de prendre des mesures proactives pour protéger les droits des citoyen·nes, avant la réalisation de projets pilotes et de déploiements à grande échelle.

Et pourtant, même dans les pays où une législation préventive est en place, celle-ci n'est pas appliquée. En Espagne, par exemple, toute « action administrative automatisée » est codifiée par la loi, qui prévoit des exigences spécifiques en matière de contrôle de la qualité et de supervision, ainsi que l'audit du système informatique et de son code source. L'Espagne bénéficie également d'une loi sur la liberté de l'information. Cependant, même avec ces lois, il est rare, d'après notre chercheur, que les organismes publics divulguent des informations détaillées sur les systèmes ADM qu'ils utilisent. De même, en France, il existe bien une loi de 2016 qui exige la transparence des algorithmes, mais là encore, en vain.

Même le fait d'ester en justice pour obtenir la transparence d'un algorithme, conformément aux dispositions spécifiques d'une loi sur la transparence des algorithmes, peut ne pas suffire à faire respecter et à protéger les droits des utilisateurs. Comme en témoigne le cas de l'algorithme de Parcoursup, destiné à classer et trier les candidats à l'université en France[6], des exceptions peuvent être prévues à la discrétion du législateur pour dégager une administration de toute responsabilité.

Ce phénomène est particulièrement troublant lorsqu'il vient s'ajouter au déficit généralisé d'aptitudes et de compétences entourant les systèmes ADM dans le secteur public, déploré par de nombreux·ses chercheur·ses. Comment les responsables publics pourraient-ils expliquer ou faire preuve d'une quelconque transparence à l'égard de systèmes qu'ils ne comprennent pas eux-mêmes ?

Récemment, plusieurs pays se sont efforcés de résoudre ce problème. L'Estonie, par exemple, a mis en place un centre de compétences consacré aux systèmes ADM afin de mieux déterminer comment ceux-ci pourraient être utilisés pour développer les services publics et, plus particulièrement, pour guider l'action du ministère des Affaires économiques et des Communications et de la chancellerie d'État pour le développement de l'administration en ligne. La Suisse a également appelé à la création d'un « réseau de compétences » dans le cadre plus large de la stratégie nationale de « Suisse numérique ».

Et pourtant, le manque de culture numérique est un problème bien connu qui touche une grande partie de la population dans plusieurs pays européens. En outre, il est difficile de faire valoir des droits dont on ignore l'existence. Les mouvements de contestation au Royaume-Uni et ailleurs, conjugués à quelques scandales très médiatisés impliquant des systèmes ADM[7], ont certainement sensibilisé la population aux risques et au potentiel de l'automatisation de la société. Mais bien qu'elle soit en hausse, cette prise de conscience n'en est qu'à ses débuts dans de nombreux pays.

Les résultats de notre étude sont clairs : si les systèmes ADM affectent déjà toutes sortes d'activités et de jugements, ils sont encore principalement déployés sans aucune sorte de débat démocratique significatif. Par ailleurs, on observe que dans l'ensemble, les mécanismes d'application et de contrôle – si tant est qu'ils existent – sont à la traîne par rapport au déploiement

La finalité même de ces systèmes n'est pas communément justifiée ou expliquée aux populations concernées, sans parler des bénéfices qu'elles sont censées en retirer. Prenons l'exemple du service proactif « AuroraAI » en Finlande : celui-ci est censé identifier automatiquement certains « événements de la vie », comme le rapportent nos chercheurs finlandais, et dans l'esprit de ses promoteurs, il doit en quelque sorte jouer le rôle d'une « nounou » qui aide les citoyens à répondre à des besoins particuliers de service public pouvant survenir en lien avec certaines circonstances de la vie, par exemple un déménagement, un changement de relations familiales, etc. Selon nos chercheurs, il est vraisemblable que ce système, au lieu de responsabiliser les individus, fasse exactement le contraire, en suggérant certaines décisions ou en limitant les options d'un individu en raison de sa conception et de son architecture.

Il est alors d'autant plus important de savoir ce que le système vise à « optimiser » en termes de services publics : « l'utilisation du service est-elle maximisée, les coûts sont-ils minimisés, le bien-être des citoyen·nes est-il amélioré ? », demandent les chercheurs. « Sur quel ensemble de critères ces décisions se fondent-elles, et qui les choisit ? » Le simple fait que n'ayons pas de réponse à ces questions

6 Cf. le chapitre sur la France

7 Voir la débâcle de l'algorithme « Buona Scuola », cf. le chapitre sur l'Italie.

fondamentales en dit long sur le degré de participation et de transparence qui est admis, même pour un système ADM potentiellement si intrusif.

/ Le piège technosolutionniste

Il existe une justification idéologique globale à tout cela. C'est ce que l'on appelle le « solutionnisme technologique », et c'est un phénomène qui affecte encore sérieusement la façon dont sont développés de nombreux systèmes ADM que nous avons étudiés. Même si cette expression est depuis longtemps dénoncée comme une idéologie fallacieuse qui perçoit chaque problème social comme un « bug » qui nécessite un « correctif » technologique[8], cette rhétorique est encore largement employée, tant dans les médias que dans les milieux politiques, pour justifier l'adoption inconditionnelle de technologies automatisées dans la vie publique.

Lorsqu'ils sont vendus comme des « solutions », les systèmes ADM passent immédiatement dans le domaine décrit par la troisième loi d'Arthur C. Clarke : la magie. Et il est difficile, voire impossible, de réglementer la magie, et plus encore de l'expliquer et de faire preuve de transparence à son égard. On peut voir la main qui se glisse dans le chapeau, et le lapin qui en ressort, mais le processus lui-même est, et *doit rester* une « boîte noire ».

De nombreux chercheurs impliqués dans le projet *L'automatisation de la société* ont dénoncé ce problème comme étant la faille fondamentale dans le raisonnement qui sous-tend nombre des systèmes ADM qu'ils décrivent. Cela implique également, comme le montre le chapitre sur l'Allemagne, que la plupart des critiques de ces systèmes sont présentées comme un rejet total de l'« innovation », dépeignant les défenseurs des droits numériques comme des « néo-luddites ». Non seulement cette attitude ignore la réalité historique du mouvement luddite, qui se préoccupait des politiques du travail et non des technologies en tant que telles, mais surtout, elle menace fondamentalement l'efficacité des mécanismes de supervision et d'application potentiels.

À l'heure où l'industrie de l'« IA » assiste à l'émergence d'un secteur de lobbying « dynamique », notamment au Royaume-Uni, cette tendance risque d'aboutir à des directives de « blanchiment éthique » et à d'autres réponses politiques qui seront inefficaces et structurellement inadaptées pour traiter les implications des systèmes ADM en matière de droits fondamentaux. Cette vision revient en définitive à supposer que nous, humains, devrions nous adapter aux systèmes ADM, bien plus que les systèmes ADM ne devraient être adaptés aux sociétés démocratiques.

Pour contrer ce raisonnement, nous ne devons pas nous abstenir de poser des questions fondamentales : les systèmes ADM peuvent-ils être compatibles avec la démocratie et déployés au profit de la société dans son ensemble, et pas seulement d'une partie de celle-ci ? Il se pourrait que certaines activités humaines – par exemple, dans le domaine de l'aide sociale – ne doivent pas faire l'objet d'une automatisation, ou que certaines technologies, notamment la reconnaissance faciale dans l'espace public, ne doivent pas être encouragées dans une quête sans fin de « leadership technologique », mais plutôt qu'elles soient interdites dans leur ensemble.

Plus encore, nous devons rejeter tout carcan idéologique qui nous empêche de poser de telles questions. Au contraire, ce dont nous avons besoin maintenant, c'est de voir les politiques changer concrètement, afin de permettre un meilleur contrôle de ces systèmes. Dans la section suivante, nous énumérons les principales exigences qui découlent de nos conclusions. Nous espérons qu'elles seront largement débattues pour être enfin mises en œuvre.

Ce n'est qu'à travers un débat démocratique informé, inclusif et étayé par des preuves que nous pourrons trouver le bon équilibre entre les avantages que les systèmes ADM peuvent apporter – et apportent – en termes de rapidité, d'efficacité, d'équité, de prévention et d'accès aux services publics, et les défis qu'ils représentent pour nos droits à tous.

Recommandations politiques

À la lumière des conclusions détaillées figurant dans l'édition 2020 du rapport *L'automatisation de la société*, nous recommandons les interventions politiques suivantes aux décideur·ses du Parlement européen et des parlements des États membres, de la Commission européenne, des gouvernements nationaux, ainsi qu'aux chercheur·ses et

8 Lire Evgeny Morozov (2014), *To Save Everything, Click Here. The Folly of Technological Solutionism*, Public Affairs, https://www.publicaffairsbooks.com/titles/evgeny-morozov/to-save-everything-click-here/9781610393706/

organisations de la société civile (organisations de défense des droits, fondations, syndicats, etc.) et du secteur privé (entreprises et associations professionnelles). Ces recommandations visent à mieux garantir que les systèmes ADM actuellement déployés et ceux qui sont en passe d'être mis en œuvre à travers l'Europe sont effectivement compatibles avec les droits fondamentaux et la démocratie :

1 Accroître la transparence des systèmes ADM

Sans être en mesure de savoir précisément comment, pourquoi et à quelle fin les systèmes ADM sont mis en place, tous les efforts visant à concilier les droits fondamentaux et les systèmes ADM sont voués à l'échec.

/ Établir des registres publics pour les systèmes ADM utilisés dans le secteur public

Nous demandons par conséquent qu'une législation soit adoptée au niveau de l'UE pour obliger les États membres à tenir des registres publics des systèmes ADM utilisés dans le secteur public.

Ces registres devront être assortis de l'obligation légale pour les responsables du système ADM de divulguer et de documenter la finalité du système, de donner une explication du modèle et sa logique sous-jacente ainsi que des informations sur les personnes qui ont développé le système. Ces informations devront être mises à disposition de manière facilement lisible et accessible, y compris sous forme de données numériques structurées basées sur un protocole standardisé.

Les autorités publiques ont la responsabilité particulière de faire la transparence sur les caractéristiques opérationnelles des systèmes ADM déployés dans l'administration publique. Cette nécessité a été mise en évidence par une récente plainte administrative en Espagne, qui fait valoir que « tout système ADM utilisé par l'administration publique devrait être rendu public par défaut ». Si elle est confirmée, cette décision pourrait faire jurisprudence en Europe.

Si les dispositifs de divulgation des systèmes ADM devraient être obligatoires pour le secteur public dans tous les cas, ces exigences de transparence devraient également s'appliquer à l'utilisation des systèmes ADM par des entités privées lorsqu'un système d'IA/ADM a un impact significatif sur un individu, un groupe spécifique ou la société dans son ensemble.

/ Créer des dispositifs d'accès aux données juridiquement contraignants pour soutenir et faciliter la recherche d'intérêt public

Pour accroître la transparence d'un système, il ne suffit pas de divulguer des informations sur sa finalité, sa logique et son créateur, et d'avoir la capacité d'analyser et de tester en profondeur ce qui entre et sort d'un système ADM. Il faut également rendre les données à partir desquelles le système a été entraîné aux chercheur·ses indépendant·es, aux journalistes et aux organisations de la société civile pour encourager la recherche d'intérêt public.

C'est pourquoi nous suggérons de créer des dispositifs d'accès aux données robustes et juridiquement contraignants, explicitement axés sur le soutien et la promotion de la recherche d'intérêt public, dans le respect de la législation sur la protection des données et de la vie privée.

En tirant les leçons des meilleures pratiques aux niveaux national et européen, ces dispositifs à plusieurs niveaux devraient inclure des systèmes de sanctions, de contrôles et de contre-pouvoirs, ainsi que des examens réguliers. Comme l'ont illustré les partenariats de partage de données privées, des préoccupations légitimes ont été exprimées concernant la vie privée des utilisateurs et la possibilité d'une désanonymisation de certains types de données.

Les responsables politiques ont tout intérêt à s'inspirer des dispositifs de partage des données de santé pour faciliter un accès privilégié à certains types de données plus détaillées, tout en veillant à ce que les données à caractère personnel soient protégées de manière adéquate (par exemple, grâce à des environnements d'exploitation sécurisés).

Bien qu'un cadre de responsabilisation efficace nécessite un accès transparent aux données de la plateforme, il s'agit là d'une exigence pour que de nombreuses méthodes d'audit soient également efficaces.

2 Instaurer un cadre de responsabilisation significatif pour les systèmes ADM

Comme l'ont montré les constatations faites en France et en Espagne, même si la transparence d'un système ADM

est exigée par la loi et/ou si des informations ont été divulguées, cela n'entraîne pas nécessairement de responsabilisation. Des mesures supplémentaires sont nécessaires pour garantir que les lois et les normes soient effectivement applicables.

/ Développer et établir des approches pour auditer efficacement les systèmes algorithmiques

Pour que la transparence ait un sens, nous devons compléter la première étape, qui consiste à établir un registre public, par des processus qui contrôlent efficacement les systèmes algorithmiques.

Le terme « audit » est largement utilisé, mais il n'y a pas de consensus sur sa définition. Dans ce contexte, nous entendons par « audit », conformément à la définition de l'ISO, un « processus systématique, indépendant et documenté visant à obtenir des preuves objectives et à les évaluer objectivement afin de déterminer dans quelle mesure les critères d'audit sont remplis ».

Nous n'avons pas encore de réponses satisfaisantes aux questions complexes[9] soulevées par l'audit des systèmes algorithmiques ; cependant, nos recherches indiquent clairement la nécessité de trouver des réponses dans le cadre d'un vaste processus d'engagement des parties prenantes et par des recherches dédiées et approfondies.

Les critères d'audit, tout comme les processus d'audit appropriés, doivent être élaborés selon une approche multipartite qui prenne activement en considération l'effet disproportionné qu'ont les systèmes ADM sur les groupes vulnérables et sollicite leur participation.

Nous demandons donc aux responsables politiques de mettre en place ces processus multipartites afin de clarifier les questions soulevées, et de mettre à disposition des sources de financement visant à permettre la participation des parties prenantes qui ont été jusqu'à présent mal représentées.

Nous demandons également la mise à disposition de ressources adéquates pour soutenir/financer des projets de recherche sur l'élaboration de modèles permettant de contrôler efficacement les systèmes algorithmiques.

/ Soutenir les organisations de la société civile en tant que gardiens des systèmes ADM

Nos conclusions indiquent clairement que le travail des organisations de la société civile est crucial pour lutter efficacement contre l'opacité des systèmes ADM. Par le biais de la recherche et du travail de sensibilisation, et souvent en coopération avec les universitaires et les journalistes, celles-ci sont intervenues à plusieurs reprises dans les débats politiques portant sur ces systèmes au cours des dernières années, veillant dans plusieurs cas à ce que l'intérêt public et les droits fondamentaux soient dûment pris en compte avant et après leur déploiement dans de nombreux pays européens.

Les acteurs de la société civile devraient donc être soutenus en tant que gardiens de l'« automatisation de la société ». En tant que tels, ils font partie intégrante de tout cadre de responsabilisation efficace pour les systèmes ADM.

/ Interdire la reconnaissance faciale qui pourrait équivaloir à une surveillance de masse

Tous les systèmes ADM ne sont pas aussi dangereux les uns que les autres, et une approche de la réglementation basée sur le risque, comme celle de l'Allemagne et de l'UE, reflète bien ce constat. Mais afin d'assurer une responsabilisation réaliste pour les systèmes identifiés comme risqués, des mécanismes efficaces de surveillance et de mise en œuvre doivent être mis en place. Cela est d'autant plus

9 En réfléchissant aux modèles potentiels d'audits algorithmiques, plusieurs questions se posent. 1) Qui/quoi (services/plateformes/produits) doit être audité ? Comment personnaliser les systèmes d'audit en fonction du type de plateforme/service ? 2) Quand un audit doit-il être entrepris par une institution publique (au niveau de l'UE, au niveau national, au niveau local), et quand peut-il être réalisé par des entités/experts privés (entreprises, société civile, chercheurs) ? 3) Comment clarifier la distinction entre l'évaluation de l'impact ex ante (c'est-à-dire au cours de la phase de conception) et ex post (c'est-à-dire en cours d'exploitation) et les défis respectifs ? 4) Comment évaluer les compromis entre les différents avantages et inconvénients de l'auditabilité (par exemple, la simplicité, la généralité, l'applicabilité, la précision, la flexibilité, l'interprétabilité, la confidentialité, l'efficacité d'une procédure d'audit peuvent être en tension) ? 5) Quelles informations doivent être disponibles pour qu'un audit soit efficace et fiable (par exemple, le code source, les données de formation, la documentation) ? Les auditeurs doivent-ils disposer d'un accès physique aux systèmes en cours de fonctionnement pour pouvoir effectuer un audit efficace ? 6) Quelle obligation de produire des preuves est nécessaire et proportionnée pour les vendeurs/prestataires de services ? 7) Comment s'assurer que l'audit est possible ? Les exigences en matière d'audit doivent-elles être prises en compte dans la conception des systèmes algorithmiques (« auditable par construction ») ? 8) Règles de publicité : lorsqu'un audit est négatif et que les problèmes ne sont pas résolus, quel doit être le comportement de l'auditeur, et dans quelle mesure cet échec peut-il être rendu public ? 9) Qui audite les auditeurs ? Comment s'assurer que les auditeurs soient tenus responsables ?

important pour les systèmes présentant un « risque élevé » de violation des droits des utilisateur·trices.

Un exemple crucial qui ressort de nos conclusions est la reconnaissance faciale. Il a été démontré que les systèmes ADM basés sur les technologies biométriques, notamment la reconnaissance faciale, constituent une menace particulièrement grave pour l'intérêt public et les droits fondamentaux, car ils ouvrent la voie à une surveillance massive et sans discernement – d'autant plus qu'ils sont malgré tout déployés à grande échelle et de manière opaque.

Nous appelons à ce que les utilisations publiques de la reconnaissance faciale qui pourraient équivaloir à une surveillance de masse soient interdites de manière décisive jusqu'à nouvel ordre, et de toute urgence, au niveau de l'UE.

Ces technologies peuvent même déjà être considérées comme illégales dans l'UE, au moins pour certaines utilisations, si elles sont déployées sans le « consentement spécifique » des sujets contrôlés. Cette interprétation juridique a été suggérée par les autorités belges, qui ont infligé une amende historique pour les déploiements de la reconnaissance faciale dans le pays.

3 Sensibiliser la population au sujet des algorithmes et renforcer le débat public sur les systèmes ADM

Une plus grande transparence des systèmes ADM ne sera véritablement utile que si ceux qui y sont confrontés, tels que les organismes de réglementation, le gouvernement et les organismes industriels, peuvent gérer ces systèmes et leur impact d'une manière responsable et prudente. En outre, les personnes concernées par ces systèmes doivent être en mesure de comprendre où, pourquoi et comment ces systèmes sont déployés. C'est pourquoi nous devons améliorer la connaissance des algorithmes à tous les niveaux, auprès des acteurs importants ainsi que du grand public, et favoriser des débats publics plus diversifiés sur les systèmes ADM et leur impact sur la société.

/ Établir des centres d'expertise indépendants sur l'ADM

Parallèlement à notre demande d'audit des algorithmes et de soutien à la recherche, nous appelons à la création de centres d'expertise indépendants sur l'ADM au niveau national pour surveiller, évaluer, mener des recherches, rédiger des rapports et fournir des conseils aux gouvernements et à l'industrie en coordination avec les organismes de réglementation, la société civile et les universités sur les implications sociétales et en matière de droits de l'homme de l'utilisation des systèmes ADM. Le rôle général de ces centres sera de créer un système de responsabilisation significatif et de renforcer les capacités.

Les centres nationaux d'expertise doivent impliquer les organisations de la société civile, les groupes de parties prenantes et les organismes chargés de l'application existants tels que les DPA et les organismes nationaux de défense des droits fondamentaux afin de profiter à tous les aspects de l'écosystème et de renforcer la confiance, la transparence et la coopération entre tous les acteurs.

En tant qu'organes officiels indépendants, les centres d'expertise joueraient un rôle central dans la coordination de l'élaboration des politiques et des stratégies nationales relatives à l'ADM et dans le renforcement des capacités (compétences) des agences de réglementation, des gouvernements et des organismes industriels existants afin de répondre à l'utilisation accrue des systèmes ADM.

Ces centres ne devraient pas avoir de pouvoirs réglementaires, mais fournir une expertise essentielle sur la meilleure façon de protéger les droits fondamentaux et de prévenir les dommages collectifs et sociétaux. Ils devraient, par exemple, aider les petites et moyennes entreprises (PME) à remplir leurs obligations d'études d'impact en matière de droits fondamentaux, notamment en inscrivant les systèmes ADM dans le registre public mentionné précédemment.

/ Promouvoir un débat démocratique inclusif et diversifié sur les systèmes ADM

Outre le renforcement des capacités et des compétences des personnes qui déploient les systèmes ADM, il est également essentiel de faire progresser la culture algorithmique du grand public par le biais d'un débat plus large et de programmes diversifiés.

Nos conclusions suggèrent que non seulement les systèmes ADM ne sont pas transparents pour le grand public lorsqu'ils sont utilisés, mais que même la décision de déployer ou non un système ADM à la base est généralement prise sans que le public en soit informé ou n'y participe.

Il est donc urgent d'inclure l'intérêt général dès le début dans la prise de décision sur les systèmes ADM.

Plus généralement, nous avons besoin d'un débat public plus diversifié sur l'impact de l'ADM. Nous ne devons pas nous contenter de laisser la parole à des groupes d'experts, mais rendre la question plus accessible au grand public. Cela signifie qu'il nous faut parler un langage autre que le langage technojudiciaire pour mobiliser le public et susciter son intérêt.

Pour ce faire, des programmes détaillés – visant à construire et faire progresser la compétence numérique – doivent également être mis en place. Si nous souhaitons favoriser un débat public informé et créer une autonomie numérique pour les citoyen·nes européens, nous devons commencer par développer et faire progresser la culture du numérique, en mettant l'accent sur les conséquences sociales, éthiques et politiques de l'adoption de systèmes ADM.

Poser les bases de l'avenir de l'ADM en Europe

Alors que les systèmes de prise de décision automatisée (ADM) occupent une place centrale dans la garantie des droits fondamentaux et dans la distribution des services publics en Europe, les institutions de l'Union Européenne sont de plus en plus conscientes de leur rôle dans l'espace publique et privé, aussi bien en termes d'opportunités que de défis.

Par Kristina Penner et Fabio Chiusi

Depuis la publication de notre premier rapport en janvier 2019, et alors même que l'Europe est encore embourbée dans un débat plus global autour de l'intelligence artificielle « digne de confiance », plusieurs institutions, du Parlement européen au Conseil de l'Europe, ont publié des documents visant à donner à l'UE et à l'Europe une orientation en vue de traiter la question de l'ADM au cours des années, voire des décennies à venir.

À l'été 2019, Ursula von der Leyen, nouvelle présidente de la Commission et « techno-optimiste » autoproclamée, s'est engagée à proposer une « législation pour une approche européenne coordonnée sur les implications humaines et éthiques de l'intelligence artificielle » et à « réglementer l'intelligence artificielle (IA) » dans les 100 jours suivant son investiture. En février 2020, la Commission européenne a publié un « livre blanc » sur l'IA contenant « des idées et des actions » – un ensemble de stratégies visant à informer les citoyen·nes et à poser les bases d'une future action législative. Celui-ci plaide également en faveur d'une « souveraineté technologique » européenne : pour reprendre les termes de Von der Leyen, cela se traduit par « la capacité que doit avoir l'Europe de faire ses propres choix, sur la base de ses propres valeurs et en respectant ses propres règles », et devrait « contribuer à faire de nous tous des techno-optimistes ».

Une deuxième initiative fondamentale ayant trait à l'ADM en Europe est le « package législatif » sur les services numériques (DSA, *Digital Services Act*), annoncée dans l'« Agenda pour l'Europe » de Von der Leyen, et censée remplacer la directive sur le commerce électronique en vigueur depuis 2000. Ce package, qui vise à « actualiser nos règles de responsabilité et de sécurité pour les plateformes, services et produits numériques, et à instaurer un marché unique numérique », devrait conduire à des débats fondamentaux sur le rôle de l'ADM dans les politiques de modération de contenu, la responsabilité des intermédiaires et la liberté d'expression de manière générale[10].

Une attention particulière est prêtée aux systèmes ADM dans une résolution approuvée par la Commission du marché intérieur et de la protection des consommateurs du Parlement européen, ainsi que dans une recommandation « sur l'impact des systèmes algorithmiques sur les droits de l'homme » du Comité des ministres du Conseil de l'Europe (une organisation distincte de l'Union Européenne, à ne pas confondre avec le Conseil Européen).

Le Conseil de l'Europe (CdE), en particulier, s'est retrouvé à jouer un rôle de plus en plus important dans le débat

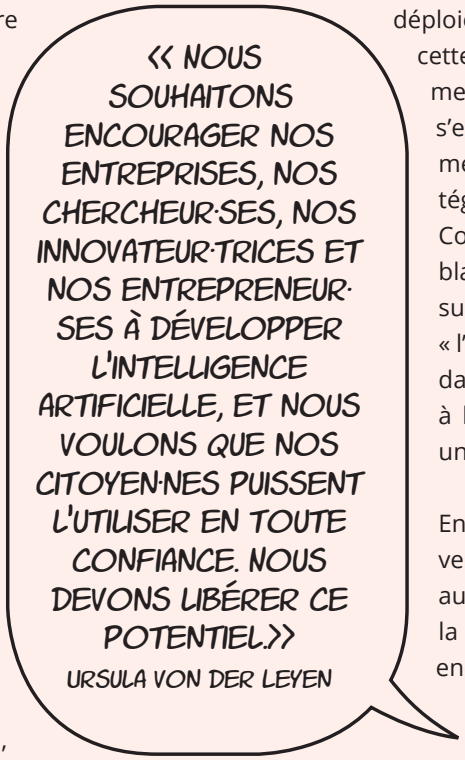
10 Des remarques et recommandations détaillées sur les systèmes ADM dans le contexte de la loi DSA peuvent être trouvées dans les conclusions du projet « Governing Platforms » d'AlgorithmWatch.

DE NOMBREUX ACTEURS SENTENT UNE TENSION FONDAMENTALE ENTRE LES IMPÉRATIFS ÉCONOMIQUES ET JURIDIQUES DANS LA MANIÈRE DONT LES INSTITUTIONS EUROPÉENNES, EN PARTICULIER LA COMMISSION, FORMULENT LEURS RÉFLEXIONS ET LEURS PROPOSITIONS SUR L'IA ET L'ADM.

politique sur l'IA au cours de l'année passée, et même si son impact réel sur les initiatives réglementaires reste à démontrer, il pourrait bien s'avérer jouer le rôle de « garant » des droits de l'homme. Cette intention ressort clairement de la recommandation intitulée « Décortiquer l'intelligence artificielle : 10 étapes pour protéger les droits de l'homme », rédigée par le Commissaire aux droits de l'homme du CdE, Dunja Mijatović, et dans les travaux du Comité ad hoc sur l'IA (CAHAI) fondé en septembre 2019.

De nombreux observateurs perçoivent une tension fondamentale entre les impératifs économiques et juridiques dans la manière dont les institutions européennes, en particulier la Commission, formulent leurs réflexions et leurs propositions sur l'IA et l'ADM. D'une part, l'Europe souhaite « accroître l'utilisation et la demande de données et de produits et services basés sur les données dans l'ensemble du marché unique », afin de devenir un « leader » des applications commerciales de l'IA, et ainsi de booster la compétitivité des entreprises européennes face à la pression croissante exercée par leurs concurrents aux États-Unis et en Chine. C'est d'autant plus important pour l'ADM, l'hypothèse étant que, grâce à cette économie « data-agile », l'UE pourra « devenir un leader de premier plan pour une société capable de prendre de meilleures décisions grâce aux données – dans les entreprises comme dans le secteur public ». Comme le souligne le livre blanc sur l'IA, « les données sont la pierre angulaire du développement économique ».

D'autre part, le traitement automatisé des données sur la santé, l'emploi et les prestations sociales d'un·e citoyen·ne est susceptible de donner lieu à des décisions aux résultats injustes et discriminatoires. Ce « côté obscur » des algorithmes utilisés dans les processus décisionnels est abordé dans la boîte à outils de l'UE à travers une série de principes. Dans le cas des systèmes à haut risque, des règles doivent garantir que les processus décisionnels automatisés sont compatibles avec les droits fondamentaux et les mécanismes de contrôle démocratiques. Il s'agit d'une approche unique que les institutions européennes qualifient de « centrée sur l'humain », diamétralement opposée à celles appliquées aux États-Unis (guidée par le profit) et en Chine (guidée par la sécurité nationale et la surveillance de masse).



Cependant, des doutes sont apparus quant à la possibilité pour l'Europe d'atteindre ces deux objectifs en même temps. La reconnaissance faciale en est une excellente illustration : même si, comme le montre ce rapport, nous avons désormais des preuves abondantes de déploiements incontrôlés et opaques de cette technologie dans la plupart des pays membres, la Commission européenne ne s'est pas montrée capable d'agir rapidement et de manière décisive pour protéger les droits des citoyens européens. Comme l'ont révélé les fuites du livre blanc de la Commission européenne sur l'IA[11], l'UE était en passe d'interdire « l'identification biométrique à distance » dans les lieux publics, avant de se défiler à la dernière minute et de promouvoir un « grand débat » sur le sujet à la place.

Entre temps, des applications controversées de l'ADM pour les contrôles aux frontières, employant notamment la reconnaissance faciale, sont toujours encouragées dans des projets financés par l'UE.

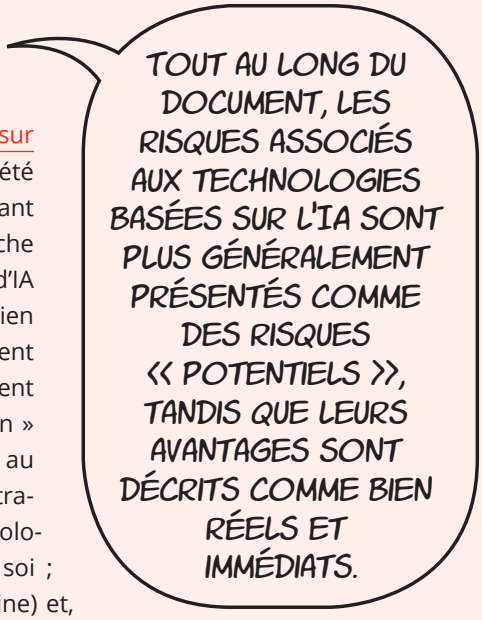
Politiques et débats

/ La stratégie européenne pour les données et le Livre blanc sur l'IA

Alors que la législation complète promise « pour une approche européenne coordonnée sur les implications humaines et éthiques de l'intelligence artificielle », annoncée par Von der Leyen dans son « Agenda pour l'Europe », n'a pas été mise en œuvre au cours des « 100 premiers jours de son mandat », la Commission européenne a publié une série de documents qui fournissent un ensemble de principes et d'idées qui devraient la guider.

11 https://www.politico.eu/article/eu-considers-temporary-ban-on-facial-recognition-in-public-spaces/

Le 19 février 2020, une « Stratégie européenne pour les données » et un « livre blanc sur l'intelligence artificielle » ont été publiés conjointement, établissant les grands principes de l'approche stratégique de l'UE en matière d'IA (qui inclut les systèmes ADM, bien qu'ils n'y soient pas explicitement mentionnés). Ces principes incluent notamment la « priorité à l'humain » (« une technologie qui fonctionne au service des personnes »), la neutralité technologique (aucune technologie n'est bonne ou mauvaise en soi ; c'est son utilisation qui le détermine) et, bien sûr, la souveraineté et l'optimisme. Comme le déclare Von der Leyen : « Nous souhaitons encourager nos entreprises, nos chercheur·ses, nos innovateur·trices et nos entrepreneur·ses à développer l'intelligence artificielle, et nous voulons que nos citoyen·nes puissent l'utiliser en toute confiance. Nous devons libérer ce potentiel. »



L'idée sous-jacente est que les nouvelles technologies ne doivent pas amener à de nouvelles valeurs. Le « nouveau monde numérique » imaginé par l'administration Von der Leyen doit protéger pleinement les droits fondamentaux et les droits civils. L'« excellence » et la « confiance », termes mis en avant dans le titre même du livre blanc, sont considérées comme les deux piliers sur lesquels un modèle européen de l'IA peut et doit reposer, se différenciant ainsi des stratégies américaine et chinoise.

Cependant, cette ambition est absente des détails du livre blanc. Par exemple, celui-ci propose une approche de la réglementation de l'IA basée sur le risque, dans laquelle la réglementation est proportionnelle à l'impact des systèmes d'IA sur la vie des citoyen·nes. « Pour les cas à haut risque, comme dans le domaine de la santé, de la police ou du transport, peut-on lire, les systèmes d'IA doivent être transparents et traçables, et garantir une supervision humaine ». Les tests et la certification des algorithmes adoptés font également partie des garde-fous devant être mis en place, et devraient devenir aussi répandus que pour les « produits cosmétiques, les voitures et les jouets ». À l'inverse, les « systèmes moins risqués » n'auront qu'à suivre des procédures de labellisation volontaire : « Les opérateurs économiques concernés se verraient alors attribuer un label de qualité pour leurs applications d'IA. »

Mais des critiques ont fait remarquer que la définition même du « risque » dans le document est à la fois circulaire et trop vague, ce qui permet à plusieurs systèmes ADM ayant un impact important de passer à travers les mailles du filet proposé[12].

Les commentaires[13] recueillis lors de la consultation publique, entre février et juin 2020, soulignent à quel point cette idée est controversée. En effet, 42,5 % des réponses conviennent que les « exigences obligatoires » devraient se limiter aux « applications d'IA à haut risque », tandis que 30,6 % doutent d'une telle limitation.

Par ailleurs, il n'existe aucune description d'un mécanisme clair pour faire respecter ces exigences, ni de description d'un processus permettant de s'en rapprocher.

Les conséquences sont immédiatement visibles pour les technologies biométriques, en particulier la reconnaissance faciale. Sur ce point, le livre blanc propose une distinction entre l'« authentification » biométrique, qui est considérée comme ne prêtant pas à controverse (par exemple, la reconnaissance faciale pour déverrouiller un smartphone), et l'« identification » biométrique à distance (comme le déploiement dans les lieux publics pour identifier les manifestants), qui pourrait causer de sérieux problèmes en matière de droits fondamentaux et de protection de la vie privée.

Seuls les cas relevant de cette dernière catégorie seraient problématiques dans le cadre du dispositif proposé par

12 « Pour donner deux exemples : VioGén, un système ADM permettant de prévoir les violences à caractère sexiste, et Ghostwriter, une application permettant de détecter la fraude aux examens, passeraient probablement entre les mailles du filet, de la réglementation, alors qu'elles comportent des risques énormes » (https://algorithmwatch.org/en/response-european-commission-ai-consultation/)

13 « Au total, 1 215 contributions ont été reçues, dont 352 au nom d'une entreprise ou d'une organisation/association commerciale, 406 de la part de citoyens (92 % de citoyens européens), 152 d'institutions universitaires/de recherche, et 73 provenant des pouvoirs publics. Les voix de la société civile étaient représentées par 160 répondants (dont 9 organisations de consommateurs, 129 ONG et 22 syndicats). 72 personnes ont contribué dans la catégorie « autres ». Les commentaires sont parvenus « du monde entier », y compris de pays tels que l'Inde, la Chine, le Japon, la Syrie, l'Irak, le Brésil, le Mexique, le Canada, les États-Unis et le Royaume-Uni. (extrait du rapport de synthèse de la consultation, accessible via le lien suivant : https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence)

l'UE. La FAQ du livre blanc explique qu'« il s'agit de la forme de reconnaissance faciale la plus intrusive, en principe interdite dans l'UE », à moins qu'un « intérêt public substantiel » ne justifie son déploiement.

Le document explicatif affirme que « l'autorisation de la reconnaissance faciale est actuellement l'exception », mais les conclusions de ce rapport contredisent manifestement cette opinion : la reconnaissance faciale semble devenir rapidement la norme. Une version préliminaire du livre blanc qui a fuité semble reconnaître l'urgence du problème, en incluant l'idée d'un moratoire de trois à cinq ans sur le recours à la reconnaissance faciale dans les lieux publics, jusqu'à ce qu'un moyen de les concilier avec les contrôles démocratiques puisse être trouvé, si tant est que cela soit possible.

Juste avant la publication officielle du livre blanc, même la commissaire européenne Margrethe Vestager a préconisé une « pause » de ces applications.

Cependant, immédiatement après l'appel de Mme Vestager, des responsables de la Commission ont ajouté que cette « pause » ne saurait empêcher les gouvernements nationaux d'utiliser la reconnaissance faciale conformément aux règles en vigueur. Pour finir, la version finale du livre blanc a évacué toute mention d'un moratoire et a appelé à un « grand débat européen sur les circonstances spécifiques, le cas échéant, qui pourraient justifier » son utilisation à des fins d'identification biométrique en direct. Parmi celles-ci, le livre blanc évoque notamment la justification, la proportionnalité, l'existence de mécanismes de protection démocratiques et le respect des droits fondamentaux.

Tout au long du document, les risques associés aux technologies basées sur l'IA sont plus généralement présentés comme des risques « potentiels », tandis que leurs avantages sont décrits comme bien réels et immédiats. Cela a conduit de nombreuses organisations de défense des droits fondamentaux[14] à déclarer que la teneur générale du livre blanc laisse entrevoir un revirement inquiétant des priorités de l'UE, faisant passer la compétitivité mondiale avant la protection des droits fondamentaux.

Certaines questions fondamentales sont toutefois abordées dans les documents : par exemple, l'interopérabilité de ces solutions et la création d'un réseau de centres de recherche axés sur les applications de l'IA visant l'« excellence » et le développement des compétences.

L'objectif est « d'attirer plus de 20 milliards d'euros d'investissements dans l'UE par an dans le domaine de l'IA au cours de la prochaine décennie ».

Un certain déterminisme technologique semble également affecter le livre blanc. « Il est essentiel, peut-on y lire, que l'administration publique, les hôpitaux, les services publics et de transport, les autorités de surveillance financière et d'autres secteurs d'intérêt public commencent rapidement à déployer des produits et des services reposant sur l'IA dans leurs activités. Un accent particulier sera mis sur les domaines des soins de santé et du transport, où la technologie est mûre pour un déploiement à grande échelle. »

Toutefois, il reste à voir si cette suggestion d'un déploiement hâtif des solutions ADM dans toutes les sphères de l'activité humaine est compatible avec les efforts de la Commission européenne visant à relever les défis structurels que posent les systèmes ADM en matière de droit et de justice.

/ Résolution du Parlement européen sur l'ADM et la protection des consommateurs

Une résolution, adoptée par le Parlement européen en février 2020, traite plus spécifiquement des systèmes ADM dans le contexte de la protection des consommateurs. Cette résolution fait remarquer à juste titre que « des avancées technologiques rapides ont lieu dans les domaines de l'intelligence artificielle, l'apprentissage automatique, les systèmes complexes fondés sur des algorithmes et les processus de prise de décision automatisés », et que « les applications, possibilités et défis découlant de ces technologies sont nombreux et concernent pratiquement tous les secteurs du marché intérieur ». Le texte souligne également la nécessité d'un « examen du cadre juridique actuel de l'UE » afin de « vérifier qu'il est à même de faire face à l'émergence de l'IA et de la prise de décision automatisée ».

Appelant à une « approche commune de l'Union européenne en matière de développement des processus de prise de décision automatisés », la résolution détaille plusieurs conditions que tout système de ce type devrait posséder

14 Parmi ceux-ci : Access Now (https://www.accessnow.org/cms/assets/uploads/2020/05/EU-white-paper-consultation_AccessNow_May2020.pdf), AI Now (https://ainowinstitute.org/ai-now-comments-to-eu-whitepaper-on-ai.pdf), EDRI (https://edri.org/can-the-eu-make-ai-trustworthy-no-but-they-can-make-it-just/) — et AlgorithmWatch (https://algorithmwatch.org/en/response-european-commission-ai-consultation/).

SI L'« IA » EST EFFECTIVEMENT UNE RÉVOLUTION NÉCESSITANT UN PAQUET DE MESURES LÉGISLATIVES DÉDIÉE, ÉLU·ES VEULENT AVOIR LEUR MOT À DIRE.

pour rester compatible avec les valeurs européennes. Les consommateur·trices devraient être « dûment informé[·e]s » sur l'impact qu'ont les algorithmes sur leur vie, et ils et elles devraient avoir accès à une personne ayant un pouvoir décisionnel afin que ces décisions puissent être vérifiées et corrigées si nécessaire. Ils et elles devraient également être informé·es « lorsque les prix des biens ou services ont été personnalisés sur la base d'une prise de décision automatisée et du profilage du comportement ».

En rappelant à la Commission européenne la nécessité d'une approche fondée sur les risques et soigneusement élaborée, la résolution précise que les mécanismes de protection doivent tenir compte du fait que les systèmes ADM « peuvent évoluer et agir d'une manière qui n'était pas envisagée lors de leur mise sur le marché », et que la responsabilité n'est pas toujours facile à attribuer lorsque le déploiement d'un système ADM entraîne un préjudice.

La résolution fait écho à l'article 22 du RGPD en notant qu'un sujet humain doit toujours être impliqué lorsque « des intérêts publics légitimes sont en jeu », et être responsable en dernier ressort des décisions prises dans le « domaine médical, juridique et comptable, ainsi que dans le secteur bancaire ». Une évaluation « correcte » des risques doit notamment précéder toute automatisation des services professionnels.

Enfin, la résolution liste des exigences détaillées en matière de qualité et de transparence dans la gouvernance des données : parmi celles-ci, « 'importance d'utiliser uniquement des données non faussées et de qualité pour améliorer les résultats des systèmes algorithmiques et renforcer la confiance et l'acceptation des consommateur[·trice]s » ; l'utilisation d'algorithmes « explicables et impartiaux » ; ainsi que la nécessité de « structures de réexamen » permettant aux personnes affectés de « réclamer un réexamen et une correction, par un être humain, des décisions automatisées qui sont définitives et permanentes ».

/ Tirer le maximum du « droit d'initiative » du Parlement européen

Dans son discours d'investiture, von der Leyen a clairement exprimé son soutien à un « droit d'initiative » pour le Parlement européen. « Lorsque cette Assemblée, statuant à la majorité de ses membres, adopte des résolutions demandant à la Commission de soumettre des propositions législatives, a-t-elle déclaré, je m'engage à répondre par un acte législatif dans le plein respect des principes de proportionnalité, de subsidiarité ainsi que de l'accord 'Mieux légiférer'. »

Si l'« IA » est effectivement une révolution nécessitant un paquet de mesures législatives dédiées, censé arriver au cours du premier trimestre 2021, les élus veulent avoir leur mot à dire. Cette volonté, associée à l'intention déclarée de von der Leyen de renforcer leurs capacités législatives, pourrait même déboucher sur ce que Politico a appelé un « moment parlementaire », des commissions parlementaires commençant à rédiger plusieurs rapports différents en conséquence.

Chaque rapport étudie des aspects spécifiques de l'automatisation dans la politique publique qui, bien qu'ils soient

destinés à façonner la législation à venir sur l'IA, sont aussi pertinents pour l'ADM.

Par exemple, dans son *cadre d'aspects éthiques en matière d'intelligence artificielle, de robotique et de technologies connexes*, la Commission des affaires juridiques appelle à la constitution d'une « Agence européenne de l'intelligence artificielle » et, dans le même temps, d'un réseau d'autorités nationales de surveillance dans chaque État membre pour s'assurer que les décisions prises en matière d'automatisation soient et demeurent éthiques.

Dans son rapport intitulé *Dr oits de propriété intellectuelle pour le développement des technologies liées à l'intelligence artificielle*, cette même commission expose sa vision pour l'avenir de la propriété intellectuelle et de l'automatisation. D'une part, le rapport préliminaire indique que « les méthodes mathématiques et les programmes d'ordinateurs ne sont pas brevetables en tant que tels, ils peuvent être intégrés à une invention technique susceptible d'être brevetable », tout en affirmant que, s'agissant de la transparence algorithmique, « la rétro-ingénierie constitue une exception au secret d'affaires ».

Le rapport va même jusqu'à envisager comment protéger « les créations techniques et artistiques générées par l'IA afin d'encourager cette forme de création », imaginant que « certaines œuvres générées par l'IA sont assimilables à des œuvres de l'esprit et que, dès lors, elles pourraient être protégées par le droit d'auteur ».

Enfin, dans un troisième document (*Intelligence artificielle et responsabilité civile*), la Commission détaille une « approche de gestion des risques » pour la responsabilité civile des technologies de l'IA. Selon ce document, « la partie qui est la mieux à même de contrôler et de gérer un risque lié à une technologie est tenue pour strictement responsable, en tant que point d'entrée unique pour les litiges ».

Des principes importants concernant l'utilisation de l'ADM dans le système de justice pénale se trouvent dans le rapport de la Commission des libertés civiles, de la justice et des affaires intérieures intitulé *L'intelligence artificielle en droit pénal et son utilisation par les autorités policières et judiciaires dans les affaires pénales*. Après avoir dressé une liste détaillée d'utilisations réelles et actuelles de l'« IA » – qui sont en réalité des systèmes ADM – par les forces de police[15], la Commission « estime qu'il est nécessaire de créer un régime clair et équitable pour l'attribution de la responsabilité juridique des conséquences négatives potentielles de ces technologies numériques avancées ».

Le rapport détaille ensuite certaines de ses dispositions : pas de décisions entièrement automatisées[16], une explication des algorithmes qui soit « intelligible pour les utilisateur[·trice]s », une « étude d'impact obligatoire sur les droits fondamentaux [...] de tout système d'IA à des fins répressives ou judiciaires » avant son déploiement ou son adoption, plus « un audit périodique obligatoire de tous les systèmes d'IA utilisés par les services répressifs et judiciaires pour tester et évaluer les systèmes algorithmiques une fois qu'ils sont en service ».

Le rapport préconise également un moratoire sur les technologies de reconnaissance faciale destinées aux forces de l'ordre, « jusqu'à ce que les normes techniques puissent être considérées comme pleinement respectueuses des droits fondamentaux, que les résultats obtenus soient non discriminatoires et que la confiance du public soit assurée quant à la nécessité et à la proportionnalité du déploiement de ces technologies ».

L'objectif est de renforcer à terme la transparence globale de ces systèmes, et de conseiller aux États membres d'avoir une « compréhension complète » des systèmes d'IA adoptés par les services répressifs et judiciaires, et – sur le modèle d'un « registre public » – de détailler « les types d'outils utilisés, les types de criminalité auxquels ils s'appliquent et les entreprises dont les outils sont utilisés ».

La Commission de la culture et de l'éducation et la Commission de la politique industrielle travaillent également sur leurs propres rapports à l'heure où nous écrivons ces lignes.

---

15 À la page 5, le rapport indique : « Les applications de l'IA utilisées par les services répressifs comprennent des applications telles que les technologies de reconnaissance faciale, la reconnaissance automatisée des plaques d'immatriculation, l'identification du locuteur, l'identification vocale, les technologies de lecture labiale, la surveillance auditive (par exemple, des algorithmes de détection des coups de feu), la recherche et l'analyse autonomes de bases de données identifiées, la prévision (police prédictive et analyse des zones sensibles), les outils de détection des comportements, les outils autonomes d'identification de la fraude financière et du financement du terrorisme, la surveillance des médias sociaux (scrapping et collecte de données pour les connexions de minage), les dispositifs de capture d'identité d'abonnés mobiles internationaux (IMSI) et les systèmes de surveillance automatisés intégrant différentes capacités de détection (comme la détection des battements de cœur et les caméras thermiques). »

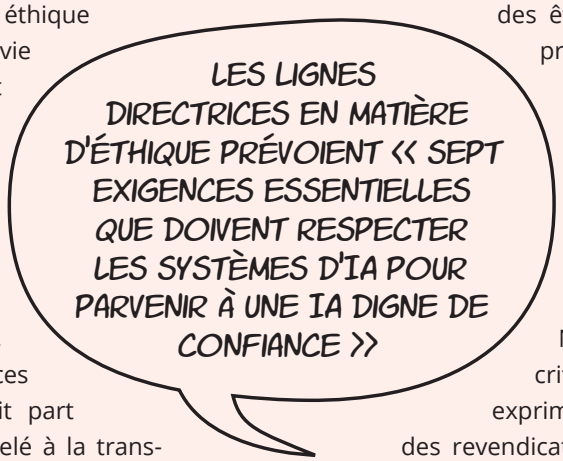
16 « Dans les contextes judiciaires et répressifs, la décision finale doit toujours être prise par un être humain. » (p. 6)

Toutes ces initiatives ont abouti à la création d'une Commission spéciale sur l'« intelligence artificielle à l'ère du numérique » (AIDA), le 18 juin 2020. Composée de 33 membres et créée pour une durée initiale de 12 mois, celle-ci « analysera la future incidence » de l'IA sur l'économie de l'UE, et « en particulier sur les compétences, l'emploi, les technologies de la finance, l'éducation, la santé, les transports, le tourisme, l'agriculture, l'environnement, la défense, l'industrie, l'énergie et l'administration en ligne ».

/ Groupe d'experts de haut niveau sur l'IA et AI Alliance

En 2018, le Groupe d'experts de haut niveau (GEHN) sur l'IA, un comité composé de 52 experts, a été constitué par la Commission européenne afin de soutenir la mise en œuvre de la stratégie européenne sur l'IA, d'identifier les principes à respecter pour parvenir à une « IA digne de confiance », et, en tant que comité directeur de l'AI Alliance, afin de créer une plateforme multipartite ouverte (qui comptait plus de 4 000 membres au moment de la rédaction de ce rapport) qui permettra d'élargir la contribution aux travaux du GEHN.

Après la publication de la première ébauche des *Lignes directrices en matière d'éthique pour une IA digne de confiance* en décembre 2018, document auquel ont répondu plus de 500 contributeurs, une version révisée a été publiée en avril 2019. Celle-ci propose une « approche centrée sur l'humain » pour parvenir à une IA légale, éthique et robuste tout au long du cycle de vie du système. Toutefois, il ne s'agit encore que d'un cadre volontaire sans recommandations concrètes et applicables en matière d'opérationnalisation, de mise en œuvre et d'application.



Les organisations de la société civile, de protection des consommateur·trices et de défense des droits ont fait part de leurs commentaires et ont appelé à la transposition des lignes directrices en droits concrets[17]. Par exemple, l'association à but non lucratif de défense des droits numériques Access Now, membre du GEHN, a demandé instamment à la Commission européenne de clarifier les modalités selon lesquelles les différentes parties prenantes pourront tester, appliquer, améliorer, approuver et faire respecter cette « IA digne de confiance », tout en reconnaissant la nécessité de déterminer les lignes rouges de l'Europe.

Dans un éditorial, deux autres membres du GEHN ont affirmé que le groupe avait « travaillé pendant un an et demi, tout cela pour que ses propositions détaillées soient essentiellement ignorées ou simplement mentionnées à titre indicatif » par la Commission européenne, qui a rédigé la version finale[18]. Ils ont également fait valoir que, puisque le groupe était initialement chargé d'identifier les risques et les « lignes rouges » de l'IA, les membres du groupe ont souligné que les systèmes d'armement autonomes, de notation des citoyens et d'identification automatisée des individus par la reconnaissance faciale étaient des applications de l'IA à éviter. Cependant, les représentants de l'industrie, qui dominent le comité[19], sont parvenus à faire supprimer ces principes avant que le projet ne soit publié.

Ce déséquilibre qui vise à mettre en évidence le potentiel de l'ADM plutôt que les risques qu'elle comporte s'observe également dans le deuxième rapport du groupe. Dans son document *Recommandations de politique et d'investissement pour une IA digne de confiance en Europe*, publié en juin 2019, le GEHN propose 33 recommandations visant à « guider l'IA digne de confiance vers la durabilité, la croissance et la compétitivité, ainsi que vers l'inclusion – tout en renforçant l'autonomie, les avantages et la protection des êtres humains ». Ce document est principalement un appel à encourager l'adoption et l'expansion de l'IA dans les secteurs privé et public en investissant dans des outils et des applications destinés à « aider les populations vulnérables » et à ne « laisser personne de côté ».

Néanmoins, et malgré toutes les critiques légitimes, ces deux directives expriment encore des préoccupations et des revendications critiques concernant les systèmes de prise de décision automatisée. Par exemple, les *Lignes directrices en matière d'éthique* prévoient « sept

17 Par exemple, le Bureau européen des unions de consommateurs (BEUC) : insérer lien vers la référence

18 Mark Coeckelbergh et Thomas Metzinger : Europe needs more guts when it comes to AI ethics, https://background.tagesspiegel.de/digitalisierung/europe-needs-more-guts-when-it-comes-to-ai-ethics

19 Le groupe était composé de 24 représentants d'entreprises, 17 universitaires, 5 organisations de la société civile et 6 autres membres, dont l'Agence des droits fondamentaux de l'Union européenne.

exigences essentielles que doivent respecter les systèmes d'IA pour parvenir à une IA digne de confiance ». Le document offre ensuite des conseils pour la mise en œuvre pratique de chacune de ces exigences : supervision et contrôle humains, robustesse et sécurité techniques, confidentialité et gouvernance des données, transparence, diversité, absence de discrimination et équité, bien-être sociétal et environnemental, et responsabilité.

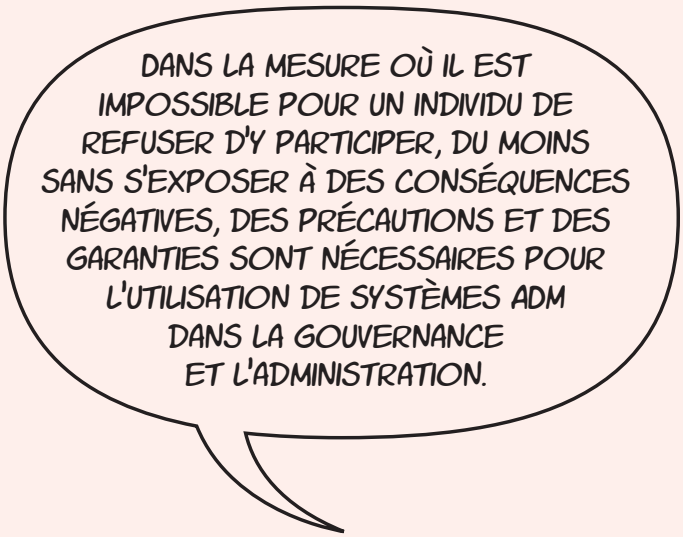
Ces directives prévoient également un projet pilote concret, appelé « liste d'évaluation pour une IA digne de confiance », qui vise à rendre ces principes généraux opérationnels. Le but est de faire adopter celle-ci « lors du développement, du déploiement ou de l'utilisation de systèmes d'IA » et de l'adapter « aux cas d'utilisation spécifiques dans lesquels le système est appliqué ».

La liste aborde de nombreuses questions liées au risque d'atteinte aux droits fondamentaux par les systèmes ADM. Celles-ci incluent l'absence de supervision et de contrôle humains, les problèmes de robustesse et de sécurité techniques, l'incapacité à éviter les traitements injustes ou à fournir un accès égal et universel à ces systèmes, et l'absence d'accès significatif aux données qui y sont introduites.

Contextuellement, la liste incluse dans les lignes directrices fournit des questions utiles pour aider ceux et celles qui déploient des systèmes ADM. Par exemple, elle préconise « une évaluation de l'incidence potentielle sur les droits fondamentaux » là où il pourrait y avoir un impact négatif sur ceux-ci. Elle demande également si « des mécanismes spécifiques de contrôle et de surveillance » ont été mis en place dans le cas de « systèmes d'autoapprentissage ou d'IA autonome », et si des processus existent « pour assurer la qualité et l'intégrité de vos données ».

Les remarques détaillées portent également sur des questions fondamentales pour les systèmes ADM, telles que leur transparence et leur explicabilité. Ces questions incluent « dans quelle mesure les décisions et donc les conséquences du système d'IA peuvent-elles être comprises ? » et « dans quelle mesure la décision du système influence-t-il les processus décisionnels de l'organisation ? ». Ces questions sont particulièrement pertinentes pour évaluer les risques posés par le déploiement de tels systèmes.

Afin d'éviter les partis pris et les effets discriminatoires, les lignes directrices préconisent « des processus de supervision permettant d'analyser et de traiter la finalité, les contraintes, les exigences et les décisions du système de



manière claire et transparente », tout en exigeant la participation des parties prenantes tout au long du processus de mise en œuvre des systèmes d'IA.

De plus, les recommandations en matière de politique et d'investissement prévoient de déterminer les lignes rouges par le biais d'un « dialogue institutionnalisé sur la politique d'IA avec les acteurs concernés », y compris des experts de la société civile. En outre, elles appellent à interdire la notation à grande échelle des individus au moyen de l'IA comme défini dans les lignes directrices en matière d'éthique, et à fixer des règles très claires et strictes pour la surveillance à des fins de sécurité nationale et d'autres fins prétendument d'intérêt public ou national. Cette interdiction inclurait les technologies d'identification biométrique et le profilage.

En ce qui concerne les systèmes de prise de décision automatisée, le document indique également que « pour parvenir à mettre en œuvre une IA digne de confiance, il sera essentiel de définir clairement si, quand et comment l'IA peut être utilisée aux fins de l'identification automatique de personnes », avertissant que « toute forme de notation des citoyen[·ne]s peut entraîner la perte de [leur] autonomie et mettre en péril le principe de non-discrimination», et « ne doit être utilisée que si elle se justifie clairement et lorsque les mesures sont proportionnées et équitables ». Il souligne en outre que « la transparence ne peut ni empêcher la discrimination, ni garantir l'équité ». Cela signifie qu'il doit être possible pour un·e citoyen·ne de ne pas participer à un mécanisme de notation, idéalement sans que cela ne lui porte préjudice.

D'une part, le document reconnaît que « certaines applications d'IA sont certes susceptibles d'apporter des avantages considérables aux individus et à la société, mais qu'elles peuvent également avoir des incidences négatives, y compris des incidences pouvant s'avérer difficiles à anti-

ciper, reconnaître ou mesurer (par exemple, en matière de démocratie, d'état de droit et de justice distributive, ou sur l'esprit humain lui-même) ». D'autre part, le groupe affirme cependant qu'il faut éviter toute réglementation inutilement prescriptive.

En juillet 2020, le GEHN sur l'IA a également présenté la version finale de sa liste d'évaluation pour une intelligence artificielle digne de confiance (ALTAI), compilée après un processus pilote ayant impliqué 350 partenaires.

La liste, qui est entièrement volontaire et sans aucune portée réglementaire, vise à traduire en actions les sept exigences énoncées dans les lignes directrices en matière d'éthique du GEHN sur l'IA. L'intention est de fournir à tous ceux qui souhaitent mettre en œuvre des solutions d'IA compatibles avec les valeurs de l'UE – par exemple, les concepteurs et développeurs de systèmes d'IA, les statisticien·nes, les responsables ou spécialistes des marchés publics et les responsables juridiques/de la conformité – une boîte à outils d'autoévaluation.

/ Conseil de l'Europe : comment protéger les droits fondamentaux dans le cadre de systèmes ADM

Parallèlement au Comité ad hoc sur l'intelligence artificielle (CAHAI), établi en septembre 2019, le Comité des ministres du Conseil de l'Europe[20] a publié un document substantiel et probant.

Envisagée comme un instrument normatif, sa *Recommandation aux États membres sur l'impact des systèmes algorithmiques sur les droits de l'homme* décrit[21] les « défis importants » qui se posent avec l'émergence et notre « recours croissant » à l'égard de ces systèmes, et qui sont pertinents « pour les sociétés démocratiques et l'État de droit ».

Ce texte, qui a fait l'objet d'une période de consultation publique avec des commentaires détaillés d'organisations de la société civile, va au-delà du livre blanc de la Commission européenne en ce qui concerne la protection des valeurs de l'UE et des droits fondamentaux.

La recommandation analyse de manière approfondie les effets et les configurations changeantes des systèmes algorithmiques (annexe A) en se concentrant sur toutes les étapes du processus qui entrent dans la fabrication d'un algorithme, c'est-à-dire l'acquisition, la conception, le développement et le déploiement continu.

Bien qu'elle suive généralement l'approche de l'« IA centrée sur l'humain » des lignes directrices du GEHN, la recommandation définit les « obligations des États » (annexe B) ainsi que les responsabilités des acteurs du secteur privé (annexe C). Par ailleurs, la recommandation ajoute des principes tels que l'« autodétermination informationnelle »[22], énumère des suggestions détaillées pour des mécanismes de responsabilisation et des recours efficaces, et exige des évaluations d'impact sur les droits fondamentaux.

Bien que le document reconnaisse clairement le « potentiel important des technologies numériques pour faire face aux défis sociétaux et pour favoriser une innovation et un développement économique socialement bénéfiques », il invite également à la prudence. Ceci afin de garantir que ces systèmes ne perpétuent pas délibérément ou accidentellement « les inégalités raciales, de genre et autres disparités au sein de la société et au sein de la population active, qui n'ont pas encore été éliminées de nos sociétés ».

Au contraire, les systèmes algorithmiques devraient être utilisés de manière proactive et sensible pour résoudre ces déséquilibres, et « accorder une attention particulière aux besoins et aux voix des groupes vulnérables ».

Mais surtout, la recommandation identifie le risque potentiellement plus élevé pour les droits fondamentaux lié à l'utilisation de systèmes algorithmiques par les États membres pour la prestation de services et de politiques

20 Le Conseil de l'Europe est à la fois « un organe gouvernemental où les approches nationales des problèmes européens sont discutées sur un pied d'égalité et un forum permettant de trouver des réponses collectives à ces défis ». Son travail inclut « les aspects politiques de l'intégration européenne, la sauvegarde des institutions démocratiques et de l'État de droit et la protection des droits de l'homme – en d'autres termes, tous les problèmes qui nécessitent des solutions paneuropéennes concertées ». Bien que les recommandations adressées aux gouvernements des États membres ne soient pas contraignantes, le Comité peut, dans certains cas, demander aux gouvernements de l'informer des suites données par eux à ces recommandations (Art. 15b des statuts). Les relations entre le Conseil de l'Europe et l'Union européenne sont définies dans (1) le *Recueil des textes régissant les relations entre le Conseil de l'Europe et l'Union européenne*, et (2) le *Mémorandum d'accord entre le Conseil de l'Europe et l'Union européenne*.

21 Sous la supervision du Comité directeur sur les médias et la société de l'information (CDMSI) et préparé par le Comité d'experts sur la dimension droits de l'Homme du traitement automatisé des données et de différentes formes d'intelligence artificielle (MSI-AUT).

22 « Les États doivent veiller à ce que la conception, le développement et le déploiement continu de systèmes algorithmiques permettent aux personnes d'être informées à l'avance du traitement des données (y compris de ses objectifs et de ses résultats possibles) et de contrôler leurs données, notamment grâce à l'interopérabilité », peut-on lire à la section 2.1 de l'annexe B.

publiques. Dans la mesure où il est impossible pour un individu de refuser d'y participer, du moins sans s'exposer à des conséquences négatives, des précautions et des garanties sont nécessaires pour l'utilisation de systèmes ADM dans la gouvernance et l'administration.

La recommandation aborde également les conflits et les difficultés potentielles découlant des partenariats public-privé dans un large éventail d'utilisations.

La recommandation exhorte ainsi les gouvernements des États membres à abandonner les processus et refuser d'utiliser un système ADM si « son opacité empêche toute supervision ou tout contrôle par un être humain » ou si les droits fondamentaux sont menacés ; et à déployer des systèmes ADM si et seulement si la transparence, la responsabilité, la légalité et la protection des droits humains peuvent être garantis « tout au long du déploiement ». Par ailleurs, le suivi et l'évaluation de ces systèmes doivent être « constants », « inclusifs et transparents », et comporter un dialogue avec toutes les parties prenantes concernées, ainsi qu'une analyse de l'impact environnemental et des autres externalités potentielles affectant « les populations et leurs environnements ».

À l'annexe A, le Conseil de l'Europe donne également une définition des algorithmes à « haut risque », dont les autres organismes pourront s'inspirer. Plus précisément, la recommandation explique que « l'expression « à haut risque » est employée en référence à l'utilisation de systèmes algorithmiques dans des processus ou des décisions susceptibles d'avoir des conséquences graves pour les individus, ou dans des situations où l'absence de solutions de rechange engendre une probabilité particulièrement élevée d'atteinte aux droits de l'homme, notamment en introduisant ou en amplifiant des inégalités distributives. ».

Le document, qui n'a pas nécessité l'unanimité des membres pour être adopté, est non contraignant.

/ Réglementation du contenu terroriste en ligne

Après une longue période de progrès laborieux, un règlement visant à empêcher la propagation de contenus terroristes en ligne a gagné du terrain en 2020. Si le règlement adopté devait inclure des outils automatisés et proactifs pour la reconnaissance et le retrait de contenus en ligne, ceux-ci relèveraient probablement de l'article 22 du RGPD.

Comme le souligne le Contrôleur européen de la protection des données (CEPD) : « étant donné que les outils automatisés, tels qu'envisagés par la proposition, pourraient non seulement conduire au retrait et à la conservation de contenus (et de données connexes) concernant la personne les ayant téléchargés, mais aussi, en fin de compte, à des poursuites pénales à son encontre, ces outils affecteraient de manière significative cette personne, auraient une incidence sur son droit à la liberté d'expression et présenteraient des risques importants pour ses droits et libertés », et, par conséquent, relèveraient de l'article 22(2).

En outre, et surtout, ce règlement nécessiterait des garanties plus substantielles que celles que la Commission prévoit actuellement. Comme l'explique le groupe de défense des droits numériques European Digital Rights (EDRi) : « Le règlement sur les contenus terroristes proposé doit être réformé en profondeur pour être à la hauteur des valeurs de l'Union et pour protéger les droits et libertés fondamentaux de ses citoyen·nes. »

Une première vague de critiques virulentes de la proposition initiale, émanant de groupes de la société civile et de commissions du Parlement européen (PE), y compris des avis et des analyses de l'Agence des droits fondamentaux de l'Union européenne (FRA), de l'EDRi, ainsi qu'un rapport critique conjoint de trois rapporteurs spéciaux des Nations unies, ont mis en évidence les menaces pesant sur le droit à la liberté d'expression et d'information, le droit à la liberté et au pluralisme des médias, la liberté d'entreprise et les droits à la vie privée et à la protection des données personnelles.

Les aspects critiqués comprennent notamment une définition trop vague du contenu terroriste, le champ d'application du règlement (qui couvre actuellement les contenus à caractère éducatif et journalistique), l'appel susmentionné à des « mesures proactives », le manque de supervision judiciaire efficace, l'insuffisance des obligations de signalement pour les services répressifs, et l'absence de garanties pour « les cas où il y a des motifs raisonnables de penser que les droits fondamentaux sont affectés » (EDRi 2019).

Le CEPD souligne que ces « garanties appropriées » devraient inclure le droit de bénéficier d'une intervention humaine et le droit d'obtenir une explication de la décision prise par des moyens automatisés (EDRi 2019).

Bien que les garanties suggérées ou exigées figurent maintenant dans le rapport préliminaire du Parlement euro-

péen sur la proposition, il reste à voir qui pourra retenir son souffle le plus longtemps avant le vote final. Lors des trilogues à huis clos entre le PE, la nouvelle CE et le Conseil de l'UE (qui a pris ses fonctions en octobre 2019), seules des modifications mineures sont encore possibles, selon un document qui a fuité.

Contrôle et réglementation

/ Premières décisions sur la conformité des systèmes ADM avec le RGPD

« Bien qu'il n'y ait pas eu de grand débat sur la reconnaissance faciale lors des négociations relatives au RGPD et à la directive "police-justice", la législation a été conçue de manière à pouvoir s'adapter au fil de temps en fonction de l'évaluation des technologies. [...] L'heure est venue pour l'UE, alors qu'elle débat du caractère éthique de l'IA et de la nécessité d'une réglementation, de déterminer si, le cas échéant, la technologie de reconnaissance faciale peut être autorisée dans une société démocratique. Si la réponse est "oui", alors seulement pourrons-nous nous pencher sur la question de savoir comment, et quelles garanties et responsabilités mettre en place. » – Wojciech Wiewiórowski, CEPD.

«Les dispositifs de reconnaissance faciale sont particulièrement intrusifs et présentent des risques majeurs d'atteinte à la vie privée et aux libertés individuelles des personnes concernées. » – (CNIL 2019)

Depuis la publication du dernier rapport *L'automatisation de la société* d'AlgorithmWatch, nous avons vu les premiers cas d'amendes et de décisions liées à des violations de la réglementation prononcées par les autorités nationales de protection des données (APD) en se basant sur le RGPD. Les études de cas suivantes illustrent toutefois les limites du RGPD dans la pratique en ce qui concerne l'article 22 relatif aux systèmes ADM, et montrent qu'il laisse le soin aux autorités de protection de la vie privée d'évaluer les affaires au cas par cas.

En Suède, un projet pilote de reconnaissance faciale, réalisé dans une classe pendant une période limitée, a été jugé contraire à plusieurs dispositions du règlement sur la protection des données (en particulier les articles 2(14) et 9(2) du RGPD) (Comité européen de la protection des données 2019).

Un cas similaire en France a été suspendu par la Commission nationale de l'informatique et des libertés (CNIL), qui a fait part de son inquiétude lorsque deux lycées ont envisagé d'introduire la technologie de reconnaissance faciale en partenariat avec la société américaine Cisco. L'avis de la CNIL est non contraignant et la procédure suit son cours[23].

L'autorisation préalable des autorités de protection des données n'est pas nécessaire pour réaliser de tels essais, le consentement des utilisateurs étant généralement considéré comme suffisant pour traiter leurs données biométriques. Et pourtant, ce n'était pas le cas en Suède, en raison d'un déséquilibre des pouvoirs entre le gestionnaire des données et les personnes concernées. Au lieu de cela, une évaluation d'impact adéquate et une consultation préalable avec l'APD ont été jugées nécessaires.

Le Contrôleur européen de la protection des données (CEPD) l'a confirmé :

« Le consentement doit être explicite ainsi que libre, informé et spécifique. Pourtant, il est clair qu'une personne ne peut pas refuser ni même accepter de se soumettre lorsqu'elle a besoin d'accéder à des espaces publics couverts par une surveillance par reconnaissance faciale. [...] Enfin, la conformité de la technologie avec des principes tels que la minimisatio n des données et l'obligation de protection des données dès la conception est plus que douteuse. La technologie de reconnaissance faciale n'a jamais été totalement précise, ce qui peut avoir de graves conséquences pour les personnes faussement identifiées, qu'il s'agisse de criminels ou autre. [...] Il serait toutefois malavisé de se focaliser uniquement sur les questions de protection de la vie privée. Il s'agit fondamentalement d'une question éthique dans une société démocratique. » (CEPD 2019)

Access Now a commenté :

« Alors que de plus en plus de projets de reconnaissance faciale se développent, nous constatons déjà que le RGPD offre des garanties utiles en matière de droits fondamentaux qui peuvent être opposées à la collecte et à l'utilisation illégales de données sensibles telles que les données biométriques. Mais le battage irresponsable et souvent infon-

23 Voir le chapitre sur la France (Kayalki 2019)

dé autour de l'efficacité de ces technologies et de l'intérêt économique sous-jacent risque de conduire les gouvernements centraux et locaux ainsi que les entreprises privées à tenter de circonvenir la loi. »-

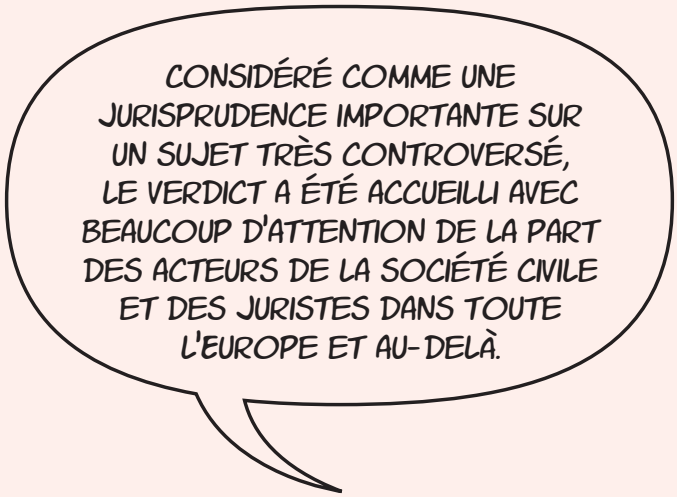
/ La reconnaissance faciale automatique utilisée par la police du sud du Pays de Galles jugée illégale

Au cours de l'année 2020, le Royaume-Uni a été témoin de la première application très médiatisée de la Law Enforcement Directive[24] concernant l'utilisation des technologies de reconnaissance faciale par la police dans les espaces publics. Considéré comme une jurisprudence importante sur un sujet très controversé, le verdict a été accueilli avec beaucoup d'attention de la part des acteurs de la société civile et des juristes dans toute l'Europe et au-delà[25].

L'affaire a été portée devant les tribunaux par Ed Bridges, un homme de 37 ans de Cardiff, qui a déclaré que son visage avait été scanné sans son consentement alors qu'il faisait ses courses de Noël en 2017, ainsi que lors d'une manifestation pacifique contre la prolifération des armes un an plus tard.

Le tribunal a d'abord affirmé la légalité de l'utilisation de la technologie de reconnaissance faciale automatisée (« RFA ») par la police du sud du Pays de Galles, la déclarant légale et proportionnée. Mais la décision a été contestée en appel par Liberty, un groupe de défense des droits civiques, et la Cour d'appel d'Angleterre et du Pays de Galles a décidé de casser le verdict du tribunal de première instance et de déclarer la technologie illégale le 11 août 2020.

En statuant contre la police du sud du Pays de Galles sur trois des cinq motifs, la Cour d'appel a constaté qu'il existait des « lacunes fondamentales » dans le cadre normatif existant autour de l'utilisation de la RFA, que son déploiement ne satisfaisait pas au principe de « proportionnalité » et, par ailleurs, qu'une évaluation adéquate de l'impact sur la protection des données (DPIA) n'avait pas été effectuée, omettant de multiples étapes cruciales.



Le tribunal n'a toutefois pas jugé que le système produisait des résultats discriminatoires sur la base du sexe ou de la race, car la police du sud du Pays de Galles n'avait pas recueilli suffisamment de données pour émettre un jugement à ce sujet[26]. Cependant, le tribunal a jugé nécessaire d'ajouter la remarque suivante : « La RFA étant une technologie nouvelle et controversée, nous espérons que toutes les forces de police qui ont l'intention de l'utiliser à l'avenir voudront bien s'assurer que tout ce qui peut raisonnablement être fait a été fait pour que le logiciel utilisé soit dépourvu de tout préjugé racial ou sexiste. »

Suite à l'arrêt de la cour, Liberty a exhorté la police du sud du Pays de Galles et d'autres forces de police de renoncer à l'utilisation des technologies de reconnaissance faciale.

L'ADM en pratique : gestion et surveillance des frontières

Alors que la Commission européenne et ses partenaires débattaient de la nécessité de réglementer ou d'interdire les technologies de reconnaissance faciale, des essais approfondis de ces systèmes étaient déjà en cours dans toute l'Europe.

24 La Law Enforcement Directive, en vigueur depuis mai 2018, « porte sur le traitement des données à caractère personnel par les gestionnaires de données à des «fins répressives» - qui ne relève pas du champ d'application du RGPD » https://www.dataprotection.ie/en/organisations/law-enforcement-directive

25 Décision rendue le 4 septembre 2019 par la Haute Cour de Cardiff dans l'affaire Bridges v. the South Wales Police (High Court of Justice 2019)

26 La police a affirmé qu'elle n'avait pas accès à la composition démographique de l'ensemble des données de formation pour l'algorithme adopté, « Neoface ». La Cour note que « le fait demeure, cependant, que la police du sud du Pays de Galles n'a jamais cherché à s'assurer, directement ou par le biais d'une vérification indépendante, que le logiciel en l'espèce ne présente pas un biais inacceptable fondé sur la race ou le sexe ».

L'UE FINANCE À HAUTEUR DE 8 MILLIONS D'EUROS CE PROJET VISANT À DÉVELOPPER « DES MÉTHODES AMÉLIORÉES POUR LA SURVEILLANCE DES FRONTIÈRES »

Cette section met en évidence un lien crucial et souvent passé sous silence entre la biométrie et les systèmes de gestion des frontières de l'UE, en montrant clairement comment des technologies susceptibles de produire des résultats discriminatoires risquent d'être appliquées aux personnes – par exemple, celles en situation de migration – qui souffrent déjà le plus de la discrimination.

/ Reconnaissance faciale et utilisation des données biométriques dans les politiques et les pratiques de l'UE

Au cours de l'année passée, la reconnaissance faciale et d'autres types de technologies d'identification biométrique ont suscité beaucoup d'intérêt de la part des gouvernements, de l'UE, de la société civile et des organisations de défense des droits, en particulier dans le domaine de l'application de la loi et de la gestion des frontières.

En 2019, la Commission européenne a chargé un consortium d'organismes publics de « dresser la carte de la situation actuelle de la reconnaissance faciale dans les enquêtes criminelles de tous les États membres de l'UE », dans le but d'« avancer vers un échange potentiel de données faciales ». Elle a demandé au cabinet de conseil Deloitte de réaliser une étude de faisabilité sur l'extension du système d'images faciales Prüm. Prüm est un système qui connecte les bases de données d'ADN, d'empreintes digitales et d'immatriculation des véhicules à l'échelle européenne pour permettre des recherches à l'échelle de plusieurs pays. La crainte est qu'une base de données paneuropéenne des visages de ses citoyens puisse être utilisée pour instaurer une surveillance omniprésente, injustifiée ou illégale.

/ Systèmes de gestion des frontières sans frontières

Comme rapporté dans la précédente édition du rapport *L'automatisation de la société*, la mise en œuvre d'un système de gestion des frontières global, interopérable et intelligent dans l'UE, initialement proposé en 2013 par la Commission, suit son cours. Bien que les nouveaux systèmes annoncés (EES, ETIAS[27], ECRIS-TCN[28]) n'entreront en fonctionnement qu'en 2022, le règlement Entry/Exit System (EES) a déjà introduit pour la première fois dans la législation européenne les images faciales comme éléments d'identification biométrique et l'utilisation de la technologie de reconnaissance faciale à des fins de vérification[29].

L'agence des droits fondamentaux de l'Union européenne (FRA) a confirmé ces changements : « Le traitement des images faciales devrait être introduit de manière plus systématique dans les systèmes informatiques utilisés à grande

27 ETIAS (EU Travel Information and Authorisation System) est le nouveau système d'exemption de visa pour la gestion des frontières de l'UE développé par eu-LISA. « Les informations soumises lors de la demande seront automatiquement traitées vis-à-vis des bases de données existantes de l'UE (Eurodac, SIS et VIS), des futurs systèmes EES et ECRIS-TCN, ainsi que des bases de données pertinentes d'Interpol. Cela permettra de vérifier à l'avance les risques potentiels en matière de sécurité, d'immigration illégale et de santé publique. » (ETIAS 2019)

28 Le système européen d'information sur les casiers judiciaires - ressortissants de pays tiers (ECRIS-TCN), qui doit être développé par eu-LISA, sera un système centralisé de recherche de concordance/non-concordance visant à compléter la base de données existante des casiers judiciaires de l'UE (ECRIS) sur les ressortissants de pays tiers condamnés dans l'Union européenne.

29 EES entrera en service au premier trimestre 2022, suivi par le système ETIAS d'ici la fin de l'année 2022 – qui devrait « changer la donne dans le domaine de la justice et des affaires intérieures (JAI) ».

échelle dans l'UE à des fins d'asile, d'immigration et de sécurité [...] une fois que les mesures juridiques et techniques nécessaires auront été prises. »

Selon Ana Maria Ruginis Andrei, de l'Agence européenne pour la gestion opérationnelle des systèmes d'information à grande échelle au sein de l'espace de liberté, de sécurité et de justice (EU-LISA), cette nouvelle architecture d'interopérabilité étendue a été « assemblée afin de forger le moteur idéal pour lutter avec succès contre les menaces à la sécurité intérieure, pour contrôler efficacement l'immigration et pour éliminer les angles morts en matière de gestion de l'identité ». En pratique, cela consiste à « conserver les empreintes digitales, les images faciales et autres données personnelles de 300 millions de ressortissants extraeuropéens, en fusionnant les données de cinq systèmes distincts » (Campbell 2020).

/ ETIAS : contrôles de sécurité automatisés aux frontières

Le système européen d'information et d'autorisation concernant les voyages (ETIAS), qui n'est pas encore entré en fonction à l'heure où nous rédigeons ce rapport, utilisera différentes bases de données pour automatiser les contrôles de sécurité numériques des voyageurs extraeuropéens (ceux qui n'ont pas besoin de visa) avant leur arrivée en Europe.

Ce système permettra de recueillir et d'analyser des données pour la « vérification avancée des risques potentiels en matière de sécurité ou d'immigration illégale » (ETIAS 2020). Son objectif est de « faciliter les contrôles aux frontières ; d'éviter les lenteurs bureaucratiques et les retards pour les voyageurs lorsqu'ils se présentent aux frontières ; et d'assurer une évaluation coordonnée et harmonisée des risques des ressortissants de pays tiers » (ETIAS 2020).

Ann-Charlotte Nygård, chef de l'Unité d'assistance technique et de renforcement des capacités de la FRA, distingue deux risques spécifiques concernant l'ETIAS : « Premièrement, l'utilisation de données qui pourraient entraîner une discrimination involontaire de certains groupes, par exemple si un demandeur est issu d'un groupe ethnique particulier présentant un risque élevé d'immigration illégale ; deuxièmement, l'évaluation du risque de sécurité sur la base de condamnations passées dans le pays d'origine. Certaines de ces condamnations antérieures pourraient être considérées comme déraisonnables par les Européens, comme les condamnations des personnes LGBT dans certains pays. Pour éviter cela, [...] les algorithmes doivent être vérifiés pour s'assurer qu'ils ne sont pas discriminatoires, et ce type de vérification doit impliquer des experts de différents domaines » (Nygård 2019).

/ iBorderCtrl : reconnaissance faciale et évaluation des risques aux frontières

iBorderCtrl était un projet impliquant les agences de sécurité de Hongrie, de Lettonie et de Grèce qui visait à « permettre des contrôles aux frontières plus rapides et plus approfondis des ressortissants de pays tiers franchissant les frontières terrestres des États de membres de l'UE ». iBorderCtrl utilisait une technologie de reconnaissance faciale, un détecteur de mensonges et un système de notation pour alerter la police aux frontières s'il jugeait qu'une personne était potentiellement dangereuse ou si son droit d'entrée paraissait douteux.

Le projet iBorderCtrl a pris fin en août 2019, et ses résultats – pour une mise en œuvre potentielle du système à l'échelle de l'UE – sont contradictoires.

Bien qu'il soit nécessaire de « déterminer jusqu'où le système ou une partie de celui-ci sera utilisé », la page « Résultats » du projet évoque « la possibilité d'intégrer les fonctionnalités similaires du nouveau système ETIAS et d'étendre les capacités liées à la procédure de passage des frontières là où les voyageurs se trouvent (bus, voiture, train, etc.). »

Toutefois, les modules auxquels il est fait référence ne sont pas spécifiés, et les outils ADM qui ont été testés n'ont pas fait l'objet d'une évaluation publique.

Dans le même temps, la page FAQ du projet confirme que le système qui a été testé n'est pas considéré comme « actuellement adapté au déploiement à la frontière [...] en raison de sa nature de prototype et de l'infrastructure technologique au niveau de l'UE ». Cela signifie que « des développements supplémentaires et une intégration au sein des systèmes existants de l'UE seraient nécessaires pour une utilisation par les autorités frontalières. »

En particulier, si le consortium iBorderCtrl a pu démontrer, en principe, le fonctionnement de cette technologie pour les contrôles aux frontières, il est également clair que les contraintes éthiques, légales et sociétales doivent être résolues avant tout déploiement à grande échelle.

/ Projets Horizon2020 associés

Plusieurs projets ultérieurs se sont focalisés sur le test et l'élaboration de nouveaux systèmes et de nouvelles technologies pour la gestion et la surveillance des frontières, dans le cadre du programme Horizon2020. Ceux-ci sont listés sur le site CORDIS de la Commission européenne, qui fournit des informations sur toutes les activités de recherche bénéficiant du soutien de l'UE qui s'y rapportent.

Le site montre que 38 projets sont actuellement en cours dans le cadre du programme/thème « H2020-EU.3.7.3. – Renforcer la sécurité par la gestion des frontières » de l'Union européenne. Son programme parent – « Sociétés sécurisées – Protection de la liberté et de la sécurité de l'Europe et de ses citoyens », qui est doté d'un budget global d'environ 1,7 milliard d'euros et finance 350 projets, prétend combattre « l'insécurité, qu'elle soit due à la criminalité, à la violence, au terrorisme, aux catastrophes naturelles ou d'origine humaine, aux cyberattaques ou aux atteintes à la vie privée, et à d'autres formes de troubles socioéconomiques qui touchent de plus en plus les citoyens » par le biais de projets visant principalement à développer de nouveaux systèmes technologiques basés sur l'IA et l'ADM.

Certains projets qui sont déjà achevés et/ou dont les applications sont déjà utilisées – par exemple, FastPass, ABC4EU, MOBILEPASS et EFFISEC – ont tous examiné les exigences en matière de « contrôles aux frontières automatisés (ABC) intégrés et interopérables », de systèmes d'identification et de portails « intelligents » à différents postes frontaliers.

TRESSPASS est un projet en cours qui a débuté en juin 2018 et prendra fin en novembre 2021. L'UE finance le projet à hauteur de près de 8 millions d'euros, et les coordinateurs d'iBorderCtrl (ainsi que de FLYSEC et XP-DITE) visent à « exploiter les résultats et les concepts mis en œuvre et testés » par iBorderCtrl et à « les développer pour créer une solution de sécurité multimodale pour le passage des frontières basée sur les risques, dans un cadre juridique et éthique solide » (Horizon2020 2019).

Le projet a pour objectif de transformer les contrôles de sécurité aux frontières en passant de l'ancienne stratégie, déclarée obsolète, « basée sur des règles » à une nouvelle stratégie « basée sur les risques ». Cette stratégie inclut l'application de technologies biométriques et de capteurs, d'un système de gestion basé sur les risques et de modèles pertinents pour évaluer l'identité, les possessions, les capacités et les intentions des individus. Elle vise à permettre des contrôles par le biais de « liens avec les systèmes existants et des bases de données externes telles que VIS/SIS/PNR » et recueille des données de toutes les sources susmentionnées à des fins de sécurité.

Un autre projet pilote, FOLDOUT, a débuté en septembre 2018 et se terminera en février 2022. L'UE finance à hauteur de 8 millions d'euros ce projet visant à développer « des méthodes améliorées pour la surveillance des frontières » afin de lutter contre l'immigration illégale, en mettant l'accent sur « la détection des personnes à travers un feuillage dense dans des climats extrêmes », en combinant « divers capteurs et technologies et en les fusionnant intelligemment en une plateforme de détection intelligente efficace et robuste » pour suggérer des scénarios de réaction. Des essais sont en cours en Grèce, en Finlande et en Guyane française.

MIRROR, pour *Migration-Related Risks caused by misconceptions of Opportunities and Requirement* (risques liés à la migration causés par des idées fausses sur les opportunités et les exigences), a débuté en juin 2019 et se poursuivra jusqu'en mai 2022. L'UE contribue à hauteur d'un peu plus de cinq millions d'euros à ce projet, qui vise à « comprendre comment l'Europe est perçue à l'étranger, détecter les divergences entre l'image et la réalité, repérer les cas de manipulation des médias et développer ses capacités à contrer ces idées fausses et les menaces pour la sécurité qui en découlent ». Sur la base d'une « analyse de la menace spécifique à la perception, le projet MIRROR combinera des méthodes d'analyse automatisée de textes, de supports multimédias et de réseaux sociaux avec des études empiriques » en vue de « développer des connaissances technologiques et des idées pratiques, [...] validées de manière approfondie avec les agences frontalières et les décideurs politiques, par exemple par le biais de projets pilotes. »

Parmi les autres projets déjà achevés, mais qui méritent d'être mentionnés, on peut citer Trusted Biometrics under Spoofing Attacks (TABULA RASA), qui a débuté en novembre 2010 et s'est achevé en avril 2014. Il visait à analyser « les faiblesses des logiciels d'identification biométrique dans le cadre de leur vulnérabilité au spoofing, diminuant ainsi l'efficacité des dispositifs biométriques » » Un autre projet, Bodega, qui a débuté en juin 2015 et s'est terminé en octobre 2018, a examiné comment mettre à profit « l'expertise du facteur humain » dans le cadre de « l'introduction de systèmes de contrôle aux frontières plus intelligents comme les portails automatisés et les systèmes de libre-service basés sur la biométrie ».

Références :

Access Now (2019) : Comments on the draft recommendation of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems https://www.accessnow.org/cms/assets/uploads/2019/10/Submission-on-CoE-recommendation-on-the-human-rights-impacts-of-algorithmic-systems-21.pdf

AlgorithmWatch (2020) : Our response to the European Commission's consultation on AI https://algorithmwatch.org/en/response-european-commission-ai-consultation/

Campbell, Zach/Jones, Chris (2020) : Leaked Reports Show EU Police Are Planning a Pan-European Network of Facial Recognition Databases https://theintercept.com/2020/02/21/eu-facial-recognition-database/

CNIL (2019) : French privacy regulator finds facial recognition gates in schools illegal https://www.biometricupdate.com/201910/french-privacy-regulator-finds-facial-recognition-gates-in-schools-illegal

Coeckelbergh, Mark/Metzinger, Thomas(2020) : Europe needs more guts when it comes to AI ethics https://background.tagesspiegel.de/digitalisierung/europe-needs-more-guts-when-it-comes-to-ai-ethics

Comité de protection des données (2020) : Law enforcement directive https://www.dataprotection.ie/en/organisations/law-enforcement-directive

Comité des ministres (2020) : Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts fo algorithmic systems https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154

Comité européen de la protection des données (2019) : Facial recognition in school renders Sweden's first GDPR fine https://edpb.europa.eu/news/national-news/2019/facial-recognition-school-renders-swedens-first-gdpr-fine_en

Commissaire aux droits de l'homme (2020) : Unboxing artificial intelligence: 10 steps to protect human rights https://www.coe.int/en/web/commissioner/-/unboxing-artificial-intelligence-10-steps-to-protect-human-rights

Commission des affaires juridiques (2020) : Rapport préliminaire : With recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies https://www.europarl.europa.eu/doceo/document/JURI-PR-650508_EN.pdf

Commission des affaires juridiques (2020) : Artificial Intelligence and Civil Liability https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf

Commission des affaires juridiques (2020) : Rapport préliminaire : On intellectual property rights for the development of artificial intelligence technologies https://www.europarl.europa.eu/doceo/document/JURI-PR-650527_EN.pdf

Commission des libertés civiles, de la justice et des affaires intérieures (2020) : Rapport préliminaire : On artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters https://www.europarl.europa.eu/doceo/document/LIBE-PR-652625_EN.pdf

Commission européenne (2018) : White paper: On Artificial Intelligence - A European approach to excellence and trust https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Commission européenne (2020) : A European data strategy https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

Commission européenne (2020) : Shaping Europe's digital future – Questions and Answers https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_264

Commission européenne (2020) : White Paper on Artificial Intelligence: Public consultation towards a European approach for excellence and trust https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence

Commission européenne (2018) : Security Union: A European Travel Information and Authorisation System - Questions & Answers https://ec.europa.eu/commission/presscorner/detail/en/MEMO_18_4362

Contrôleur européen de la protection des données (2019) : Facial recognition: A solution in search of a problem? https://edps.europa.eu/press-publications/press-news/blog/facial-recognition-solution-search-problem_de

Delcker, Janosch(2020) : Decoded: Drawing the battle lines — Ghost work — Parliament's moment https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-drawing-the-battle-lines-ghost-work-parliaments-moment/?utm_source=POLITICO.EU&utm_campaign=5a7d137f82-EMAIL_CAMPAIGN_2020_09_09_08_59&utm_medium=email&utm_term=0_10959edeb5-5a7d137f82-190607820

EDRi (2019) : FRA and EDPS: Terrorist Content Regulation requires improvement for fundamental rights https://edri.org/our-work/fra-edps-terrorist-content-regulation-fundamental-rights-terreg/

ETIAS (2020) : European Travel Information and Authorisation System (ETIAS) https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders/etias_en

ETIAS (2019) : European Travel Information and Authorisation System (ETIAS) https://www.eulisa.europa.eu/Publications/Information%20Material/Leaflet%20ETIAS.pdf

Groupe d'experts de haut niveau sur l'intelligence artificielle (2020) : Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Horizon2020 (2019) : robusT Risk basEd Screening and alert System for PASSengers and luggage https://cordis.europa.eu/project/id/787120/reporting

High Court of Justice (2019) : Bridges v. the South Wales Police https://www.judiciary.uk/wp-content/uploads/2019/09/bridges-swp-judgment-Final03-09-19-1.pdf

Hunton Andrew Kurth (2020) : UK Court of Appeal Finds Automated Facial Recognition Technology Unlawful in Bridges v South Wales Police https://www.huntonprivacyblog.com/2020/08/12/uk-court-of-appeal-finds-automated-facial-recognition-technology-unlawful-in-bridges-v-south-wales-police/

Kayalki, Laura (2019) : French privacy watchdog says facial recognition trial in high schools is illegal https://www.politico.eu/article/french-privacy-watchdog-says-facial-recognition-trial-in-high-schools-is-illegal-privacy/

Kayser-Bril, Nicolas (2020) : EU Commission publishes white paper on AI regulation 20 days before schedule, forgets regulation https://algorithmwatch.org/en/story/ai-white-paper/

Leyen, Ursula von der (2019) : A Union that strives for more - My agenda for Europe https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf

Leyen, Ursula von der (2020) : Paving the road to a technologically sovereign Europe https://delano.lu/d/detail/news/paving-road-technologically-sovereign-europe/209497

Leyen, Ursula von der (2020) : Shaping Europe's digital future https://twitter.com/eu_commission/status/1230216379002970112?s=11

Leyen, Ursula von der (2019) : Opening Statement in the European Parliament Plenary Session by Ursula von der Leyen, Candidate for President of the European Commission https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_19_4230

Nygård, (2019) : The New Information Architecture as a Driver for Efficiency and Effectiveness in Internal Security https://www.eulisa.europa.eu/Publications/Reports/eu-LISA%20Annual%20Conference%20Report%202019.pdf

Parlement européen (2020) : Artificial intelligence: EU must ensure a fair and safe use for consumers https://www.europarl.europa.eu/news/en/press-room/20200120IPR70622/artificial-intelligence-eu-must-ensure-a-fair-and-safe-use-for-consumers

Parlement européen (2020) : On automated decision-making processes: ensuring consumer protection and free movement of goods and services https://www.europarl.europa.eu/doceo/document/B-9-2020-0094_EN.pdf

Police du sud du Pays de Galles (2020) : Automated Facial Recognition https://afr.south-wales.police.uk/

RGPD (Art 22) : Automated individual decision-making, including profiling https://gdpr-info.eu/art-22-gdpr/

Sabbagh, Dan (2020) : South Wales police lose landmark facial recognition case https://www.theguardian.com/technology/2020/aug/11/south-wales-police-lose-landmark-facial-recognition-case

Valero, Jorge (2020) : Vestager: Facial recognition tech breaches EU data protection rules https://www.euractiv.com/section/digital/news/vestager-facial-recognition-tech-breaches-eu-data-protection-rules/

SUISSE

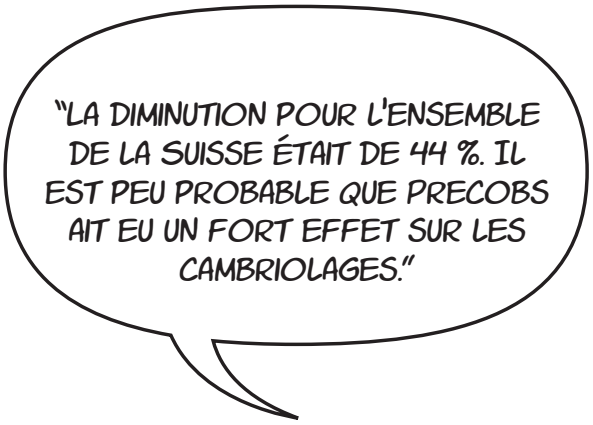
ARTICLE

PAGE 93

RECHERCHE

PAGE 97

Pour en savoir plus, lisez le chapitre de la recherche sous la rubrique « Diagnostic et traitement du cancer »



Police algorithmique : l'outillage des polices suisses baigne dans l'opacité

Un examen de trois systèmes automatisés utilisés par la police et la justice en Suisse pose de sérieuses questions. L'opacité qui les entoure est telle qu'il est impossible d'estimer l'ampleur du problème.

Par Nicolas Kayser-Bril

En Suisse, les autorités de police et de justice utilisent, selon un décompte, plus de vingt systèmes automatisés différents pour estimer ou prédire les comportements inappropriés. La police et la justice sont des compétences principalement cantonales. Chaque *canton* peut déployer ses propres systèmes.

Sur la base d'une série de reportages publiés par la SRF en 2018, nous en avons examiné trois.

/ Prédire les cambriolages

Precobs est utilisé en Suisse depuis 2013. L'outil est vendu par une société allemande qui ne fait aucun mystère de sa filiation avec « Minority Report », une histoire de science-fiction où les « precogs » prédisent certains crimes avant qu'ils ne se produisent. (L'intrigue tourne autour des échecs fréquents des précogs et leur dissimulation par la police.)

Precobs tente de prédire les cambriolages à partir de données historiques, en partant de l'hypothèse que les cambrioleurs opèrent souvent dans de petites zones. Selon cette théorie, si plusieurs cambriolages sont détectés dans un quartier, la police peut y mettre fin en patrouillant le quartier plus souvent.

Quatre cantons utilisent Precobs: Zurich, Zoug, Argovie et Bâle-Campagne, qui représentent près d'un tiers de la population suisse. Les cambriolages y ont chuté de façon spectaculaire depuis le milieu des années 2010. La police d'Argovie s'est même plainte en avril 2020 qu'il y avait désormais trop peu de cambriolages pour que Precobs puisse être utilisé.

Mais le nombre de cambriolages a diminué dans tous les cantons suisses. Les trois cantons qui utilisent Precobs sont loin d'être ceux où la diminution est la plus forte. Entre les années 2012-2014 (où les cambriolages étaient à leur apogée) et les années 2017-2019 (où Precobs était utilisé dans les quatre cantons), le nombre de cambriolages a diminué dans tous les cantons, pas seulement dans les trois qui utilisaient le logiciel. Les cambriolages ont diminués de 40 % à Zurich et en Argovie, et de 46 % à Bâle-Campagne et à Zoug. La diminution pour l'ensemble de la Suisse était de 44 %. Il est peu probable que Precobs ait eu un fort effet sur les cambriolages.

Un rapport de 2019 de l'Université de Hambourg n'a trouvé aucune preuve de l'efficacité des solutions de police algorithmique, y compris Precobs. Aucun document public n'a pu être trouvé mentionnant le coût du système pour les autorités suisses, mais la ville de Munich a payé 100 000 € pour l'installation de Precobs – frais de fonctionnement non compris.

/ Prédire la violence contre les femmes

Six cantons (Glaris, Lucerne, Schaffhouse, Soleure, Thurgovie et Zurich) utilisent le système Dyrias-Intimpartner pour calculer la probabilité qu'un homme agresse sa partenaire. Dyrias signifie « système dynamique pour l'analyse des risques ». Le logiciel est également conçu et vendu par une société allemande.

Selon un article publié en 2018 par la SRF, Dyrias demande aux policiers de répondre par « oui » ou par « non » à 39 questions concernant un suspect. L'outil produit ensuite un score sur une échelle de 1 à 5, allant de « inoffensif » à « dangereux ». Bien que le nombre total de personnes testées par l'outil soit inconnu, un décompte effectué par la SRF a montré que 3 000 personnes avaient été qualifiées de « dangereuses » en 2018 (mais cette appellation n'est pas forcément issue d'une utilisation de Dyrias).

La société qui produit Dyrias affirme que le logiciel identifie correctement 8 personnes potentiellement dangereuses sur 10. Cependant, une autre étude a examiné les faux positifs, des individus qualifiés de dangereux qui étaient en fait inoffensifs, et a révélé que 6 personnes sur 10 signalées par le logiciel comme dangereuses auraient dû être qualifiées d'inoffensives. En d'autres termes, Dyrias ne peut se vanter de bons résultats que parce qu'il ne prend aucun risque et colle très généreusement l'étiquette « dangereux » aux individus. (Le fabricant de Dyrias conteste ces résultats.)

Même si les performances de l'outil étaient améliorées, ses effets resteraient impossibles à évaluer. Justyna Gospodinov, codirectrice de BIF-Frauenberatung, une organisation qui soutient les victimes de violence domestique, a déclaré à AlgorithmWatch que, même si la coopération avec la police s'améliore et que l'évaluation systématique des risques est une bonne chose, elle ne pouvait rien dire sur Dyrias. Lorsque BIF-Frauenberatung accueille une nouvelle femme victime de violence, ils ne savent pas si le logiciel a été utilisé ou non, dit-elle.

/ Prédire la récidive

Depuis 2018, toutes les autorités judiciaires des cantons germanophones utilisent ROS (acronyme de « Risikoorientierter Sanktionenvollzug » ou exécution des peines de prison en fonction du risque). L'outil étiquette les détenus A lorsqu'ils ne présentent aucun risque de récidive, B lorsqu'ils pourraient commettre une nouvelle infraction ou C lorsqu'ils pourraient commettre un crime violent. Les détenus peuvent être testés plusieurs fois mais, lors de tests ultérieurs, ils ne peuvent que régresser de la catégorie A à B ou C, et non progresser de C ou B vers A.

Un reportage de la SRF a révélé que seulement un quart des détenus de la catégorie C avaient commis un nouveau crime après leur libération (un taux de faux positifs de 75%), et que seulement un sur cinq de ceux qui avaient commis un nouveau crime était en catégorie C (un taux de faux négatifs de 80%), d'après une étude réalisée en 2013 par l'Université de Zurich. Une nouvelle version de l'outil a été publiée en 2017 mais n'a pas encore été auditée.

Les cantons francophones et italophones travaillent sur une alternative au ROS, qui devrait être déployée en 2022. S'il conserve les 3 catégories, leur outil intégrera les résultats d'entretiens avec les détenus évalués.

/ Mission impossible

Les spécialistes des sciences sociales peuvent prédire le futur pour des groupes de personnes. En 2010, l'office suisse des statistiques a prédit que la population résidente de la Suisse atteindrait 8,5 millions d'ici 2020. Dix ans plus tard, la population suisse est estimée à 8,6 millions. Malgré ce bon résultat au niveau de la population dans son ensemble, aucun·e scientifique n'essaierait de prédire la date du décès d'un individu donné. La vie est tout simplement trop compliquée.

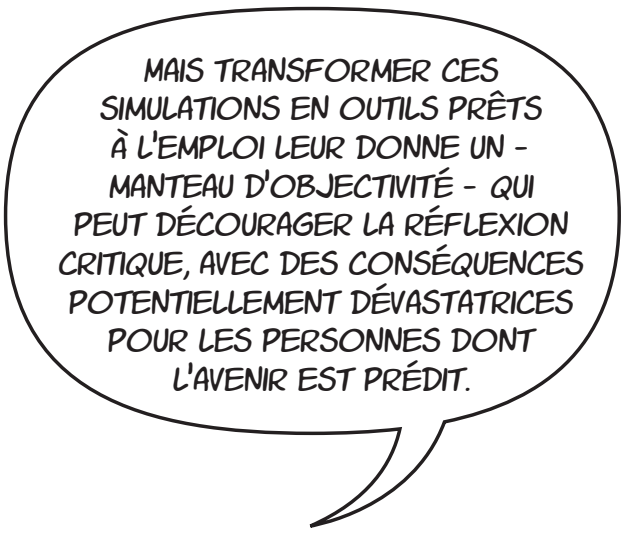
À cet égard, la démographie n'est pas différente de la criminologie. Malgré les affirmations de certains commerciaux, il est probable que la prédiction du comportement individuel soit impossible. En 2017, un groupe de scientifiques voulut en avoir le cœur net. Ils et elles ont demandé à 160 équipes de chercheur·ses de prédire les performances scolaires, la probabilité d'être expulsé de leur domicile et quatre autres résultats pour des milliers d'adolescent·es, sur la base de données collectées depuis leur naissance. Des milliers de points de données étaient disponibles pour chaque enfant. Les résultats, publiés en avril 2020, invitent à la modestie. Non seulement aucune équipe n'a pu prédire un résultat avec une précision quelconque, mais celles qui ont utilisé l'intelligence artificielle n'ont pas fait mieux que les équipes qui n'ont utilisé que quelques variables et des modèles statistiques basiques.

Moritz Büchi, chercheur à l'Université de Zurich, est le seul Suisse à avoir participé à cette expérience. Contacté par e-mail, il explique que, bien que la criminalité ne faisait pas partie des résultats examinés, les informations tirées de l'expérience s'appliquent probablement aux prédictions de criminalité. Toutes les prédiction ne doivent pas être abandonnées pour autant, nous dit M. Büchi. Mais transformer ces simulations en outils prêts à l'emploi leur donne un „manteau d'objectivité" qui peut décourager la réflexion critique, avec des conséquences potentiellement dévastatrices pour les personnes dont l'avenir est prédit.

Precobs, qui ne tente pas de prédire le comportement d'individus spécifiques, n'entre pas dans la même catégorie, a-t-il ajouté. Augmenter le nombre de patrouille pourrait effectivement avoir un effet dissuasif sur les cambrioleurs. Cependant, la détection des zones à risque repose sur des données historiques. Cela pourrait créer un cercle vicieux où les quartier qui ont concentré l'attention des policiers dans le passé seraient automatiquement indiqués par le logiciel comme étant des zones à risques dans le futur.

/ Effets dissuasifs

Malgré l'absence de preuve de leur efficacité, et malgré des preuves de la quasi-impossibilité de prédire les comportements individuels, les autorités suisses continuent d'utiliser des outils supposés prédire les comportements individuels.



Leur popularité est sans doute en partie due à leur opacité. Il existe très peu d'informations publiques sur Precobs, Dyrias et ROS. Les personnes sur lesquelles ces outils sont utilisés, qui sont en général pauvres, disposent rarement des ressources financières nécessaires pour remettre en cause ces systèmes automatisés. Leurs avocat·es se concentrent le plus souvent sur la vérification des faits allégués par le parquet.

Timo Grossenbacher, le journaliste qui a enquêté sur ROS et Dyrias pour la SRF en 2018, a déclaré à AlgorithmWatch que trouver des personnes affectées par ces systèmes était « presque impossible ». Ce n'est pas faute de cas: le ROS à lui seul est utilisé sur des milliers de détenus chaque année. Mais son opacité empêche les journalistes et la société civile d'en analyser les résultats et les effets.

Sans plus de transparence, ces systèmes pourraient avoir un « effet dissuasif » sur la société, selon M. Büchi de l'Université de Zurich. « Ces systèmes pourraient dissuader les gens d'exercer leurs droits et pourraient les amener à modifier leurs comportements », écrit-il. « C'est une forme d'obéissance anticipée. Conscients de la possibilité de se faire (injustement) attraper par ces algorithmes, les gens peuvent avoir tendance à accroître la conformité avec les normes sociétales perçues. L'expression personnelle et les modes de vie alternatifs pourraient en pâtir. »

Suisse

PAR LA PROFESSEURE NADJA BRAUN BINDER, DOCTEUR EN DROIT, ET CATHERINE EGLI

Contexte

La Suisse est un pays résolument fédéral, avec une séparation des pouvoirs prononcée. Par conséquent, les innovations techniques dans le secteur public sont souvent d'abord développées dans les cantons. L'introduction d'un système d'identité électronique (eID) en est un exemple. Au niveau fédéral, le processus législatif nécessaire à la mise en œuvre de l'eID n'est pas encore achevé, alors qu'un système d'identité électronique officiellement confirmé est déjà en vigueur dans un canton. En 2017, dans le cadre de la stratégie cantonale suisse de gouvernement électronique, le canton de Schaffhouse est devenu le premier à mettre en place une identité numérique pour ses résidents. Grâce à cette identité électronique, les citoyen·nes peuvent, entre autres, déposer une demande de permis de pêche en ligne, calculer les obligations fiscales d'un bénéfice immobilier ou d'une déclaration de capital, ou demander une prolongation pour leur déclaration d'impôt. Par ailleurs, un compte adulte peut être ouvert auprès de l'Autorité de protection des enfants et des adultes, et les médecins peuvent faire une demande de crédit pour les patients hospitalisés en dehors de leur district. Un autre exemple, qui a débuté dans le cadre d'un projet pilote dans le même canton en septembre 2019, permet aux résident·es de commander des extraits (via un smartphone) auprès du registre des poursuites. Ces services sont en constante expansion (Schaffhauser 2020). Bien que l'eID lui-même ne soit pas un processus d'ADM, il est une condition préalable essentielle pour accéder aux services gouvernementaux numériques, et pourrait également faciliter l'accès aux procédures automatisées, par exemple dans le domaine fiscal. Le fait qu'un seul canton ait davantage progressé sur cette voie que le gouvernement suisse est un phénomène typique pour la Suisse.

La démocratie directe est un autre élément caractéristique de l'État suisse. Par exemple, le processus législatif relatif à un eID national n'est pas encore achevé parce qu'un référendum va être organisé sur le projet de loi parlementaire correspondant (eID - Referendum 2020). Ceux qui ont demandé le référendum ne s'opposent pas fondamentalement à un eID officiel, mais ils souhaitent empêcher les entreprises privées de délivrer l'eID et de mettre la main sur des données sensibles à caractère privé.

Un autre élément dont il faut tenir compte est la situation économique favorable de la Suisse. Celle-ci permet de réaliser de grands progrès dans des domaines particuliers, comme les décisions automatisées utilisées en médecine, et dans de nombreux secteurs de la recherche. Bien qu'il n'existe pas de stratégie d'IA ou d'ADM centralisée en Suisse, en raison de la structure fédérale distincte et de la répartition des responsabilités entre les différents ministères au niveau fédéral, la recherche sectorielle réalisée est compétitive à l'échelle mondiale.

Catalogue d'applications de l'ADM

/ Diagnostic et traitement du cancer

À l'heure actuelle, la Suisse étudie l'utilisation de la prise de décision automatisée en médecine, et c'est pourquoi l'ADM a été davantage développée dans le secteur des soins de santé que dans d'autres domaines. Aujourd'hui, plus de 200 types de cancer différents sont recensés et près de 120 médicaments sont disponibles pour les traiter. Chaque année, de nombreux diagnostics de cancer sont établis et, comme chaque tumeur a son propre profil particulier,

avec des mutations génétiques qui favorisent la croissance de la tumeur, cela pose problème aux oncologues. Une fois qu'ils ont établi un diagnostic et déterminé la mutation génétique éventuelle, ceux-ci doivent étudier une littérature médicale toujours plus abondante afin de choisir le traitement le plus efficace. C'est la raison pour laquelle les Hôpitaux universitaires de Genève sont les premiers hôpitaux d'Europe à utiliser l'outil d'IBM Watson Health, Watson for Genomics, pour déterminer les meilleures options thérapeutiques et proposer des traitements aux patients cancéreux. Les médecins doivent encore examiner les mutations génétiques et décrire où et combien d'entre elles sont présentes, mais Watson for Genomics peut utiliser ces informations pour effectuer des recherches dans une base de données d'environ trois millions de publications. Le programme crée ensuite un rapport classant les altérations génétiques de la tumeur du patient et présentant les thérapies et les essais cliniques pertinents associés. Jusqu'à présent, les oncologues devaient faire ce travail eux et elles-mêmes, au risque de passer à côté d'une éventuelle méthode de traitement. Aujourd'hui, le programme informatique peut se charger de la recherche, mais les oncologues doivent encore vérifier soigneusement la liste de publications générée par le programme, pour ensuite décider d'une méthode de traitement. Ainsi, Watson for Genomics fait gagner beaucoup de temps lors de l'analyse et fournit des informations complémentaires importantes. À Genève, le rapport produit par cet outil d'ADM est utilisé lors de la préparation du Tumor Board, ou réunion de concertation pluridisciplinaire (RCP) en cancérologie, dans le cadre de laquelle les médecins prennent connaissance des traitements proposés par Watson for Genomics et en discutent en séance plénière afin d'élaborer conjointement une stratégie de traitement pour chaque patient (Schwerzmann/Arroyo 2019).

L'ADM est également utilisée à l'hôpital universitaire de Zurich, car elle se prête particulièrement bien aux tâches répétitives, principalement en radiologie et en pathologie, et est ainsi utilisée pour calculer la densité mammaire. Lors d'une mammographie, un algorithme informatique analyse automatiquement les images radiologiques et classe le tissu mammaire dans les catégories A, B, C ou D (une grille d'analyse des risques reconnue au niveau international). En analysant le risque en fonction de la densité mammaire, l'algorithme aide grandement les médecins à évaluer le risque de cancer du sein chez une patiente, puisque la densité mammaire est l'un des facteurs de risque les plus importants du cancer du sein. Cette utilisation de l'ADM à des fins d'analyse d'images médicales est désormais une pratique courante à l'hôpital universitaire de Zurich. En outre, des recherches sont en cours en vue de mettre au point des algorithmes avancés pour l'interprétation des images échographiques (Lindner 2019).

Cela étant, plus d'un tiers des cancers du sein ne sont pas détectés lors d'un examen de dépistage par mammographie. C'est pourquoi des recherches sont en cours pour déterminer comment l'ADM peut aider à l'interprétation des images échographiques. L'interprétation des images échographiques mammaires contraste fortement avec la mammographie numérique standard – qui dépend dans une large mesure de l'observateur et nécessite des radiologues bien formé·es et expérimenté·es. C'est pourquoi un programme de l'hôpital universitaire de Zurich a étudié comment l'ADM pouvait faciliter et normaliser l'imagerie échographique. Ce faisant, le processus humain de prise de décision est simulé en fonction du système d'imagerie mammaire, de rapports et de données. Cette technique est très précise et, à l'avenir, cet algorithme pourrait être utilisé pour imiter le processus humain de prise de décision et devenir la norme pour la détection, la mise en évidence et la classification des lésions mammaires aux ultrasons (Ciritisis a.o. 2019 p. 5458–5468).

/ Chatbot à l'Institut d'assurance sociale

Pour simplifier et assister les communications administratives, certains cantons utilisent également des « chatbots », ou agents conversationnels. Un chatbot a notamment été testé en 2018 au « Sozialversicherungsanstalt des Kantons St. Gallens » (Institut d'assurance sociale du canton de Saint-Gall, SVA St. Gallen). Le SVA St. Gallen est un centre d'excellence pour tous types d'assurances sociales, et inclut notamment un système de réduction de prime pour l'assurance maladie. L'assurance maladie est obligatoire en Suisse et couvre tous·tes les résident·es en cas de maladie, de maternité et d'accident, offrant à tous·tes la même gamme de prestations. Elle est financée par les cotisations (primes) des citoyens. Les primes varient selon l'assureur et dépendent du lieu de résidence, du type d'assurance requis et ne sont pas liées au niveau de revenu. Toutefois, grâce aux subventions des cantons (réduction des primes), les citoyen·nes à faible revenu, les enfants et les jeunes adultes en formation à temps plein paient souvent des primes réduites. Il revient aux cantons de décider de qui pourra bénéficier d'une réduction (FOPH 2020).

Vers la fin de chaque année, le SVA St. Gallen reçoit environ 80 000 demandes de réduction de prime. Pour réduire la

charge de travail liée à ce déluge de demandes, l'institut a testé un chatbot via Facebook Messenger. L'objectif de ce projet pilote était de proposer aux clients une méthode de communication alternative. Le premier assistant administratif numérique a été conçu pour fournir aux clients des réponses automatiques aux questions les plus importantes concernant les réductions de primes. Par exemple : que sont les réductions de primes et comment peut-on en bénéficier ? Puis-je demander une réduction de prime ? Existe-t-il des cas particuliers et comment dois-je procéder ? Comment la réduction de prime est-elle calculée et versée ? En outre, si cela était indiqué, le chatbot pouvait renvoyer les clients vers d'autres services proposés par le SVA St. Gallen, notamment le calculateur de réduction de prime et le formulaire d'inscription interactif. Bien que le chatbot ne prenne pas la décision finale d'accorder une réduction de prime, il peut néanmoins réduire le nombre de demandes en informant les citoyens qui n'y ont pas droit que leur demande risque d'être rejetée. Il joue également un rôle important dans la diffusion de l'information (Ringeisen/Bertolosi-Lehr/Demaj 2018 S.51-65).

Au vu des retours positifs de ce premier test, le chatbot a été intégré au site web du SVA St. Gallen en 2019, et l'institut prévoit d'étendre progressivement le chatbot à d'autres produits d'assurance. Il est ainsi envisagé d'utiliser le chatbot pour des services liés aux cotisations à l'assurance-vieillesse et survivants, à l'assurance-invalidité et à l'assurance perte de gain (IPV-Chatbot 2020).

/ Système pénal

En Suisse, l'exécution des peines et la justice pénale reposent sur un système à plusieurs niveaux. Dans le cadre de ce système, les détenus bénéficient généralement d'une liberté croissante au fil de leur incarcération. Il s'agit donc d'un processus de collaboration entre le pouvoir exécutif, les établissements pénitentiaires, les prestataires de thérapie et les services de probation. Bien entendu, les risques d'évasion et de récidive sont des facteurs décisifs lorsqu'il s'agit d'accorder plus de liberté aux détenus. Ces dernières années, et en réponse à des affaires dans lesquelles des criminels ont commis plusieurs actes de violence et délits sexuels graves, le modèle ROS (Risk-Oriented Sanctioning, ou exécution des sanctions orientée vers les risques) a été instauré. Le principal objectif du modèle ROS est de prévenir la récidive en harmonisant l'exécution des peines et les mesures prises aux différents échelons du système répressif, en mettant systématiquement l'accent sur la prévention de la récidive et la réinsertion. Le modèle ROS se divise en quatre phases : triage, évaluation, planification et avancement. Lors du triage, les cas sont classés en fonction de la nécessité d'évaluer le risque de récidive. Sur la base de cette classification, une analyse différenciée de chaque cas est effectuée au cours de l'étape d'évaluation. Au cours de la phase de planification, ces résultats sont exploités pour élaborer un plan d'exécution individuel pour la sanction de l'individu correspondant, qui sera continuellement réévalué au cours de la phase d'avancement (ROSNET 2020).

Le triage joue un rôle décisif au début de ce processus – tant pour le détenu qu'en termes d'ADM, car il est effectué par un outil de filtrage automatisé appelé FaST (Fall-Screening-Tool). Le programme FaST répartit automatiquement tous les cas dans des catégories A, B et C. La catégorie A indique qu'une évaluation n'est pas nécessaire ; la catégorie B, qu'il y a un risque général de délinquance ultérieure, et la catégorie C indique un risque de délinquance violente ou sexuelle. Cette classification est déterminée à partir des casiers judiciaires et repose sur des facteurs de risque statistiques généraux, notamment l'âge, les infractions violentes commises avant l'âge de 18 ans, les inscriptions du juge des enfants, le nombre de condamnations antérieures, la catégorie de l'infraction, les peines, la délinquance polymorphe, la période sans infraction après la remise en liberté et la violence domestique. Si des facteurs de risque qui, selon

LE MODÈLE ROS SE DIVISE EN QUATRE PHASES : TRIAGE, ÉVALUATION, PLANIFICATION ET AVANCEMENT.

des conclusions scientifiques, ont un lien spécifique avec des délits violents ou sexuels sont avérés, alors on applique la classification C. Si les facteurs de risque constatés ont un lien spécifique avec la délinquance générale, on applique la catégorie B. Si aucun ou presque aucun facteur de risque n'est trouvé, on applique la catégorie A. Par conséquent, la classification se compose d'éléments (facteurs de risque) sous forme de réponse fermée, chacun d'entre eux ayant une pondération différente (points). Si un facteur de risque est donné, ses points sont inclus dans la valeur totale. Pour connaître le résultat global, les éléments pondérés et confirmés sont additionnés pour obtenir le score final, qui se traduit par une classification dans la catégorie A, B ou C, qui, à son tour, servira de base pour décider si une évaluation supplémentaire est nécessaire (étape 2). Cette classification est effectuée de manière entièrement automatique par l'application d'ADM. Cependant, il est important de noter qu'il ne s'agit pas d'une analyse de risque, mais d'un outil permettant de filtrer les cas nécessitant une évaluation plus poussée (ROSNET 2020, Treuhardt/Kröger 2019 p. 76-85, Treuhardt/Kröger 2018 p. 24-32).

Néanmoins, la classification du triage a un effet sur la manière dont les responsables d'une institution particulière prennent des décisions et sur les évaluations qui doivent être effectuées. Elle détermine également le « profil de problème » du délinquant en ce qui concerne la planification de l'exécution des peines et des mesures à prendre (étape 3). Cette planification définit notamment toute facilitation éventuelle de l'exécution, telle que l'exécution en milieu ouvert, en semi-liberté ou dans un hébergement externe. En outre, aucune application d'ADM ne figure dans les autres étapes du modèle ROS. Le programme FaST n'est donc utilisé que dans la phase de triage.

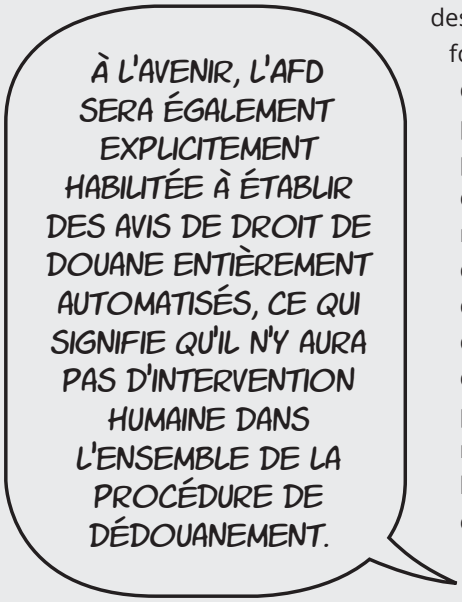
/ Police préventive

Dans certains cantons, en particulier à Bâle-Campagne, dans le canton d'Argovie et à Zurich, la police utilise des logiciels pour aider à prévenir les délits. Pour ce faire, elle s'appuie sur le logiciel commercial « PRECOBS » (Pre-Crime Observation System), qui sert uniquement à pronostiquer les cambriolages de domiciles. Ce type de criminalité relativement courante a fait l'objet de recherches scientifiques approfondies et les autorités policières disposent généralement d'une base de données solide portant sur la répartition spatiale et temporelle des cambriolages ainsi que les caractéristiques de ces délits. En outre, ces délits dénotent un auteur professionnel et présentent donc une probabilité de récidive supérieure à la moyenne. De plus, des modèles de pronostic correspondants peuvent être créés en utilisant une quantité relativement limitée de données. PRECOBS est donc basé sur l'hypothèse que les cambrioleurs frappent plusieurs fois en un bref laps de temps s'ils ont déjà réussi leur coup à un certain endroit.

Le logiciel est utilisé pour repérer certains motifs caractéristiques dans les rapports de police sur les cambriolages, tels que la manière dont les auteurs procèdent et le moment et le lieu où ils frappent. Par la suite, PRECOBS crée un rapport prévisionnel pour les zones où il existe un risque accru de cambriolage dans les 72 heures à venir. La police envoie alors des patrouilles ciblées dans la zone. PRECOBS génère donc des prévisions sur la base de décisions principalement saisies et n'utilise pas de méthodes d'apprentissage automatique. Bien qu'il soit prévu d'étendre à l'avenir PRECOBS à d'autres délits (tels que le vol de voiture ou le vol à la tire) et de créer ainsi de nouvelles fonctionnalités, il convient de noter que l'utilisation de la police prédictive en Suisse se limite actuellement à un domaine relativement restreint et clairement défini du travail de police préventif (Blur 2017, Leese 2018 p. 57-72).

/ Formalités douanières

Au niveau fédéral, on pense que l'ADM doit être particulièrement présente à l'Administration fédérale des douanes (AFD), car ce service est déjà fortement automatisé. Ainsi, l'évaluation des déclarations en douane est en grande partie déterminée par voie électronique. La procédure d'évaluation peut être divisée en quatre étapes : procédure d'examen sommaire, acceptation d'une déclaration en douane, vérification et inspection, suivies d'une décision d'évaluation. La procédure d'examen sommaire constitue un contrôle de plausibilité et est effectuée directement par le système utilisé dans le cas des déclarations en douane électroniques. Une fois le contrôle de plausibilité électronique effectué, le système de traitement des données ajoute automatiquement la date et

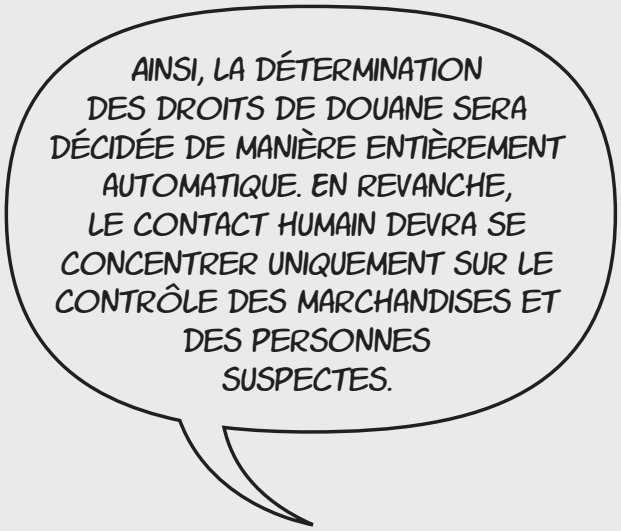


l'heure d'acceptation à la déclaration en douane électronique, ce qui signifie que la déclaration en douane a été acceptée. Jusqu'à ce stade, la procédure se déroule sans aucune intervention humaine de la part des autorités. Toutefois, le bureau de douane peut ensuite procéder à une inspection complète ou aléatoire des marchandises déclarées. À cette fin, le système informatisé effectue une sélection basée sur une analyse de risque. La dernière étape de la procédure est la délivrance de la décision d'évaluation. On ne sait pas si cette décision d'évaluation peut déjà être délivrée sans intervention humaine. Cependant, le programme DaziT permettra de clarifier cette incertitude.

Le programme DaziT est une mesure fédérale visant à numériser toutes les procédures douanières d'ici 2026 afin de simplifier et d'accélérer le passage des frontières. Les relations des autorités frontalières avec leurs clients en matière de circulation des biens et des personnes seront fondamentalement repensées. Les clients qui se comportent correctement devront pouvoir remplir leurs formalités par voie numérique, indépendamment de l'heure et du lieu. Si la mise en œuvre exacte du programme DaziT en est encore au stade de la planification, la révision de la loi sur les douanes correspondant à DaziT est incluse dans la révision de la loi fédérale sur la protection des données. Ceci est expliqué plus en détail ci-dessous, et devrait permettre de clarifier l'incertitude mentionnée précédemment concernant la procédure d'évaluation douanière automatisée : à l'avenir, l'AFD sera également explicitement habilitée à établir des avis de droit de douane entièrement automatisés, ce qui signifie qu'il n'y aura pas d'intervention humaine dans l'ensemble de la procédure de dédouanement. Ainsi, la détermination des droits de douane sera décidée de manière entièrement automatique. En revanche, le contact humain devra se concentrer uniquement sur le contrôle des marchandises et des personnes suspectes EZV 2020).

/ Assurance-accidents et assurance militaire

Lors de la révision de la loi sur la protection des données (expliquée plus en détail ci-dessous), il a été décidé que les compagnies d'assurance-accidents et d'assurance militaire seront habilitées à traiter automatiquement les données à caractère personnel. Il reste à savoir quelles activités automatisées les compagnies d'assurance envisagent d'utiliser à l'avenir. Toutefois, elles pourraient, par exemple, utiliser des algorithmes pour évaluer les dossiers médicaux des assurés. Grâce à ce système entièrement automatisé, les primes pourraient être calculées, et les décisions relatives aux demandes de prestations pourraient être prises et coordonnées avec d'autres prestations sociales. Il est prévu que ces organismes soient autorisés à prendre des décisions automatisées.



/ Reconnaissance automatique des véhicules

Ces dernières années, les hommes politiques et le grand public se sont préoccupés de l'utilisation de systèmes automatiques, tels que les caméras qui photographient les plaques d'immatriculation des véhicules, les lisent par reconnaissance optique de caractères et les comparent avec une base de données. Cette technologie peut être utilisée à diverses fins, mais pour l'instant, en Suisse, elle n'est utilisée que dans un cadre limité (EJPD 2019). Au niveau fédéral, le système de détection automatique des véhicules et de surveillance de la circulation n'est utilisé que comme un outil tactique en fonction de la situation et de l'évaluation des risques, ainsi que de considérations économiques, et uniquement aux frontières entre États (parlament.ch 2020). Le demi-canton de Bâle-Campagne a adopté une base juridique pour l'enregistrement automatique des plaques d'immatriculation des véhicules et leur rapprochement avec les bases de données correspondantes (EJPD 2019).

/ Répartition des élèves de primaire

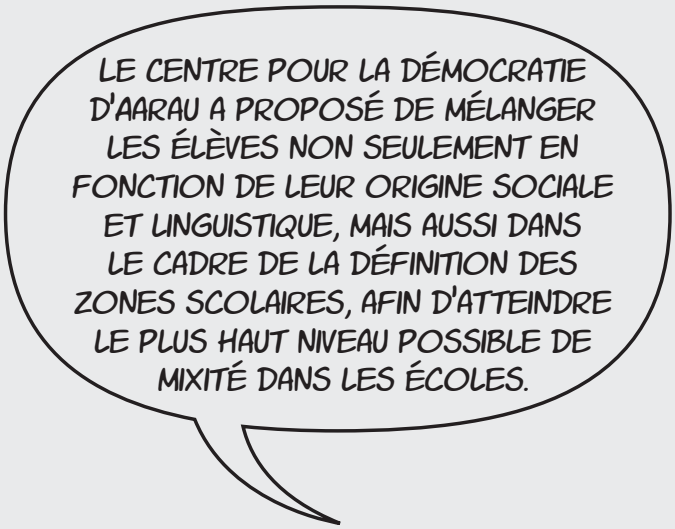
Un autre algorithme qui a été développé, mais n'est pas encore utilisé, est conçu pour répartir les élèves de primaire. Des études internationales montrent que la ségrégation sociale et ethnique entre les écoles urbaines est en constante augmentation. Cette situation est problématique, car la composition sociale et ethnique des écoles a

un effet avéré sur les performances des élèves, quelle que soit leur origine. Dans aucun autre pays de l'OCDE, ces « effets de composition » ne sont aussi prononcés qu'en Suisse. La composition différente des écoles est principalement due à la ségrégation entre les quartiers résidentiels et aux cartes scolaires correspondantes. C'est pourquoi le Centre pour la démocratie d'Aarau a proposé de mélanger les élèves non seulement en fonction de leur origine sociale et linguistique, mais aussi dans le cadre de la définition des zones scolaires, afin d'atteindre le plus haut niveau possible de mixité dans les écoles. Afin d'optimiser ce processus, un nouvel algorithme détaillé a été élaboré qui pourrait être utilisé à l'avenir pour faciliter la répartition des écoles et la planification des classes. L'algorithme a été formé pour reconstruire la carte scolaire et pour étudier la composition sociale de chaque école en utilisant les données du recensement des élèves de la première à la troisième année dans le canton de Zurich. Des données sur la circulation, le réseau de trottoirs et de voies piétonnes, les passages souterrains et les passerelles ont également été prises en compte. Ces données pourraient être utilisées pour calculer les endroits où les élèves doivent être placés pour obtenir davantage de mixité dans les classes. Dans le même temps, la capacité des bâtiments scolaires ne sera pas dépassée et le temps nécessaire pour se rendre à l'école restera raisonnable (ZDA 2019).

Politique, encadrement et débat public sur l'ADM

/ La structure fédérale de la Suisse comme circonstance prédominante

Dans les rapports sur la politique en Suisse, il convient de souligner la structure fédérale qui prévaut. Cette situation a déjà fait l'objet de réflexions dans les exemples d'ADM mentionnés précédemment. La Suisse est un État fédéral, composé de 26 États fédérés (cantons) très autonomes, qui à leur tour accordent à leurs municipalités une grande marge de manœuvre. Par conséquent, le débat politique et public sur l'ADM dépend largement du gouvernement correspondant, ce qui ne peut être décrit de manière exhaustive dans le présent rapport. De plus, cette fragmentation en matière de politique, de réglementation et de recherche fait courir le risque de travailler en parallèle sur des questions identiques, ce qui explique aussi pourquoi la Confédération s'efforce de mettre en place une coordination avancée, comme indiqué ci-dessous. Cependant, le gouvernement fédéral a l'entière responsabilité de certains domaines juridiques pertinents et de la gouvernance politique, qui lie légalement tous les gouvernements de Suisse et a donc un impact sur l'ensemble de la population. C'est pourquoi nous présentons ci-dessous les différents aspects du débat politique actuel au niveau fédéral.



/ Gouvernement et Parlement

À l'heure actuelle, le rôle de l'ADM dans la société, que l'on désigne généralement sous le nom d'IA, est principalement traité dans le cadre d'une discussion plus large sur la numérisation. Le gouvernement fédéral n'a pas de stratégie spécifique en ce qui concerne l'IA ou l'ADM, mais ces dernières années, il a lancé la stratégie « Suisse numérique », dans le cadre de laquelle tous les aspects de l'IA seront intégrés. Plus généralement, le cadre juridique national en matière de numérisation sera simultanément ajusté par la révision de la loi fédérale sur la protection des données (LPD).

/ Suisse numérique

En 2018, et dans un contexte de numérisation croissante au-delà des services gouvernementaux, la Confédération a lancé la stratégie « Suisse numérique ». Celle-ci se focalise notamment sur les développements actuels dans le domaine de l'intelligence artificielle (BAKOM 2020). Le groupe de coordination interdépartementale « Suisse numérique » (GCI Suisse numérique), avec son unité de gestion « bureau d'affaires Suisse numérique », est responsable de la stra-

tégie, en particulier de sa coordination et de sa mise en œuvre (Digital Switzerland 2020).

Dans le cadre de la stratégie « Suisse numérique », le Conseil fédéral a mis en place un groupe de travail sur le thème de l'IA et l'a chargé de lui présenter un rapport sur les défis associés à l'IA. Le Conseil fédéral a pris acte de ce rapport en décembre 2019 (SBFI 2020). En plus d'aborder les principaux enjeux de l'IA – à savoir la traçabilité et les erreurs systématiques dans les données ou les algorithmes –, le rapport décrit la nécessité de prendre des mesures concrètes. Il est admis que tous les défis, y compris ce besoin d'action, dépendent fortement du domaine en question ; c'est pourquoi le rapport a examiné 17 domaines de manière plus approfondie, tels que l'IA dans les soins de santé, l'administration et la justice (SBFI 2020 b).

En principe, les défis posés par l'IA en Suisse ont, d'après le rapport, déjà été largement reconnus et abordés dans divers domaines politiques. Néanmoins, le rapport interdépartemental identifie un certain besoin d'agir, raison pour laquelle le Conseil fédéral a adopté quatre mesures : dans le domaine du droit international et sur l'utilisation de l'IA dans la formation de l'opinion publique et la prise de décision, des rapports supplémentaires seront commandés pour une clarification approfondie. Par ailleurs, des mesures permettant d'améliorer la coordination de l'utilisation de l'IA dans l'administration fédérale seront examinées. Plus particulièrement, la création d'un réseau de compétences, axé sur les aspects techniques de l'application de l'IA dans l'administration fédérale, sera étudiée. Enfin, une politique relative à l'IA sera prise en compte en tant que composante essentielle de la stratégie « Suisse numérique ». Dans ce contexte, le Conseil fédéral a décidé de poursuivre le travail interdépartemental et d'élaborer des orientations stratégiques pour la Confédération d'ici au printemps 2020 (SBFI 2020 b, SBFI 2020 c).

En outre, lors de sa séance du 13 mai 2020, le Conseil fédéral a décidé de créer un Centre national de compétences en matière de sciences des données. L'Office fédéral de la statistique (OFS) instituera ce centre interdisciplinaire le 1er janvier 2021. Le nouveau centre viendra soutenir l'administration fédérale dans la mise en œuvre de projets dans le domaine de la science des données. À cette fin, le transfert de connaissances au sein de l'administration fédérale ainsi que l'échange avec les cercles scientifiques, les instituts de recherche et les organismes responsables de l'application pratique seront encouragés. Le centre d'excellence contribuera notamment à la production d'informations transparentes tout en tenant compte de la protection des données. Le raisonnement qui sous-tend le nouveau centre est étayé par une déclaration du Conseil fédéral, qui affirme que la science des données gagne sans cesse en importance, tout particulièrement dans l'administration publique. D'après le Conseil fédéral, la science des données comprend des calculs « intelligents » (algorithmes) permettant d'automatiser certaines tâches complexes (Bundesrat 2020).

/ Mesures réglementaires

Comme la loi fédérale sur la protection des données est aujourd'hui rendue obsolète, en raison de l'évolution rapide de la technologie, le Conseil fédéral envisage d'adapter la LPD à cette nouvelle réalité technologique et sociale, et notamment d'améliorer la transparence du traitement des données et de renforcer l'autodétermination des personnes concernées à l'égard de leurs données. Parallèlement, cette révision totale devrait permettre à la Suisse de ratifier la convention révisée du Conseil de l'Europe ETS 108 sur la protection des données et d'adopter la directive (UE) 680/2016 relative à la protection des données dans le domaine des poursuites pénales, ce qu'elle est tenue de faire dans le cadre de l'accord de Schengen. Par ailleurs, cette révision devrait permettre de rapprocher l'ensemble de la législation suisse sur la protection des données des exigences du règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (RGPD). Cette révision est actuellement débattue au Parlement (EJPD 2020).

S'il est évident que la révision totale de la loi fédérale sur la protection des données va entraîner une révision de l'ensemble de la législation et de tous ses différents aspects, une nouvelle disposition présente un intérêt tout particulier à l'égard de l'ADM. Dans le cas de ce que l'on appelle les « décisions individuelles automatisées », il devrait y avoir une obligation d'informer la personne concernée si la décision a des conséquences juridiques ou des effets importants. La personne concernée devrait également pouvoir demander à faire réexaminer la décision par une personne physique, ou à être informée de la logique sur laquelle se fonde cette décision. Ainsi, une réglementation différenciée sera prévue pour les décisions des organismes fédéraux. En conséquence, même si les organismes fédéraux doivent également signaler la décision individuelle automatisée comme telle, la possibilité pour la personne concernée de demander un réexamen par une personne peut être limitée par d'autres lois fédérales. À la différence du RGPD de l'UE, il n'est ni interdit de

prendre des décisions automatisées, ni possible d'exiger de ne pas faire l'objet d'une telle décision (SBFI 2020 b).

/ Société civile, universités et autres organisations

Par ailleurs, un certain nombre de forums en Suisse étudient, débattent et travaillent sur la transformation numérique et ses opportunités, ses défis, ses besoins et son éthique. La plupart d'entre eux abordent cette question de manière généraliste, bien que certains traitent spécifiquement de l'ADM ou de l'IA.

/ Instituts de recherche

La Suisse compte un certain nombre de centres de recherche réputés et établis de longue date qui étudient la technologie de l'intelligence artificielle. Ces instituts incluent le Swiss AI Lab IDSIA à Lugano (SUPSI 2020) et l'Institut de recherche Idiap à Martigny (Idiap 2020) , ainsi que les centres de recherche des écoles polytechniques fédérales de Lausanne (EPFL) (EPFL 2020). En outre, des initiatives privées telles que le Groupe pour l'intelligence artificielle et les sciences cognitives (SGAICO) viennent compléter ces initiatives universitaires, réunissant chercheurs et utilisateurs et favorisant le transfert de connaissances, le développement de la confiance et l'interdisciplinarité (SGAICO 2020).

/ Financement de la recherche par le gouvernement

La Confédération se penche également sur le sujet de l'IA par le biais du financement de la recherche. Par exemple, le gouvernement fédéral investit actuellement dans deux programmes de recherche nationaux par l'intermédiaire du Fonds national suisse de la recherche scientifique (FNS) (SNF 2020). : d'une part, le programme de recherche national 77 « Transformation numérique » (NRP 77) (NRP77 2020) . Le premier étudie les interdépendances et les effets concrets de la transformation numérique en Suisse, et se focalise sur l'éducation et l'apprentissage, l'éthique, la fiabilité, la gouvernance, l'économie et le marché du travail (NFP 77 2020). Le deuxième vise à établir une base scientifique pour une utilisation efficace et appropriée de grandes quantités de données. En conséquence, ces projets de recherche examinent les questions de l'impact social des technologies de l'information et abordent des applications concrètes (NRP 75 2020).

Un autre institut travaillant dans ce domaine est la Fondation pour l'évaluation des choix technologiques (TA-Swiss). TA-Swiss est un centre d'excellence des Académies suisses des arts et des sciences, dont la mission est définie dans la loi fédérale sur la recherche. Il s'agit d'un organe consultatif, financé par le secteur public, qui a commandé diverses études sur l'IA. La plus pertinente est une étude publiée le 15 avril 2020 sur l'utilisation de l'IA dans différents domaines (consommation, travail, éducation, recherche, médias, administration publique et justice). D'après cette étude, une loi distincte sur l'utilisation de l'IA n'est pas jugée nécessaire. Toutefois, les citoyen·nes, les consommateur·trices et les employé·es, dans leurs relations avec l'État, les entreprises ou leur employeur, doivent être informés de manière aussi transparente que possible sur l'utilisation de l'IA. Lorsque des institutions publiques ou des entreprises utilisent l'IA, elles doivent le faire dans le respect de règles claires, et de manière compréhensible et transparente. (Christen, M. et al. 2020).

/ Digital Society Initiative

La Digital Society Initiative a été lancée en 2016. Il s'agit d'un centre d'excellence de l'Université de Zurich pour la réflexion critique sur tous les aspects de la société numérique. Son objectif est de réfléchir et d'aider à façonner la numérisation de la société, de la démocratie, de la science, de la communication et de l'économie. En outre, il vise à réfléchir de manière critique et à modeler le changement actuel de la pensée amené par la numérisation dans une perspective d'avenir et à positionner l'Université de Zurich comme un centre d'excellence pour la réflexion critique sur tous les aspects de la société numérique à l'échelle nationale et internationale (UZH 2020).

/ Digitale Gesellschaft

La Digitale Gesellschaft (Société numérique) est une société à but non lucratif et une association généraliste consacrée à la protection des citoyens et des consommateurs à l'ère numérique. Depuis 2011, elle œuvre en tant qu'organisation de la société civile en faveur d'une sphère publique durable, démocratique et libre, et vise à défendre les droits fondamentaux dans un monde numérique interconnecté (Digitale Gesellschaft 2020)

/ Autres organisations

Plusieurs autres organisations suisses valent également d'être mentionnées. Ces organisations se focalisent sur la numérisation en règle générale, particulièrement dans un

contexte économique, par exemple l'Association suisse des télécommunications (asut) (Asut 2020), digitalswitzerland (Castle 2020), la Swiss Data Alliance (Swiss Data Alliance 2020) et Swiss Fintech Innovations (SFTI 2020).

Conclusions principales

L'ADM est utilisée dans diverses branches du secteur public en Suisse, mais de manière généralement non centralisée ou globale. Seuls quelques cantons utilisent l'ADM dans le travail de police, par exemple, et les systèmes utilisés varient. L'avantage d'une telle approche, c'est que les cantons ou le gouvernement fédéral peuvent bénéficier de l'expérience d'autres cantons. L'inconvénient, c'est que cela peut contribuer à des pertes d'efficacité. Il existe des fondements juridiques sélectifs, mais aucune loi uniforme sur l'ADM ou la gouvernance électronique, ni rien de semblable. Il n'y a pas non plus de stratégie spécifique en matière d'IA ou d'ADM, mais récemment, on a accordé une attention plus particulière à l'amélioration de la coordination, que ce soit entre différents ministères au niveau fédéral, ou entre le gouvernement fédéral et les cantons. Les méthodes d'apprentissage automatique ne sont pas utilisées dans les activités de l'État au sens strict, par exemple dans le travail de police ou dans le système de justice pénale, pour autant qu'on puisse en juger. En outre, à ce même niveau, l'ADM est utilisée ou évoquée de manière sélective, mais pas de manière globale. Dans le secteur public au sens plus large, l'ADM est utilisée plus souvent et plus largement. Un bon exemple en est le déploiement dans le système de santé suisse, l'hôpital universitaire de Genève étant devenu le premier hôpital d'Europe à utiliser l'ADM pour proposer des traitements aux patients cancéreux.

Références:

Asut (o. J.): in: *asut.ch,* [online] https://asut.ch/asut/de/page/index.xhtml [30.01.2020]

Bundesamt für Kommunikation BAKOM (o. J.): Digitale Schweiz, in: *admin.ch,* [online] https://www.bakom.admin.ch/bakom/de/home/digital-und-internet/strategie-digitale-schweiz.html [30.01.2020].

Der Bundesrat (o.J.): Der Bundesrat schafft ein Kompetenzzentrum für Datenwissenschaft, In: *admin.ch,* [online] https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-79101.html [15.05.2020].)

Christen, M. et al. (2020): Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz, in: TA-Swiss, [online] https://www.ta-swiss.ch/themen-projekte-publikationen/informationsgesellschaft/kuenstliche-intelligenz/ [15.05.2020].

Ciritsis, Alexander / Cristina Rossi / Matthias Eberhard / Magda Marcon / Anton S. Becker / Andreas Boss (2019): Automatic classification of ultrasound breast lesions using a deep convolutional neural network mimicking human decision-making, in: European Radiology, Jg. 29, Nr. 10, S. 5458–5468, doi: 10.1007/s00330-019-06118-7.

digitalswitzerland (Castle, Danièle Digitalswitzerland (2019): Digitalswitzerland - Making Switzerland a Leading Digital Innovation Hub, in: digitalswitzerland, [online] https://digitalswitzerland.com [30.01.2020])

Digital Switzerland (2020): (Ofcom, Federal Office Of Communications (o. J.): Digital Switzerland Business Office, in: *admin.ch,* [online] https://www.bakom.admin.ch/bakom/en/homepage/ofcom/organisation/organisation-chart/information-society-business-office.html [30.01.2020].)

EPFL (o. J.): in: *epfl,* [online] https://www.epfl.ch/en/ [30.01.2020]

EJPD (o. J.): Stärkung des Datenschutzes, in: *admin.ch,* [online] https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/datenschutzstaerkung.html [30.01.2020c].

E-ID Referendum (o.J.): in: e-id-referendum.ch/, [online] https://www.e-id-referendum.ch [31.1.2020].

EJPD (o. J.): Stärkung des Datenschutzes, in: *admin.ch,* [online] https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/datenschutzstaerkung.html [30.01.2020c].

EJPD Eidgenössisches Justiz- und Polizeidepartement (2019): Änderung der Geschwindigkeitsmessmittel-Verordnung (SR 941.261) Automatische Erkennung von Kontrollschildern, in: *admin.ch,* [online] https://www.admin.ch/ch/d/gg/pc/documents/3059/Erl_Bericht_de.

EZV (2020): EZV, Eidgenössische Zollverwaltung (o. J.): Transformationsprogramm DaziT, in: *admin.ch,* [online] https://www.ezv.admin.ch/ezv/de/home/themen/projekte/dazit.html [30.01.2020].

ETH Zurich - Homepage (o. J.): in: *ETH Zurich - Homepage | ETH Zurich,* [online] https://ethz.ch/en.html [30.01.2020].

Federal office of public health FOPH (2020): (Health insurance: The Essentials in Brief (o. J.): in: *admin.ch,* [online] https://www.bag.admin.ch/bag/en/home/versicherungen/krankenversicherung/krankenversicherung-das-wichtigste-in-kuerze.html [13.02.2020].)

Geschäft Ansehen (o. J.): in: *parlament.ch,* [online] https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20143747 [30.01.2020].

Heinhold, Florian (2019): Hoffnung für Patienten?: Künstliche Intelligenz in der Medizin, in: *br.ch,* [online] https://www.br.de/br-fernsehen/sendungen/gesundheit/kuenstliche-intelligenz-ki-medizin-102.html [30.01.2020].

Idiap Research Institute (o. J.): in: *Idiap Research Institute, Artificial Intelligence for Society,* [online] https://www.idiap.ch/en [30.01.2020]

Der IPV-Chatbot – SVA St.Gallen (o. J.): in: *svasg.ch,* [online] https://www.svasg.ch/news/meldungen/ipv-chatbot.php [30.01.2020].

Leese, Matthias (2018): Predictive Policing in der Schweiz: Chancen, Herausforderungen Risiken, in: *Bulletin zur Schweizerischen Sicherheitspolitik,* Jg. 2018, S. 57–72.

Lindner, Martin (2019): KI in der Medizin: Hilfe bei einfachen und repetitiven Aufgaben, in: Neue Zürcher Zeitung, [online] https://www.nzz.ch/wissenschaft/ki-in-der-medizin-hilfe-bei-einfachen-und-repetitiven-aufgaben-ld.1497525?reduced=true [30.01.2020

Medinside (o. J.): in: *Medinside,* [online] https://www.medinside.ch/de/post/in-genf-schlaegt-der-computer-die-krebsbehandlung-vor [14.02.2020].

NRP 75 Big Data (o. J.): in: *SNF,* [online] http://www.snf.ch/en/researchinFocus/nrp/nfp-75/Pages/default.aspx [30.01.2020].

NFP [Nr.] (o. J.): in: *nfp77.ch,* [online] http://www.nfp77.ch/en/Pages/Home.aspx [30.01.2020]

NRP 75 Big Data (o. J.): in: *SNF,* [online] http://www.snf.ch/en/researchinFocus/nrp/nfp-75/Pages/default.aspx [30.01.2020].

NFP [Nr.] (o. J.): in: *nfp77.ch,* [online] http://www.nfp77.ch/en/Pages/Home.aspx [30.01.2020]

Ringeisen, Peter / Andrea Bertolosi-Lehr / Labinot Demaj (2018): Automatisierung und Digitalisierung in der öffentlichen Verwaltung: digitale Verwaltungsassistenten als neue Schnittstelle zwischen Bevölkerung und Gemeinwesen, in: Yearbook of Swiss Administrative Sciences, Jg. 9, Nr. 1, S. 51–65, doi: 10.5334/ssas.123.

ROSNET > ROS allgemein (o. J.): in: *ROSNET,* [online] https://www.rosnet.ch/de-ch/ros-allgemein [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (o. J.): Künstliche Intelligenz, in: *admin.ch,* [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (2019): Herausforderungen der künstlichen Intelligenz - Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, in: *admin.ch,* [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (o. J.): Künstliche Intelligenz, in: *admin.ch,* [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (2019): Herausforderungen der künstlichen Intelligenz - Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, in: *admin.ch,* [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

Schaffhauser eID+ - Kanton Schaffhausen (o. J.): in: *sh.ch,* [online] https://sh.ch/CMS/Webseite/Kanton-Schaffhausen/Beh-rde/Services/Schaffhauser-eID--2077281-DE.html [30.01.2020].

SGAICO - Swiss Group for Artificial Intelligence and Cognitive Science (2017): in: *SI Hauptseite,* [online] https://swissinformatics.org/de/gruppierungen/fg/sgaico/ [30.01.2020]

SNF, [online] http://www.snf.ch/en/Pages/default.aspx [30.01.2020]

Srf/Blur;Hesa (2017): Wie «Precobs» funktioniert - Die wichtigsten Fragen zur «Software gegen Einbrecher», in: Schweizer Radio und Fernsehen (SRF), [online] https://www.srf.ch/news/schweiz/wie-precobs-funktioniert-die-wichtigsten-fragen-zur-software-gegen

SUPSI - Dalle Molle Institute for Artificial Intelligence - Homepage (o. J.): in: *idsia,* [online] http://www.idsia.ch [30.01.2020].

Swissdataaliance (o. J.): in: *swissdataalliance,* [online] https://www.swissdataalliance.ch [30.01.2020].

Swiss Fintech Innovations (SFTI introduces Swiss API information platform (2019): in: *Swiss Fintech Innovations – Future of Financial Services,* [online] https://swissfintechinnovations.ch [30.01.2020]).

Schwerzmann, Jacqueline Amanda Arroyo (2019): Dr. Supercomputer - Mit künstlicher Intelligenz gegen den Krebs, in: Schweizer Radio und Fernsehen (SRF), [online] https://www.srf.ch/news/schweiz/dr-supercomputer-mit-kuenstlicher-intelligenz-gegen-den-krebs

Treuthardt, Daniel / Melanie Kröger / Mirjam Loewe-Baur (2018): Der Risikoorientierte Sanktionenvollzug (ROS) – aktuelle Entwicklungen, in: *Schweizerische Zeitschrift für Kriminologie,* Jg. 2018, Nr. 2, S. 24–32.

Treuthardt, Daniel / Melanie Kröger (2019): Der Risikoorientierte Sanktionenvollzug (ROS) – empirische Überprüfung des Fall-Screening-Tools (FaST), in: *Schweizerische Zeitschrift für Kriminologie,* Jg. 2019, Nr. 1–2, S. 76–85.);

ZDA (2019): Durchmischung in städtischen Schulen, in: *zdaarau.ch,* [online] https://www.zdaarau.ch/dokumente/SB-17-Durchmischung-Schulen-ZDA.pdf [30.01.2020].

Équipe

/ Beate Autering

Mise en page et conception graphique

Beate Autering est une graphiste freelance. Elle est diplômée en design et dirige le studio beworx avec Tiger Stangl. Ensemble, il et elle créent des dessins, des graphiques et des illustrations et fournissent également des services d'édition d'images et de postprocessing. Parmi leurs clients figurent iRights, mdsCreative, Agentur Sehstern, Patrimoine mondial de l'UNESCO et visitBerlin.

/ Nadja Braun Binder

Autrice du chapitre sur la Suisse

Nadja Braun Binder a étudié le droit à l'Université de Berne et y a obtenu son doctorat. Son parcours universitaire l'a amenée à l'Institut de recherche en administration publique de Spire en 2011, où elle a mené des recherches sur l'automatisation des procédures administratives, entre autres. En 2017, elle a été habilitée par l'Université allemande des sciences administratives de Speyer. Elle a ensuite travaillé en tant que professeure assistante à la faculté de droit de l'Université de Zurich jusqu'en 2019. Depuis, Nadja est professeure de droit public à l'Université de Bâle. Ses recherches portent sur les questions juridiques liées à la numérisation au sein du gouvernement et de l'administration. Elle mène actuellement une étude sur l'utilisation de l'intelligence artificielle dans l'administration publique du canton de Zurich.

/ Fabio Chiusi

Éditeur, auteur de de l'introduction et du chapitre sur l'Europe

Photo: Julia Bornkessel

Fabio Chiusi travaille chez AlgorithmWatch en tant que corédacteur et chef de projet pour l'édition 2020 du rapport L'automatisation de la société. Après une décennie dans le journalisme technologique, il a commencé à travailler comme consultant et assistant de recherche dans le domaine des données et de la politique (Tactical Tech) et de l'IA dans le journalisme (Polis LSE). Il a coordonné le rapport Persuasori Social sur la réglementation des campagnes politiques sur les réseaux sociaux pour le projet PuntoZero, et a travaillé en tant que collaborateur technico-politique au sein de la Chambre des députés du Parlement italien pendant la législature actuelle. Fabio est chercheur au Nexa Center for Internet & Society à Turin et professeur adjoint à l'université de Saint-Marin, où il enseigne le journalisme et les nouveaux médias, l'édition et les médias numériques. Il est l'auteur de plusieurs essais sur la technologie et la société. Le dernier en date, « Io non sono qui. Visioni e inquietudini da un futuro presente » (DeA Planeta, 2018), est actuellement en cours de traduction en polonais et en chinois. Il écrit également en tant que journaliste spécialisé dans la politique technologique pour le blog collectif ValigiaBlu.

/ Samuel Daveti

Auteur de bandes dessinées

Samuel Daveti est un membre fondateur de l'association culturelle Double Shot. Il est l'auteur du roman graphique en langue française Akron le guerrier (Soleil, 2009), et le coordinateur du volume anthologique Fascia Protetta (Double Shot, 2009). En 2011, il est devenu membre fondateur du collectif de bandes dessinées autoproduites Mammaiuto. Samuel a également écrit Un Lungo Cammino (Mammaiuto, 2014 ; Shockdom, 2017), qui fera l'objet d'un film réalisé par la société de médias Brandon Box. En 2018, il a écrit The Three Dogs, illustré par Laura Camelli, qui a remporté le prix Micheluzzi au Napoli Comicon 2018 et le prix Boscarato du meilleur webcomic au Festival de la bande dessinée de Trévise.

/ Catherine Egli

Autrice du chapitre sur la Suisse

Catherine Egli a récemment obtenu un double master bilingue en droit des Universités de Bâle et de Genève. Sa thèse portait sur la prise de décision individuelle automatisée et la nécessité d'une révision de la Loi fédérale sur la procédure administrative sur le sujet. Parallèlement à ses études, elle a travaillé pour la chaire de la professeure Nadja Braun Binder en menant des recherches sur des questions juridiques liées à la prise de décision automatisée. Ses sujets de recherche préférés incluent la division des pouvoirs, la numérisation de l'administration publique et la démocratie numérique.

/ Sarah Fischer

Rédactrice du rapport

Sarah Fischer est chef de projet pour le projet « Éthique des algorithmes » à la Bertelsmann Stiftung, où elle est principalement responsable des études scientifiques. Elle a précédemment travaillé comme post-doctorante dans le cadre du programme « Confiance et communication dans un monde numérisé » à l'université de Münster, où elle s'est concentrée sur le thème de la confiance dans les moteurs de recherche. Dans ce même groupe de formation à la recherche, elle a obtenu son doctorat avec une thèse sur la confiance dans les services de santé sur Internet. Elle a étudié les sciences de la communication à l'université Friedrich Schiller de Iéna, et est le coauteur des articles « Where Machines can err. Sources of error and responsibilities in processes of algorithmic decision making » et « What Germany knows and believes about algorithms ».

/ Leonard Haas

Rédacteur adjoint

Leonard Haas travaille comme assistant de recherche chez AlgorithmWatch. Il est notamment responsable de la conception, de la mise en œuvre et de la maintenance de l'inventaire mondial des directives éthiques pour l'IA. Il est étudiant en master dans le domaine des sciences sociales à l'Université Humboldt de Berlin et détient deux licences de l'Université de Leipzig en humanités numériques et en sciences politiques. Ses recherches portent sur l'automatisation du travail et de la gouvernance. En outre, il s'intéresse à la politique des données d'intérêt public et aux luttes du travail dans l'industrie technologique.

/ Graham Holliday

Relecteur

Photo: Josh White

Graham Holliday est un rédacteur, auteur et professeur de journalisme indépendant. Il a occupé plusieurs postes à la BBC pendant près de deux décennies et a été correspondant de Reuters au Rwanda. Il travaille comme rédacteur pour les émissions Parts Unknown et Roads & Kingdoms de CNN – le journal international de la correspondance étrangère. Ses deux premiers livres, publiés par feu Anthony Bourdain, ont fait l'objet de recensions dans le New York Times, le Los Angeles Times, le Wall Street Journal, le Publisher's Weekly, le Library Journal et sur la radio NPR, entre autres médias.

/ Nicolas Kayser-Bril

Éditeur, auteur de l'article p.93

Photo: Julia Bornkessel

Nicolas Kayser-Bril est un journaliste spécialiste des données qui travaille pour AlgorithmWatch en tant que reporter. Il a été le pionnier des nouvelles formes de journalisme en France et en Europe et est l'un des plus grands experts dans le domaine du datajournalisme. Il intervient régulièrement dans des conférences internationales, enseigne le journalisme dans des écoles de journalisme françaises et dispense des formations dans les salles de rédaction. Journaliste et développeur autodidacte (ainsi que diplômé en économie), il a commencé par développer de petites applications interactives basées sur des données pour le journal Le Monde à Paris en 2009. Il a ensuite constitué l'équipe de datajournalisme d'OWNI en 2010, avant de cofonder et de gérer Journalism++ de 2011 à 2017. Nicolas est également l'un des principaux contributeurs au Guide du datajournalisme, l'ouvrage de référence pour la vulgarisation du datajournalisme dans le monde.

/ Anna Mätzener

Éditrice

Anna Mätzener est directrice générale de AlgorithmWatch Suisse. Elle a obtenu son doctorat en mathématiques à l'Université de Zürich, où elle a aussi étudié la philosophie et la philologie italienne. Avant de rejoindre AlgorithmWatch Suisse, elle était éditrice d'une maison d'édition scientifique internationale, spécialisée en mathématiques et en histoire des sciences. Elle a aussi enseigné les mathématiques dans un lycée de Zurich.

/ Lorenzo Palloni

Auteur de bandes dessinées

Lorenzo Palloni est un dessinateur de bandes dessinées, auteur de plusieurs romans graphiques et webcomics, écrivain primé, et l'un des fondateurs du collectif d'auteurs de bandes dessinées Mammaiuto. Il travaille actuellement sur des romans pour les marchés français et italien. Lorenzo est également professeur d'écriture de scénarios et de storytelling à la Scuola Internazionale di Comics di Reggio Emilia (École internationale de bande dessinée de Reggio Emilia).

/ Kristina Penner

Autrice du chapitre sur l'Europe

Photo: Julia Bornkessel

Kristina Penner est la conseillère exécutive d'AlgorithmWatch. Ses recherches portent sur les systèmes de sécurité sociale, la notation sociale et les impacts sociétaux de l'ADM, ainsi que sur la durabilité des nouvelles technologies d'un point de vue holistique. Son analyse du système de gestion des frontières de l'UE s'appuie sur son expérience antérieure en matière de recherche et de conseil sur le droit d'asile. Son expérience inclut également des projets sur l'utilisation des médias dans la société civile et le journalisme sensible aux conflits, ainsi que l'implication des parties prenantes dans les processus de paix aux Philippines. Elle est titulaire d'un master en études internationales/recherche sur la paix et les conflits de l'université Goethe de Francfort.

/ Alessio Ravazzani

Auteur de bandes dessinées

Alessio Ravazzani est un graphiste éditorial, dessinateur et illustrateur qui collabore avec les plus prestigieux éditeurs de bandes dessinées et de romans graphiques en Italie. Il est auteur au sein du collectif Mammaiuto, dont il est membre depuis sa fondation.

/ Matthias Spielkamp

Éditeur

Photo: Julia Bornkessel

Matthias Spielkamp est cofondateur et directeur exécutif d'AlgorithmWatch. Il a témoigné devant plusieurs commissions du Bundestag allemand sur l'IA et l'automatisation. Matthias est membre du conseil d'administration de la section allemande de Reporters sans frontières et des conseils consultatifs de la Stiftung Warentest et du Whistleblower Network. Il a été membre de ZEIT Stiftung, de Stiftung Mercator et de l'American Council on Germany. Matthias a fondé le magazine en ligne mobilsicher.de, qui traite de la sécurité des appareils mobiles et compte plus de 170 000 lecteurs par mois. Il a écrit et édité des livres sur le journalisme numérique et la gouvernance de l'Internet et a été nommé l'un des 15 architectes bâtissant un avenir axé sur les données par Silicon Republic. Il est titulaire d'une maîtrise de journalisme de l'université du Colorado à Boulder et d'une maîtrise de philosophie de l'université libre de Berlin.

/ Beate Stangl

Mise en page

Beate Stangl arbeitet als Diplomdesignerin in Berlin und gestaltet mit Schwerpunkt Editorial Design u.a. für beworx, Friedrich-Ebert-Stiftung, Buske Verlag, UNESCO Welterbe Deutschland e.V., Agentur Sehstern, iRights Lab, Landesspracheninstitut Bochum.

/ Marc Thümmler

Coordinateurs des publications

Photo: Julia Bornkessel

Marc Thümmler est responsable des relations publiques et de la sensibilisation chez AlgorithmWatch. Il est titulaire d'un master en études des médias, a travaillé comme producteur et monteur pour une société cinématographique, et a géré des projets pour la Deutsche Kinemathek et l'organisation de la société civile Gesicht Zeigen. En plus de ses fonctions principales chez AlgorithmWatch, Marc a participé à la campagne de financement et de crowdsourcing OpenSCHUFA, et il a coordonné le premier numéro du rapport L'automatisation de la société, publié en 2019.

ORGANISATIONS

/ AlgorithmWatch Suisse

AlgorithmWatch est une organisation de recherche et de plaidoyer à but non lucratif qui s'engage à surveiller et analyser les systèmes de prise de décision algorithmique ou automatisée (ADM) et leur impact sur la société. Si l'utilisation prudente des systèmes ADM peut profiter aux individus et à la société, elle comporte par ailleurs de grands risques. Afin de protéger l'autonomie humaine, les droits fondamentaux et afin maximiser le bien public, nous considérons qu'il est crucial de que les systèmes ADM soient responsables devant les institutions démocratiques. L'utilisation de systèmes ADM qui affectent de manière significative les droits individuels et collectifs doit être publique, de manière claire et accessible. Les individus doivent également être en mesure de comprendre comment les décisions sont prises et de les contester si nécessaire. Par conséquent, nous travaillons à permettre aux citoyen·nes de mieux comprendre les systèmes ADM et de développer des moyens de parvenir à une gouvernance démocratique de ces processus – avec un mélange de technologies, de réglementations et d›institutions de contrôle appropriées. Avec cela, nous nous efforçons de contribuer à une société juste et inclusive et de maximiser les avantages des systèmes ADM pour la société au sens large.

https://algorithmwatch.ch/fr/

ALGORITHM WATCH / CH

/ Bertelsmann Stiftung

La Bertelsmann Stiftung œuvre pour la promotion sociale et l'inclusion pour tous. Elle s'est engagée à faire progresser cet objectif grâce à des programmes visant à améliorer l'éducation, à façonner la démocratie, à faire progresser la société, à promouvoir la santé, à dynamiser la culture et à renforcer les économies. Par ses activités, la Bertelsmann Stiftung veut encourager les citoyens à contribuer au bien commun. Fondée en 1977 par Reinhard Mohn, cette fondation à but non lucratif détient la majorité des actions de Bertelsmann SE & Co. KGaA. La Bertelsmann Stiftung est une fondation privée non partisane.
Avec son projet « Ethics of Algorithms », la Bertelsmann Stiftung examine de près les conséquences de la prise de décision algorithmique dans la société dans le but de s'assurer que ces systèmes sont utilisés au service de la société. L'objectif est d'aider à informer et à faire progresser les systèmes algorithmiques qui facilitent une plus grande inclusion sociale. Cela implique de s'engager pour ce qui est le mieux pour une société plutôt que pour ce qui est techniquement possible – afin que les décisions informées par les machines puissent servir au mieux l'humanité

https://www.bertelsmann-stiftung.de/en

BertelsmannStiftung

/ Förderfonds Engagement Migros

Le fonds de soutien Engagement Migros permet le développement de projets pionniers qui ouvrent de nouvelles voies en expérimentant des solutions innovantes dans une société en mutation. Cette approche pragmatique combine soutien financier et services de coaching dans le cadre du Pionierlab. Engagement Migros existe grâce à l'apport annuel de quelque dix millions de francs des entreprises du groupe Migros; depuis 2012, il constitue un complément au Pour-cent culturel Migros. Plus d'informations sous:

https://www.engagement-migros.ch

ENGAGEMENT
UN FONDS DE SOUTIEN DU GROUPE MIGROS

Vita quotidiana nella società automatizzata. I sistemi che automatizzano processi decisionali sono diventati mainstream: che fare?

Di Fabio Chiusi

Il materiale contenuto in questo documento è aggiornato al 30 settembre 2020. Non è stato possibile includere avvenimenti e sviluppi occorsi in data posteriore.

In un grigio pomeriggio di agosto, a Londra, gli studenti erano furiosi. Si erano riversati a centinaia in piazza del Parlamento, in protesta, mostrando nei loro cartelli e slogan di schierarsi allo stesso tempo dalla parte di un inusuale alleato, i loro professori, e contro un altrettanto inusuale bersaglio, un algoritmo.

A causa della pandemia di COVID-19, le scuole erano state chiuse già in marzo, in Gran Bretagna. Consci della diffusione del virus in rapida espansione in tutta Europa durante l'estate del 2020, gli studenti ben sapevano che i loro esami di fine anno sarebbero stati cancellati, e la loro valutazione sarebbe di conseguenza — a qualche modo — mutata. Ciò che non avrebbero potuto immaginare, tuttavia, è che migliaia di loro avrebbero finito per ricavarne voti inferiori a quelli attesi.

Gli studenti riuniti in protesta sapevano a chi dare la colpa, come reso evidente dai loro slogan e canti: il sistema di decision-making automatico ("automated decision-making", d'ora in avanti "ADM") adottato dall'Ofqal (Office of Qualifications and Examinations Regulation, l'autorità competente a verificare la correttezza di esami e voti). L'autorità intendeva produrre una migliore valutazione, basata su dati, sia per le valutazioni degli esami per ottenere il GCSE (General Certificate of Secondary Education) che per quelle degli esami "A level" (o General Certificate of Education Advanced Level)[1], in modo che "la distribuzione dei voti segua un andamento simile a quello degli altri anni, così che gli studenti dell'anno in corso non debbano patire uno svantaggio sistemico a seguito delle circostanze presentatesi quest'anno".

Il governo voleva evitare gli eccessi di ottimismo[2] che, secondo le sue valutazioni, si sarebbero prodotti facendo ricorso al solo giudizio umano: in confronto alle serie storiche, i voti sarebbero risultati troppo alti. Ma il tentativo di essere "il più equi possibile, equi per gli studenti che non hanno potuto dare gli esami quest'estate" aveva invece fallito miseramente, e in quel grigio giorno di protesta d'agosto, gli studenti continuavano ad accorrere, cantare, mostrare cartelli che esprimevano un urgente bisogno di giustizia sociale. Alcuni di loro erano semplicemente disperati, altri si erano lasciati andare al pianto.

"Basta rubarci il futuro", recitava un cartello, facendo il verso alle proteste per la giustizia climatica di Fridays for the Future. Altri, tuttavia, erano più specificamente tarati sulle falle del sistema di ADM per l'assegnazione dei voti: "Valuta il mio lavoro, non il mio codice postale", siamo "studenti, non statistiche", recavano scritto, denunciando così i risultati discriminatori del sistema[3].

Un canto levatosi in seguito ha finito, tuttavia, per definire il futuro stesso della protesta: intonava "*Fuck the algorithm*", "fanculo l'algoritmo". Terrorizzati che il governo stesse automatizzando con noncuranza — e in modo opaco — il loro futuro, indipendentemente dalle loro reali abilità e sforzi, gli studenti avevano preso a gridare il loro diritto di non vedersi indebitamente restringere le opportunità di vita da un insieme di righe di codice mal programmate. Reclamavano diritto di parola, e ciò che avevano da dire, in effetti, andava ascoltato.

Perché gli algoritmi non sono né "neutrali" né oggettivi, anche se tendiamo a pensarlo. Replicano, invece, gli assunti e le credenze di chi decide di adottarli e programmarli. È sempre un umano dunque, non "gli algoritmi" o i sistemi di ADM, a essere responsabile sia delle buone che della cattive decisioni algoritmiche — o almeno, così dovrebbe essere. La macchina sarà pure sinistra, ma il fantasma al suo interno è sempre umano. E gli esseri umani sono complicati, perfino più degli algoritmi.

Gli studenti in protesta non erano ingenui al punto di credere che tutte le loro preoccupazioni dipendessero esclusivamente dall'algoritmo, in ogni caso. Non stavano intonando cori contro "l'algoritmo" in un raptus di determinismo tecnologico: erano piuttosto motivati dall'urgenza di promuovere e proteggere la giustizia sociale. Da questo punto di vista, le loro proteste ricordano piuttosto quelle dei Luddisti. Proprio come il movimento per i diritti dei lavoratori che, nel XIX secolo, distruggeva telai meccanici e altre macchine da lavoro industriale, gli studenti sanno che i sistemi di ADM sono questione di potere, e non dovrebbero venire scambiati per tecnologie "oggettive". Per questo cantavano "giustizia alla classe lavoratrice", chiedendo le dimissioni del ministro della Salute e ritraendo il sistema di ADM incriminato come "classismo in purezza" e "classismo spudorato".

1 Nel sistema scolastico del Regno Unito, il "GCSE" è un esame affrontato dagli studenti di circa 16 anni; gli esami "A level" si tengono di norma dopo un ulteriore biennio di studio, e servono a decidere l'ingresso nel sistema educativo universitario.

2 "La letteratura suggerisce che, nello stimare la probabilità dei voti che gli studenti otterranno, gli insegnanti tendono a essere ottimisti (anche se non in tutti i casi)", scrive l'Ofqual, cfr. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/909035/6656-2_-_Executive_summary.pdf

3 Cfr. il capitolo sulla Gran Bretagna per i dettagli.

Infine, gli studenti sono riusciti ad abolire il sistema che avrebbe messo le loro opportunità professionali e di vita in pericolo: in una incredibile giravolta, il governo britannico ha infatti dismesso il sistema di ADM rivelatosi fallace, e deciso di affidarsi al giudizio degli insegnanti.

Ma c'è altro in questa storia, oltre alla vittoria dei manifestanti. L'esempio sottolinea infatti come sistemi mal progettati, implementati e controllati che riproducono bias e discriminazioni umane non possano mettere a frutto i potenziali benefici dei sistemi di ADM, per esempio in termini di comparabilità ed equità.

Più chiaramente rispetto a molte battaglie del passato, questa protesta rivela che non stiamo più solamente "automatizzando la società", ma che l'abbiamo già automatizzata.

Finalmente, qualcuno se ne è accorto.

/ Da "Automating Society" alla società automatizzata

Quando, nel gennaio 2019, abbiamo lanciato la prima edizione di questo rapporto, abbiamo deciso di chiamarlo "Automating Society", cioè "automatizzando la società", perché allora i sistemi di ADM reperibili in Europa erano principalmente nuovi, sperimentali e sconosciuti. Soprattutto, erano l'eccezione, piuttosto che la norma.

La situazione è cambiata rapidamente. Come mostrato chiaramente nei molteplici esempi raccolti per questo rapporto dal nostro straordinario network di ricercatori, l'adozione di sistemi di ADM è fortemente incrementata in poco più di un anno. I sistemi di ADM riguardano oggi ogni tipo di attività umana e, in particolare, la distribuzione di servizi essenziali a milioni di cittadini europei — nonché il loro effettivo accesso ai propri diritti.

L'opacità testarda che continua a circondare l'uso sempre crescente di sistemi di ADM ha reso anche più urgente moltiplicare, di conseguenza, i nostri sforzi. Per questo abbiamo aggiunto quattro paesi (Estonia, Grecia, Portogallo e Svizzera) ai dodici già analizzati nell'edizione precedente di questo rapporto, portando il totale a sedici paesi. Anche se ben lontana da rappresentare un'analisi esaustiva, ciò ci consente di fornire un ritratto più ampio dello scenario dell'ADM in Europa. Considerando l'impatto che questi sistemi possono avere sulla vita di ogni giorno, e quanto in profondità sfidino le nostre intuizioni — quando non le norme e regole — circa il rapporto tra democrazia e automazione, crediamo si tratti di uno sforzo indispensabile.

Ciò è vero a maggior ragione nel contesto della pandemia di COVID-19, un momento storico in cui abbiamo testimoniato l'adozione (perlopiù affrettata) di una pletora di sistemi di ADM mirati a contribuire alla salute pubblica attraverso dati e automazione. Abbiamo ritenuto il fenomeno talmente importante da farne l'oggetto di un "rapporto-anteprima" (*preview report*), pubblicato[4] ad agosto 2020 sempre all'interno del progetto 'Automating Society'.

Perfino in Europa, quando si tratta di adottare sistemi di ADM, il limite è la fantasia. Si pensi anche solo ad alcuni dei casi raccontati in questo rapporto, che vanno ad aggiungersi ai tanti — dal welfare all'educazione, dal sistema sanitario a quello giudiziario — di cui abbiamo già dato conto nell'edizione precedente. Nelle pagine che seguono, e per la prima volta, forniremo aggiornamenti e sviluppi su questi casi in tre modi: tramite articoli giornalistici, attraverso una sezione di ricerca che elenca i diversi esempi raccolti, e infine con storie illustrate. Abbiamo ritenuto che questi sistemi di ADM siano — e saranno sempre più — talmente importanti nelle vite di ciascuno di noi da richiedere ogni possibile sforzo per comunicare come funzionino, e *cosa ci facciano* davvero, così da raggiungere ogni possibile tipologia di pubblico. Dopotutto, i sistemi di ADM ci riguardano tutti.

O almeno, dovrebbero. Abbiamo visto, per esempio, come un nuovo servizio automatizzato, e proattivo, distribuisca i bonus famiglia in Estonia. I genitori non devono nemmeno più fare domanda: lo Stato registra semplicemente tutte le informazioni di un neonato fin dalla nascita, raccogliendole in diversi database. Il risultato è che i genitori ricevono il bonus automaticamente, se ne hanno diritto.

In Finlandia, l'identificazione di fattori di rischio individuali correlati all'esclusione sociale dei giovani è automatizzata attraverso uno strumento sviluppato dal gigante giapponese Fujitsu. In Francia, i dati dei social network possono essere analizzati per addestrare algoritmi di machine learning che vengono impiegati per scovare frodi fiscali.

L'Italia sta sperimentando la "giustizia predittiva", che fa ricorso all'automazione per aiutare i giudici a individuare, nella giurisprudenza passata, tendenze utili a valutare

4 'Automated Decision-Making Systems in the COVID-19 Pandemic: A European Perspective', https://algorithmwatch.org/en/project/automating-society-2020-covid19/

il caso in esame. E, in Danimarca, il governo ha provato a controllare ogni click di tastiera e mouse sui computer degli studenti durante gli esami di fine anno, causando — anche qui — una considerevole protesta studentesca che ha portato all'abbandono del sistema, per il momento.

/ È tempo di correggere i torti dell'ADM

In linea di principio, i sistemi di ADM hanno il potenziale di arrecare beneficio alle vite dei cittadini— processando per esempio enormi moli di dati, accompagnando le persone nei loro processi decisionali, e fornendo loro applicazioni personalizzate.

In pratica, tuttavia, abbiamo reperito solo un numero esiguo di casi che dimostri in modo convincente tale impatto positivo.

Per esempio, il sistema VioGén, adottato in Spagna sin dal 2007 per valutare il rischio di violenze domestiche, pur se lontano dalla perfezione mostra "indici di prestazione ragionevoli" e ha contribuito a proteggere diverse donne da abusi.

In Portogallo, un sistema centralizzato e automatico adottato per combattere frodi associate a prescrizioni mediche ha apparentemente ridotto le frodi dell'80% in un solo anno. Un sistema simile, usato in Slovenia contro le frodi fiscali, si è dimostrato utile agli ispettori, a quanto dichiara il fisco sloveno[5].

Il riconoscimento facciale, quasi assente dall'edizione 2019, sta venendo sperimentato e adottato a ritmi allarmanti in tutta Europa.

Se si amplia lo sguardo allo stato attuale dei sistemi di ADM in Europa, si scopre che gli esempi positivi, che arrecano chiari benefici, sono rari. Nel corso di tutto il rapporto, si descrive piuttosto come la stragrande maggioranza degli usi tenda ad aumentare i rischi per i cittadini, invece di essere loro d'aiuto. Ma, per giudicare davvero i reali impatti positivi e negativi di questi sitemi, abbiamo bisogno di maggiore trasparenza circa i loro scopi, e di più dati sul funzionamento dei sistemi di ADM sperimentati e adottati.

Il messaggio per i decisori politici non potrebbe essere più chiaro. Se vogliamo davvero sfruttare al massimo il potenziale dei sistemi di ADM, rispettando insieme i diritti umani e la democrazia, il tempo di intervenire, renderli trasparenti e correggerne i torti è ora.

/ Riconoscimento facciale, riconoscimento facciale ovunque

Strumenti diversi stanno venendo adottati in paesi diversi. Una tecnologia, tuttavia, è oramai comune a molti: il riconoscimento facciale. È ciò che può essere definito lo sviluppo più nuovo, rapido e preoccupante tra quelli evidenziati in questo rapporto. Il riconoscimento facciale, quasi assente dall'edizione 2019, sta venendo sperimentato e adottato a ritmi allarmanti in tutta Europa. Nell'anno e poco più trascorso dal nostro ultimo rapporto, il riconoscimento facciale ha fatto il suo ingresso in scuole, stadi, aeroporti, e perfino nei casinò. È stato così usato come strumento di polizia predittiva, per arrestare criminali, contro il razzismo e, in risposta alla pandemia di COVID-19, per garantire il rispetto delle norme di distanziamento sociale, sia via app che attraverso videosorveglianza "intelligente".

L'adozione di nuovi sistemi di riconoscimento facciale prosegue nonostante il cumularsi delle prove scientifiche circa la loro scarsa accuratezza. E ogniqualvolta emerga un problema, i loro proponenti cercano di trovare un modo di aggirarlo. In Belgio, un sistema di riconoscimento facciale utilizzato dalla polizia è ancora "parzialmente attivo", anche se un divieto temporaneo è stato emanato da un organo di controllo, l'Oversight Board for Police Information. E, in Slovenia, l'uso di tecnologie di riconoscimento facciale da parte della polizia è stato legalizzato cinque anni dopo che gli agenti avevano cominciato a utilizzarle.

5 Cfr. il capitolo sulla Slovenia per i dettagli.

Questa tendenza, se incontrastata, rischia di normalizzare l'idea di essere costantemente osservati senza avere alcuna trasparenza su chi osserva, finendo così per cristallizzare uno status quo di sorveglianza di massa pervasiva. Ecco perché, su questo, molti all'interno della comunità delle organizzazioni per i diritti civili avrebbero accolto con favore una risposta politica molto più aggressiva da parte delle istituzioni europee[6].

Perfino il proprio sorriso è ora parte di un sistema di ADM sperimentato in alcune banche polacche: più l'impiegato sorride, migliore è la ricompensa. E a essere sotto controllo non sono solo i volti. In Italia, un sistema di sorveglianza sonora è stato proposto come strumento anti-razzismo in tutti gli stadi di calcio.

/ Le scatole nere sono ancora scatole nere

Un risultato allarmante del nostro rapporto è che, mentre le cose sono rapidamente cambiate per quanto riguarda il tasso di adozione dei sistemi di ADM, lo stesso non si può dire della loro trasparenza. Nel 2015, il docente della Brooklyn Law School, Frank Pasquale, popolarizzò l'idea che una società connessa basata su sistemi algoritmici opachi fosse una "black box society", ossia una società fatta di "scatole nere", i cui contenuti e dinamiche di funzionamento sono inconoscibili. Cinque anni dopo, sfortunatamente, la metafora ancora regge — e si applica a tutti i paesi studiati in questo rapporto, senza eccezioni: non c'è abbastanza trasparenza sui sistemi di ADM, né nel settore pubblico, né in quello privato. La Polonia addirittura obbliga all'opacità, con la legge che ha introdotto il suo sistema automatizzato per riconoscere conti bancari usati per attività illegali ("STIR"). La legge stabilisce che rivelare gli algoritmi e gli indicatori di rischio adottati può comportare fino a cinque anni di carcere.

Se da un lato rigettiamo l'idea che tutti questi sistemi siano intrinsecamente malvagi — abbracciamo, al contrario, una prospettiva pragmatica, basata su fatti ed evidenze scientifiche (*evidence-based*) — è dall'altro indubbiamente male essere incapaci di valutarne il funzionamento e l'impatto sulla base di conoscenze accurate e fattuali. Anche solo perché l'opacità ostacola gravemente la capacità di raccogliere prove necessarie a formulare un giudizio informato sull'adozione stessa dei sistemi di ADM.

Quando vi si aggiungono le difficoltà che entrambi i nostri ricercatori e giornalisti hanno incontrato nell'accedere a qualunque dato realmente significativo su questi sistemi, si comprende quanto lo scenario dipinto sia problematico per chiunque desideri tenerli sotto controllo, e garantire che lo sviluppo dei sistemi di ADM sia compatibile con i diritti fondamentali, lo stato di diritto, e la democrazia.

/ Sfidare lo status quo algoritmico

Di fronte a tutto questo, come sta reagendo l'Unione Europea? Anche se i documenti strategici prodotti dalla Commissione UE, sotto la guida di Ursula Von der Leyen, fanno più generico riferimento all'"intelligenza artificiale" piuttosto che parlare direttamente di sistemi di ADM, ciò non significa che non contengano lodevoli intenzioni, a partire da quella di promuovere e realizzare una "AI degna di fiducia" ("trustworthy AI") che veda al centro la persona umana ("people first")[7].

Eppure di fatto, come descritto nel capitolo sull'Europa, l'approccio complessivo dell'Unione dà la priorità all'imperativo, commerciale e geopolitico, di guidare la "rivoluzione dell'AI" piuttosto che a quello di assicurarsi che i suoi prodotti siano coerenti con le tutele democratiche, una volta adottati come strumenti di policy.

Questa mancanza di coraggio politico, specialmente evidente nella decisione di accantonare l'idea di una moratoria sull'uso di tecnologie di riconoscimento facciale dal vivo in luoghi pubblici nel suo pacchetto di norme sull'AI, è sorprendente, in particolar modo in un frangente storico in cui diversi stati membri stanno testimoniando un numero crescente di cause — e sconfitte — giudiziarie per sistemi di ADM da loro adottati troppo frettolosamente, finendo per impattare in modo negativo sui diritti dei cittadini.

Storico è un caso proveniente dall'Olanda, dove degli attivisti per i diritti civili sono riusciti a portare in giudizio un sistema automatizzato opaco e invasivo, chiamato SyRI, che avrebbe dovuto individuare casi di frode al sistema del welfare, e vincere. Non solo, infatti, a febbraio 2020 la corte dell'Aia ha sospeso il sistema, ritenendolo in violazione della Convenzione europea dei diritti dell'uomo. Il caso ha anche istituito un precedente: secondo la sentenza, i governi hanno la "speciale responsabilità" di salvaguardare i diritti umani, qualora adottino tali sistemi di ADM. Fornire

6 Come argomentato in dettaglio nel capitolo sull'Europa.

7 Si veda il capitolo sull'Europa, e in particolare la sezione dedicata al 'Libro bianco sull'AI' della Commissione Europea.

Un risultato allarmante del nostro rapporto è che, mentre le cose sono rapidamente cambiate per quanto riguarda il tasso di adozione dei sistemi di ADM, lo stesso non si può dire della loro trasparenza.

la tanto agognata trasparenza algoritmica ne è considerata parte cruciale.

Più in generale, dall'uscita del nostro primo rapporto, i media e gli attivisti della società civile si sono imposti quali forze trainanti verso la responsabilizzazione dei sistemi di ADM. In Svezia, per esempio, sono stati dei giornalisti a forzare la pubblicazione del codice del "sistema Trelleborg", creato per prendere decisioni completamente automatizzate in relazione a domande di sussidi sociali. A Berlino, il progetto-pilota per un sistema di riconoscimento facciale nella stazione ferroviaria di Südkreuz non ha condotto a una vera e propria implementazione in tutto il paese solo grazie alla rumorosa opposizione degli attivisti — talmente rumorosa da influenzare le posizioni dei principali partiti e, da ultimo, l'agenda politica del governo.

Gli attivisti greci di Homo Digitalis hanno dimostrato che nessun reale viaggiatore ha mai partecipato alle sperimentazioni del sistema chiamato 'iBorderCtrl', un progetto finanziato dall'UE il cui obiettivo era utilizzare sistemi di ADM per i controlli ai confini, chiarendo così che le effettive capacità di molti tra questi sistemi sono di frequente grandemente esagerate. Nel frattempo, in Danimarca, un sistema di profilazione per l'identificazione precoce di rischi associati a famiglie e figli con "fragilità" (il cosiddetto "modello Gladsaxe") veniva a sua volta fermato grazie al lavoro di accademici, giornalisti e dell'Autorità nazionale per la protezione dei dati personali (DPA).

Le Authority per la privacy hanno giocato un ruolo importante anche in altri paesi. In Francia, la DPA nazionale ha stabilito che due progetti per la sorveglianza sonora e il riconoscimento facciale negli istituti scolastici superiori fossero entrambi illegali. In Portogallo, ha rifiutato di approvare l'adozione di un sistema di videosorveglianza nelle città di Leiria e Portimão, perché ritenuto in violazione del principio di proporzionalità, e perché avrebbe costituito l'equivalente di un "monitoraggio sistematico su larga scala", un "tracciamento delle persone e delle loro abitudini e comportamenti", e di una "identificazione degli individui a partire da dati correlati a caratteristiche fisiche". In Olanda, poi, l'autorità Garante dei dati personali ha chiesto più trasparenza circa gli algoritmi predittivi utilizzati dalle agenzie governative.

Da ultimo, alcuni paesi hanno fatto ricorso alla figura del difensore civico (*ombudsperson*), a caccia di consigli. In Danimarca, il suo supporto ha contribuito a sviluppare strategie e linee guida per un uso etico dei sistemi di ADM nel settore pubblico. In Finlandia, il vice-difensore civico parlamentare ha considerato illegale la valutazione automatica del livello di tassazione.

E ciononostante, data la continua diffusione di tali sistemi in tutta Europa, viene da chiedersi: questo livello di controllo è sufficiente? Quando per esempio il difensore civico polacco ha messo in discussione la legalità del sistema di riconoscimento dei sorrisi adottato in una banca (e menzionato sopra), la decisione non ha impedito una ulteriore e successiva sperimentazione nella città di Sopot, né raffreddato l'interesse di diverse aziende, ancora decise ad adottarlo.

/ Mancano auditing, conseguenze, capacità e spiegazioni

L'attivismo è principalmente una misura reattiva. Il più delle volte, gli attivisti possono reagire solo quando un sistema di

ADM sta già venendo sperimentato, o è addirittura già stato adottato. Nel tempo necessario a organizzare una risposta, i diritti dei cittadini rischiano di essere già stati indebitamente intaccati. Ciò può accadere perfino in presenza delle forme di protezione garantite, in molti casi, dalla normativa europea e degli stati membri. È per questa ragione che garanzie proattive — e preventive, anteriori alla loro sperimentazione e adozione — sono così importanti per l'effettiva salvaguardia dei diritti dei cittadini.

Eppure, perfino nei paesi in cui una qualche forma di legislazione che includa tutele proattive è in vigore, l'effettiva applicazione della legge, molto semplicemente, non si verifica. In Spagna, per esempio, l'"azione amministrativa automatizzata" è codificata per legge, con annessi obblighi in termini di controllo qualità e supervisione, così come di auditing del sistema informatico e del suo codice sorgente. La Spagna si è anche dotata di una legge per l'accesso all'informazione (*freedom of information*). Eppure, nonostante queste leggi, solo raramente, scrive il nostro ricercatore, gli organi pubblici condividono informazioni dettagliate circa i sistemi di ADM che utilizzano. Allo stesso modo, in Francia esiste una legge che dal 2016 obbliga alla trasparenza algoritmica ma, di nuovo, senza esito.

Nemmeno portare un algoritmo di fronte a una corte giudiziaria, sulla base di specifiche disposizioni contenute in una legge per la trasparenza algoritmica, è abbastanza per proteggere davvero i diritti degli utenti. Come dimostrato dal caso francese dell'algoritmo di Parcoursup per la selezione di studenti universitari[8], le eccezioni per mettere al riparo una amministrazione pubblica da ogni forma di *accountability* si possono ricavare a piacimento.

Ciò è particolarmente problematico quando vi si aggiunga un contesto caratterizzato dalla endemica mancanza, nella pubblica amministrazione, di capacità e competenze riguardanti i sistemi di ADM da tempo lamentata da molti dei nostri ricercatori. E come potrebbero i pubblici ufficiali spiegare o fornire alcuna reale trasparenza su sistemi che non comprendono?

Di recente, alcuni paesi hanno tentato di affrontare il problema. L'Estonia, per esempio, ha predisposto un centro per competenze (*competence center*) rilevanti per i sistemi di ADM, per meglio comprendere come potrebbero venire usati per sviluppare nuovi servizi pubblici e, più nello specifico, informare le operazioni del Ministero per gli Affari economici e le comunicazioni, e della Cancelleria di Stato per lo sviluppo dell'e-government. Anche la Svizzera ha proposto la creazione di una "rete di competenze" (*competence network*) all'interno della più ampia cornice della strategia nazionale, chiamata "Svizzera Digitale".

Ma ciononostante, la ben nota mancanza di alfabetismo digitale resta un problema per buona parte della popolazione in diversi paesi europei. Per di più, è difficile chiedere il rispetto di diritti che non si sa di avere. Le proteste in Gran Bretagna e altrove, insieme a diversi scandali pubblici a base di sistemi di ADM[9], hanno di certo contribuito a innalzare il livello di consapevolezza sia dei rischi che delle opportunità derivanti dall'automatizzazione della società. Ma per quanto in crescita, questa consapevolezza muove ancora solo i primi passi in molti paesi.

I risultati della nostra ricerca sono chiari: per quanto i sistemi di ADM già influenzino ogni sorta di attività e giudizio, è ancora loro concesso di essere principalmente adottati senza alcun reale dibattito democratico. Inoltre è la norma, piuttosto che l'eccezione, osservare meccanismi di tutela e controllo in palese ritardo rispetto all'adozione dei sistemi che dovrebbero controllare — se e quando esistono.

Nemmeno lo scopo di questi sistemi viene comunemente giustificato o spiegato alle popolazioni che ne sono affette, né tantomeno vengono illustrati i benefici che ne dovrebbero loro derivare. Si pensi al servizio proattivo "AuroraAI" in Finlandia: dovrebbe identificare automaticamente gli "eventi di una vita" (*life events*), come riportano i nostri ricercatori finlandesi. Nella mente dei proponenti dovrebbe funzionare come una sorta di "tata", capace di aiutare i cittadini a soddisfare i loro bisogni di servizi pubblici in concomitanza con precisi eventi della propria vita — per esempio, cambiare residenza, cambiare stato nelle relazioni familiari, etc. Una forma di spiacevole "spinta gentile" (*nudging*) potrebbe essere all'opera in questo caso, scrivono i ricercatori, stando a dire che, invece di avvantaggiare le persone, il sistema potrebbe finire per fare l'esatto opposto, suggerendo alcune decisioni o limitando le opzioni a disposizione di un individuo già con il proprio design o la sua stessa architettura.

È di conseguenza anche più importante sapere cosa, più di preciso, stia venendo "ottimizzato" in termini di servizi pubblici: "è massimizzato l'uso del servizio, sono minimizzati i costi, o viene migliorato il benessere del cittadino?",

8 Cfr. il capitolo sulla Francia.

9 Si pensi al fiasco dell'algoritmo de "La buona scuola" in Italia, cfr. il relativo capitolo.

Un risultato allarmante del nostro rapporto è che, mentre le cose sono rapidamente cambiate per quanto riguarda il tasso di adozione dei sistemi di ADM, lo stesso non si può dire della loro trasparenza.

chiedono i ricercatori. "Su quale insieme di criteri si basano queste decisioni, e chi li stabilisce?" Il semplice fatto che non si abbiano risposte a queste fondamentali domande la dice lunga sul reale grado di partecipazione e trasparenza concesso, perfino per un sistema di ADM tanto invasivo.

/ La trappola tecno-soluzionista

C'è una più ampia giustificazione ideologica per tutto questo. È chiamata "soluzionismo tecnologico", e affligge ancora in profondità il modo in cui i sistemi di ADM vengono studiati e sviluppati. Anche se il termine è da tempo criticato come sinonimo di una ideologia fallace, che concepisce ogni problema sociale come un "bug" in attesa di essere aggiustato ("fix") tramite la tecnologia[10], questa retorica è ancora ampiamente sfruttata — sia nei media che in ambienti di policy — per giustificare l'adozione acritica di tecnologie di automazione nella vita pubblica.

Quando dipinti come "soluzioni", i sistemi di ADM fanno immediatamente ingresso nei territori meglio descritti dalla Terza Legge di Arthur C. Clarke: quelli in cui sono indistinguibili dalla magia. Ed è difficile, se non impossibile, imporre delle regole alla magia, così come renderla trasparente o spiegarla. Ciò che si vede è la mano frugare nel cappello e, di conseguenza, l'apparire di un coniglio, ma il procedimento è e *deve rimanere* una "black box", inconoscibile.

Diversi tra i ricercatori coinvolti nel progetto 'Automating Society' hanno individuato in questa ideologia l'errore fondamentale che informa la logica con cui vengono concepiti i sistemi di ADM descritti nel rapporto. Ciò implica anche, come mostrato nel capitolo sulla Germania, che gran parte delle critiche a tali sistemi vengano descritte come puro e semplice rigetto dell'"innovazione" tutta, mentre i sostenitori dei diritti digitali non sarebbero che "neo-luddisti". Ciò tuttavia non solo dimentica la realtà storica del movimento luddista, che si occupava di politiche del lavoro e non di mere tecnologie, ma minaccia inoltre, in un senso perfino più elementare, di compromettere l'efficacia dei meccanismi di controllo ipotizzati.

In un frangente storico in cui l'industria dell'AI sta testimoniando l'emergenza di un settore lobbistico particolarmente "vitale", a partire dalla Gran Bretagna, ciò potrebbe tradursi nell'adozione di comodo di mere linee guida per l'etica ("ethics-washing", il tentativo di prevenire norme e regole con un surplus di autoregolamentazione volontaria) e in altre risposte di policy inefficaci e strutturalmente inadeguate ad affrontare le implicazioni sui diritti umani dei sistemi di ADM. Ciò significherebbe, in ultima analisi, sposare l'assunto secondo cui siamo noi esseri umani a doverci

10 Si veda in proposito Evgeny Morozov (2014), *To Save Everything, Click Here. The Folly of Technological Solutionism*, Public Affairs, https://www.publicaffairsbooks.com/titles/evgeny-morozov/to-save-everything-click-here/9781610393706/

adattare ai sistemi di ADM, molto più che questi ultimi a dover essere modellati secondo i principi delle società democratiche.

Per contrastare questa narrativa, non dovremmo trattenerci dal porre domande fondamentali: per esempio, se i sistemi di ADM siano compatibili con la democrazia e possano essere impiegati per arrecare beneficio alla società tutta, e non solo ad alcuni. Potrebbe darsi, infatti, che certe attività umane — per dirne una tipologia, quelle collegate al welfare — non debbano essere assoggettate ad automazione, o che certe tecnologie — a partire dal riconoscimento facciale dal vivo in luoghi pubblici — non debbano essere incentivate nella ricerca, senza fine, della "leadership sull'AI", ma al contrario messe al bando.

Anche più importante è rigettare qualunque cornice ideologica ci impedisca di porre simili domande. Al contrario: ciò di cui abbiamo bisogno ora è che cambino alcune scelte politiche concrete — così che sia consentito un più approfondito scrutinio di questi sistemi. Nella sezione successiva elencheremo le principali richieste emerse dai nostri risultati di ricerca. Speriamo che siano ampiamente dibattute, e infine implementate.

Solo attraverso un dibattito informato, inclusivo, basato su evidenze scientifiche — un dibattito davvero democratico, dunque — riusciremo a trovare il giusto bilanciamento tra i benefici che i sistemi di ADM posso arrecare — e arrecano — in termini di rapidità, efficienza, equità, migliore prevenzione e accesso ai servizi pubblici, e le sfide che pongono ai diritti di tutti.

Che fare: raccomandazioni di policy

Alla luce dei risultati dettagliati nell'edizione 2020 del rapporto 'Automating Society', consigliamo ai policy-maker nel Parlamento Europeo e nei Parlamenti degli stati membri, alla Commissione UE, ai governi nazionali, ai ricercatori, alle organizzazioni della società civile (organizzazioni di *advocacy*, fondazioni, sindacati, etc.), e al settore privato (aziende e associazioni commerciali) di adottare il seguente insieme di interventi di policy. Queste raccomandazioni hanno l'obiettivo di meglio assicurare che i sistemi di ADM attualmente adottati e in via di adozione in tutta Europa siano effettivamente coerenti con il rispetto dei diritti umani e delle regole democratiche:

1. Più trasparenza per i sistemi di ADM

Senza la possibilità di sapere precisamente come, perché e a quali fini i sistemi di ADM vengano adottati, tutti gli altri sforzi per riconciliarli con i diritti fondamentali sono destinati a fallire.

/ Creare registri pubblici per i sistemi di ADM utilizzati nel settore pubblico

Chiediamo, di conseguenza, che venga adottata una legge di rango europeo che obblighi gli stati membri alla creazione di registri pubblici per i sistemi di ADM usati nel settore pubblico.

Dovrebbero inoltre essere accompagnati da obblighi di legge, per i responsabili dei sistemi di ADM, riguardanti la trasparenza e la documentazione dello scopo del sistema, una spiegazione del modello adottato (logica inclusa), e informazioni su chi l'ha sviluppato. Tutte queste informazioni devono essere rese disponibili in formati facilmente leggibili e accessibili, inclusi dati digitali strutturati e basati su protocolli standardizzati.

Le autorità pubbliche dovrebbero avere la responsabilità specifica di rendere trasparenti le componenti operative dei sistemi di ADM adottati nelle amministrazioni pubbliche — come sottolineato da un recente reclamo amministrativo in Spagna, che argomenta che "ogni sistema di ADM usato da un'amministrazione pubblica dovrebbe essere reso pubblico di default". Se confermato, il giudizio potrebbe diventare un precedente, in Europa.

/ Introdurre schemi legalmente vincolanti per l'accesso ai dati, a supporto della ricerca nell'interesse pubblico

Aumentare il grado di trasparenza fornito non richiede solamente la divulgazione di informazioni circa scopo, logica e creatore di un sistema, né la capacità di analizzarne nel dettaglio, e mettere alla prova, gli input e gli output. Richiede infatti anche che i dati attraverso cui i suoi algoritmi vengono addestrati, e i risultati da loro prodotti, siano resi accessibili a ricercatori indipendenti, giornalisti e organizzazioni della società civile, nell'interesse pubblico.

Ecco perché suggeriamo l'introduzione di schemi solidi e legalmente vincolanti per l'accesso ai dati, esplicitamente mirati a supportare e promuovere la ricerca a fini di pubblico interesse, e nel pieno rispetto della normativa sulla protezione dei dati e la privacy.

Facendo tesoro delle migliori esperienze (*best practices*) a livello nazionale ed europeo, questi schemi a più livelli dovrebbero includere sistemi di sanzioni, di tutele (*checks and balances*), e revisioni periodiche. Come illustrato dalle partnership per la condivisione dei dati con soggetti privati, ci sono legittime preoccupazioni in termini di privacy degli utenti e di possibile deanonimizzazione di certe tipologie di dati.

I policy-maker dovrebbero fare propri gli insegnamenti derivanti dai modelli di condivisione dei dati sanitari, così da rendere più semplice dare accesso privilegiato a certe tipologie di dati, più granulari, e al contempo garantire che i dati personali siano protetti in modo adeguato (per esempio, attraverso ambienti operativi sicuri).

E se ottenere uno schema di responsabilizzazione efficace richiede di certo un accesso trasparente ai dati in possesso delle piattaforme, quest'ultimo è un requisito indispensabile anche all'efficacia di svariati approcci di auditing.

2. Creare uno schema per una reale responsabilizzazione in tema di sistemi di ADM

Come dimostrato da quanto documentato per Spagna e Francia, perfino quando la trasparenza di un sistema di ADM diventa norma di legge e/o delle informazioni al suo riguardo vengono effettivamente divulgate, ciò non necessariamente comporta reale *accountability.* Affinché disposizioni di legge e altri requisiti vengano effettivamente rispettati, servono passi ulteriori.

/ Sviluppare e adottare approcci per un effettivo audit dei sistemi algoritmici

Per garantire che la trasparenza sia reale, c'è bisogno di completare il primo passo — la creazione di un registro pubblico — con processi che consentano un effettivo *audit* dei sistemi algoritmici.

Il termine "auditing", in italiano associato a "revisione", "verifica" e "controllo", è di largo utilizzo, ma non c'è consenso intorno a una sua comune definizione. Per noi, in questo contesto va compreso in accordo con la definizione dell'ISO: un "processo sistematico, indipendente e documentato per ottenere evidenze oggettive e valutarle oggettivamente, così da determinare in che misura i criteri dell'audit sono stati rispettati"[11].

Non disponiamo ancora di risposte soddisfacenti per tutte le complesse domande[12] sollevate dall'auditing dei sistemi algoritmici. Tuttavia, i nostri risultati mostrano chiaramente il bisogno di trovarle attraverso un ampio processo di coinvolgimento dei portatori di interesse, e tramite un lavoro di ricerca approfondito e specifico.

Dovrebbero essere sviluppati sia criteri che processi appropriati a costruire un sistema di auditing efficace, attraverso un approccio multi-stakeholder che tenga attivamente in considerazione l'effetto sproporzionato che i sistemi di ADM hanno sui gruppi sociali più vulnerabili — e ne solleciti, di conseguenza, la partecipazione.

Chiediamo ai policy-maker, di conseguenza, di dare inizio a questi processi con i portatori di interesse, così da chiarire le domande illustrate, e rendere disponibili le fonti di finanziamento necessarie a consentire la partecipazione degli stakeholder finora non rappresentati in modo adeguato.

Chiediamo inoltre la predisposizione di risorse adeguate a supportare e finanziare progetti di ricerca sullo sviluppo di modelli efficaci di auditing per i sistemi algoritmici.

11 https://www.iso.org/obp/ui/#iso:std:iso:19011:ed-3:v1:en

12 Pensando a potenziali modelli di auditing algoritmico, emergono diverse domande. 1) Chi o cosa (servizi, piattaforme, prodotti) dovrebbe essere oggetto dell'auditing? Come personalizzare i sistemi di auditing alla tipologia di piattaforma o servizio? 2) Quando dovrebbe essere responsabilità di una istituzione pubblica (a livello europeo, nazionale o locale), e quando invece può essere fatto da soggetti privati e da esperti (in campo commerciale, nella società civile o nella ricerca)? 3) Come chiarire la distinzione tra valutare un impatto ex-ante (i.e. nella fase di design) ed ex-post (i.e. quando è già in funzione), e le rispettive sfide? 4) Come valutare i trade-off tra virtù e vizi dell'auditing (per esempio, semplicità, generalità, applicabilità, precisione, flessibilità, interpretabilità, privacy, efficacia di una procedura di auditing possono essere in conflitto)?; 5) Quali informazioni devono essere disponibile affinché un audit sia efficace e affidabile (codice sorgente, dati di training, documentazione)? Gli ispettori necessitano di avere accesso fisico ai sistemi durante il loro funzionamento per compiere un audit efficace? 6) Quali obblighi di produrre prove sono necessari e proporzionati per venditori e fornitori di servizi? 7) Come possiamo assicurare che l'auditing sia possibile? C'è bisogno che i requisiti dell'auditing vengano considerati già nella fase di modellazione (*design*) del sistema algoritmico ("auditable by construction")? 8) Regole di pubblicità: quando un audit è negativo, e i problemi non sono ancora risolti, quale dovrebbe essere il comportamento degli auditor, in che modo dovrebbe essere reso pubblico il fallimento di un audit? 9) Chi controlla i controllori? Come assicurarsi che siano davvero responsabili delle proprie azioni?

/ Promuovere le organizzazioni della società civile a watchdog dei sistemi di ADM

I risultati della nostra ricerca indicano chiaramente che il lavoro delle organizzazioni della società civile è cruciale per sfidare efficacemente l'opacità dei sistemi di ADM. Attraverso ricerca e attivismo, e spesso in cooperazione con istituzioni accademiche e giornalisti, tali entità sono negli ultimi anni ripetutamente intervenute nei dibattiti politici su automazione e democrazia, riuscendo in molti casi ad assicurare che l'interesse pubblico e i diritti fondamentali fossero debitamente tenuti in considerazione, sia prima che dopo l'adozione di sistemi di ADM, in diversi paesi europei.

La società civile dovrebbe essere dunque supportata, in qualità di "cane da guardia" (*watchdog*) della società automatizzata. Come tale, la società civile è parte integrante di qualunque schema di *accountability* dei sistemi di ADM voglia dirsi efficace.

/ Mettere al bando il riconoscimento facciale capace di sorveglianza di massa

Non tutti i sistemi di ADM sono ugualmente pericolosi, e un approccio regolatorio basato sul rischio, come quello tedesco o della UE, lo riconosce correttamente. Ma se l'intento è fornire reale *accountability* per sistemi identificati come rischiosi, vanno insieme creati meccanismi efficaci di controllo e implementazione. Ciò è anche più importante per sistemi considerati "ad alto rischio" di violare i diritti degli utenti.

Un esempio cruciale, emerso dai risultati della nostra ricerca, è il riconoscimento facciale. I sistemi di ADM che sono basati su tecnologie biometriche, incluso il riconoscimento facciale, si sono rivelati una minaccia particolarmente seria per l'interesse pubblico e i diritti umani, visto che aprono la strada a forme di sorveglianza di massa indiscriminata — in particolare modo quando, come ora, vengono adottati perlopiù in modo non trasparente.

Chiediamo che gli usi pubblici di tecnologie di riconoscimento facciale capaci di sorveglianza di massa siano strettamente, e urgentemente, proibiti fino a data da destinarsi, a livello europeo.

Tali tecnologie potrebbero perfino essere considerate già illegali nell'UE, almeno per certi usi, se impiegate senza "consenso specifico" dei soggetti inquadrati. Questa lettura legale è stata suggerita dalle autorità in Belgio, quando hanno emesso una prima, storica multa per l'indebita adozione di un sistema di riconoscimento facciale nel paese.

3. Accrescere l'alfabetismo algoritmico e rafforzare il dibattito pubblico sui sistemi di ADM

Maggiore trasparenza sui sistemi di ADM può essere realmente utile solo se chi la deve realmente affrontare — legislatori, governi, organi di settore — è in grado di confrontarsi con l'impatto di tali sistemi in modo responsabile e prudente. Inoltre, coloro i quali subiscono quell'impatto devono poter essere in grado di comprendere quando, perché e come quei sistemi siano stati adottati. È per questa ragione che dobbiamo accrescere l'alfabetismo algoritmico, a tutti i livelli, sia tra i più importanti portatori di interesse che nel pubblico generalista, e rafforzare i dibattiti pubblici sui sistemi di ADM e il loro impatto sulla società, rendendoli maggiormente plurali.

/ Creare centri di competenze sull'ADM indipendenti

Insieme alle nostre richieste sull'auditing algoritmico e il supporto alla ricerca, chiediamo vengano costruiti dei centri di competenze sull'ADM indipendenti, a livello nazionale, per monitorare, valutare e condurre ricerca sui sistemi di ADM. Tali centri dovrebbero insieme fornire consigli al governo e al settore privato, in coordinamento con legislatori, società civile e accademia, circa le conseguenze per la società e i diritti umani derivanti dall'uso di tali sistemi. Il ruolo complessivo di questi centri è la creazione di un sistema funzionante di *accountability*, oltre alla formazione di competenze specifiche.

I centri nazionali di competenze dovrebbero coinvolgere le organizzazioni della società civile, i gruppi dei portatori di interesse, e gli organismi di *enforcement* attualmente esistenti, tra cui le Autorità per la protezione dei dati personali (DPAs) e gli organismi nazionali deputati al rispetto dei diritti umani, così da arrecare il massimo beneficio all'intero ecosistema e promuovere fiducia, trasparenza, e cooperazione tra tutte le parti in causa.

In qualità di organismi indipendenti, questi centri di expertise avrebbero un ruolo centrale nel coordinare gli sviluppi di policy e le strategie nazionali correlate all'ADM, così come nel contribuire all'ampliamento delle competenze e abilità

esistenti tra legislatori, governi e organi di settore, in risposta all'aumentato utilizzo dei sistemi di ADM.

Tali centri non dovrebbero disporre di poteri regolatori, ma piuttosto fornire le competenze necessarie a proteggere i diritti umani degli individui, e impedire danni sociali e alla collettività. Dovrebbero, per esempio, aiutare le piccole e medie imprese a soddisfare i loro obblighi in termini di "due diligence" sui diritti umani, incluse le valutazioni sui diritti umani e di impatto algoritmico, e l'iscrizione al registro pubblico dei sistemi di ADM sopra discusso.

/ Promuovere un dibattito democratico plurale e inclusivo sui sistemi di ADM

Oltre a rafforzare abilità e competenze di chi adotta i sistemi di ADM, è altrettanto vitale promuovere l'alfabetismo algoritmico nel pubblico generalista attraverso un più ampio dibattito e programmi pluralisti.

I nostri risultati suggeriscono non solo che i sistemi di ADM restino opachi al pubblico quando sono già utilizzati, ma che perfino la decisione di adottare o meno un sistema di ADM venga di norma presa a insaputa e senza alcuna partecipazione della cittadinanza.

C'è dunque un urgente bisogno di includere il pubblico (e l'interesse pubblico) nei processi decisionali sui sistemi di ADM fin dal principio.

Più in generale, c'è bisogno di un dibattito pubblico maggiormente plurale sugli impatti dell'ADM. Dobbiamo andare oltre il semplice relazionarsi a gruppi di esperti, e garantire invece che la questione risulti più accessibile a un pubblico più ampio. Ciò significa parlare una lingua diversa da quella meramente tecno-giudiziaria, ingaggiare il pubblico, e stimolarne la curiosità.

Per riuscirci, dovrebbero essere predisposti programmi dettagliati per fare e promuovere alfabetismo digitale. Se l'obiettivo è un dibattito pubblico informato per la creazione di vera autonomia digitale per i cittadini europei, dobbiamo cominciare a costruirlo, e promuoverlo davvero, con una attenzione specifica alle conseguenze sociali, etiche e politiche dell'adozione dii sistemi di ADM.

Alle radici del futuro dell'ADM in Europa

Se i sistemi di decision-making automatizzato (ADM) hanno assunto un ruolo crescente nella distribuzione di diritti e servizi in Europa, anche le istituzioni in tutto il Continente hanno preso a riconoscerne sempre più la presenza nella vita pubblica, sia in termini di opportunità che di sfide.

Di Kristina Penner e Fabio Chiusi

Dal nostro primo rapporto nel gennaio 2019 — e nonostante l'Unione Europea sia ancora indaffarata nel più ampio dibattito su un'intelligenza artificiale "degna di fiducia" ("Trustworthy AI") — diversi organi, dal Parlamento UE al Consiglio d'Europa, hanno pubblicato documenti il cui obiettivo è mettere l'UE e l'Europa tutta nelle condizioni di affrontare l'ADM negli anni, se non nelle decadi, a venire.

Nell'estate del 2019 la neoeletta presidente della Commissione, Ursula Von der Leyen, una "ottimista tecnologica" dichiarata, si è impegnata a proporre "un pacchetto legislativo per un approccio europeo coordinato sulle implicazioni umane ed etiche dell'intelligenza artificiale", e a "regolare l'AI" entro i primi cento giorni in carica. Invece, a febbraio 2020, la Commissione Europea ha pubblicato un 'Libro bianco' sull'AI ('White Paper On Artificial Intelligence - A European approach to excellence and trust') contenente "idee e azioni" — ovvero, un pacchetto strategico che ha l'obiettivo di informare i cittadini e aprire la strada per futuri interventi legislativi. Il documento argomenta anche in favore di una "sovranità tecnologica" europea: nelle parole della stessa Von der Leyen, ciò si traduce nella "capacità che l'Europa deve possedere di fare le proprie scelte, basate sui propri valori, rispettando le proprie regole", e dovrebbe inoltre "contribuire a renderci tutti tecno-ottimisti".

Un secondo fondamentale progetto è significativo per l'ADM in Europa: il Digital Services Act (DSA) annunciato nell"Agenda per l'Europa' di Von der Leyen, e concepito per rimpiazzare la direttiva E-Commerce in vigore fin dal 2000. Il suo scopo è "aggiornare le nostre regole di responsabilità e sicurezza per le piattaforme, i servizi e i prodotti digitali, e completare il nostro Mercato Unico Digitale" — alimentando così un fondamentale dibattito sul ruolo dell'ADM nelle politiche di moderazione dei contenuti, in quelle sulla responsabilità degli intermediari, e per la libertà di espressione[13] più in genere.

Una problematizzazione esplicita dei sistemi di ADM è reperibile nella Risoluzione approvata dalla Commissione sul mercato interno e la protezione dei consumatori del Parlamento UE, e in una Raccomandazione "sugli impatti in termini di diritti umani dei sistemi algoritmici" del Comitato dei Ministri del Consiglio d'Europa.

Il Consiglio d'Europa (CoE), in particolare, ha mostrato nel corso dell'ultimo anno di ricoprire un ruolo sempre più importante nel dibattito su come governare l'AI, e anche se il suo reale impatto sulle proposte di regolamentazione resta da dimostrare, si potrebbe sostenere che funga da "guardiano" dei diritti umani. Ciò è maggiormente evidente nella Raccomandazione, 'Unboxing Artificial Intelligence: 10 steps to protect Human Rights', della Commissaria per i diritti umani del CoE, Dunja Mijatović, e nei lavori dell'Ad Hoc Committe on AI (CAHAI) creato nel settembre 2019.

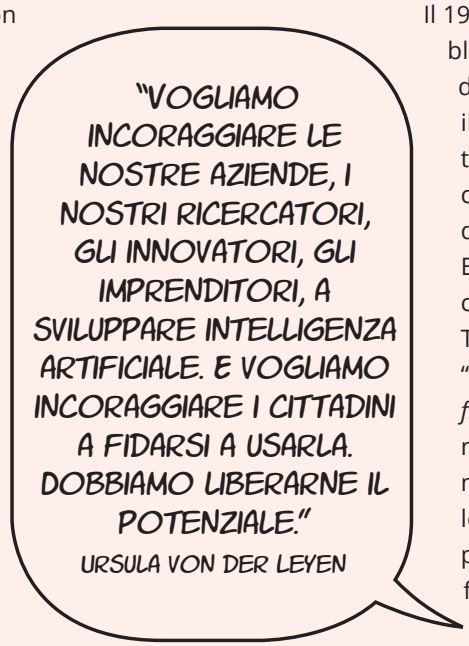
13 Considerazioni dettagliate e raccomandazioni circa l'uso di sistemi di ADM nel contesto del DSA possono essere reperite tra i risultati di un altro progetto di AlgorithmWatch, 'Governing Platforms', https://algorithmwatch.org/en/project/governing-platforms/

DIVERSI OSSERVATORI VEDONO UNA FONDAMENTALE TENSIONE TRA GLI IMPERATIVI AFFARISTICI E QUELLI DEI DIRITTI NEL MODO IN CUI LE ISTITUZIONI EUROPEE, E IN PARTICOLAR MODO LA COMMISSIONE, STANNO CONCEPENDO LE LORO RIFLESSIONI E PROPOSTE SU AI E ADM.

Diversi osservatori vedono una fondamentale tensione tra gli imperativi affaristici e quelli dei diritti nel modo in cui le istituzioni europee, e in particolar modo la Commissione, stanno concependo le loro riflessioni e proposte su AI e ADM.

Da un lato, l'Europa vorrebbe "incrementare l'uso, e la domanda, di dati e prodotti e servizi resi possibili dai dati in tutto il Mercato Unico", diventando così "leader" nelle applicazioni commerciali dell'AI, e potenziando la competitività delle aziende UE a fronte delle crescenti pressioni dai rivali negli Stati Uniti e in Cina. Tutto questo è ancora più significativo per l'ADM, visto che l'assunto di fondo è che, attraverso un'economia "*data-agile*", l'UE "possa diventare un modello e una guida per una società che grazie ai dati riesce a prendere decisioni migliori — negli affari come nel settore pubblico". Come scrive il White Paper sull'AI, "i dati sono la linfa vitale dello sviluppo economico".

Dall'altro lato, tuttavia, l'analisi automatica di dati riguardanti la salute, il lavoro e i diritti sociali di un cittadino può portare a decisioni dagli esiti discriminatori e iniqui. Questo "lato oscuro" degli algoritmi nei processi di decisione è affrontato nel pacchetto di strumenti regolatori (*toolbox*) proposto dall'UE attraverso una serie di principi. Nel caso di sistemi ad alto rischio, dovrebbero essere imposte regole per garantire che i processi decisionali automatici siano compatibili con i diritti umani e con effettive tutele democratiche. Questo è un approccio che le istituzioni UE definiscono "umano-centrico" e insieme unico, perché fondamentalmente opposto a quelli applicati negli Stati Uniti (dove il criterio è il profitto) e in Cina (dove invece lo sono la sicurezza nazionale e la sorveglianza di massa).



Tuttavia, sono emersi dubbi circa le reali possibilità per l'Europa di raggiungere entrambi gli obiettivi allo stesso tempo. Il caso del riconoscimento facciale è quello più comunemente portato a esempio: anche se, come mostrato da questo rapporto, disponiamo ormai di svariate prove che ne testimoniano l'adozione opaca e incontrollata in buona parte dei paesi membri, la Commissione UE non ha reagito rapidamente e in modo deciso per proteggere i diritti dei cittadini europei. Come rivelato in un leaked draft del 'Libro bianco sull'AI' della Commissione, l'UE aveva considerato l'idea di mettere al bando i sistemi di "identificazione biometrica da remoto" nei luoghi pubblici, prima di cambiare idea all'ultimo minuto e preferivi, al contrario, la chiamata a un "ampio dibattito" sulla materia.

Nel frattempo, diversi progetti finanziati dall'UE continuano a promuovere controversi sistemi di ADM per il controllo dei confini, anche dotati di riconoscimento facciale.

Scelte politiche e dibattiti politici

/ La Strategia europea sui dati e il Libro bianco sull'AI

Se da un lato il promesso pacchetto di norme "per un approccio europeo coordinato sulle implicazioni umane ed etiche dell'intelligenza artificiale", annunciato nell'"Agenda per l'Europa' di Von der Leyen, non è stato poi effettivamente proposto entro "i primi cento giorni in carica", dall'altro la Commissione UE ha quantomeno pubblicato una serie di documenti che stabiliscono l'insieme di principi e idee che dovrebbero informarlo.

Il 19 febbraio 2020, infatti, sono stati pubblicati insieme la 'Strategia europei sui dati' ('A European Strategy for Data') e il già menzionato 'Libro bianco sull'intelligenza artificiale', i due documenti che illustrano i principi fondamentali dell'approccio strategico dell'Unione Europea all'AI (e ai sistemi di ADM, anche se non esplicitamente menzionati). Tra questi, spiccano l'idea di mettere "al primo posto" le persone (*people first,* inteso come sinonimo di una "tecnologia al servizio delle persone"), la neutralità tecnologica (nessuna tecnologia è buona o cattiva di per sé; ciò, piuttosto, dipende dall'uso che se ne fa) e, naturalmente, la "sovranità tecnologica" e il tecno-ottimismo. Nelle parole di Von der Leyen: "Vogliamo incoraggiare le nostre aziende, i nostri ricercatori, gli innovatori, gli imprenditori, a sviluppare intelligenza artificiale. E vogliamo incoraggiare i cittadini a fidarsi a usarla. Dobbiamo liberarne il potenziale".

L'idea di fondo è che a nuove tecnologie non debbano accompagnarsi nuovi valori. Il "nuovo mondo digitale" ipotizzato dall'amministrazione Von der Leyen intende infatti rispettare a pieno i diritti umani e civili. "Eccellenza" e "fiducia", evidenziati fin dal titolo stesso del Libro bianco, vengono considerate i due pilastri su cui un modello europeo per l'AI può e deve poggiare, differenziandolo sia da quello statunitense che da quello cinese.

E tuttavia, all'ambizione non corrisponde adeguato livello di dettaglio nel Libro bianco. Per esempio, il White Paper della Commissione illustra un approccio alla regolamentazione dell'AI basato sul rischio, in cui le regole sono proporzionali alla severità dell'impatto di un sistema di "AI" sulle vite dei cittadini. "Per i casi ad alto rischio, come nella sanità, nella sorveglianza (*policing*) e nei trasporti", si legge, "i sistemi di AI dovrebbero essere trasparenti, tracciabili e garantire supervisione umana". Tra le salvaguardie che dovrebbero venire predisposte figurano anche una fase di test e la certificazione degli algoritmi adottati, con l'obiettivo di renderle comuni come per "cosmetici, vetture o giocattoli". Al contrario, "sistemi meno rischiosi" prevedono unicamente l'osservanza di schemi di controllo qualità volontari: "Gli operatori economici coinvolti si vedrebbero di conseguenza riconosciuti un marchio di qualità per le loro applicazioni di AI".

Diversi critici, tuttavia, hanno sottolineato che la definizione stessa di "rischio" nel Libro bianco è insieme circolare e troppo vaga, finendo per consentire a diversi sistemi di ADM altamente impattanti di insinuarsi tra le maglie dello schema di regole proposto[14].

I commenti[15] pervenuti nel corso della consultazione pubblica tenutasi tra febbraio e giugno 2020 evidenziano quanto questa idea sia controversa, con il 42,5% delle risposte a favore di "requisiti obbligatori" per le sole "applicazioni di AI ad alto rischio", e il 30,6% che invece esprimono dubbi al riguardo.

Inoltre, non vi è nel documento descrizione chiara di un meccanismo per fare poi realmente rispettare i requisiti esposti, né quantomeno di un processo che porti a costruirlo davvero.

Le conseguenze sono immediatamente visibili per le tecnologie biometriche, e per il riconoscimento facciale in particolare. Su questo, il Libro bianco propone una distinzione tra "autenticazione" biometrica, considerata non controversa (per esempio, il riconoscimento facciale per sbloccare uno smartphone), e "identificazione" biometrica da remoto (per esempio, per identificare manifestanti in protesta in una pubblica piazza), che può sollevare serie preoccupazioni in termini di privacy e diritti umani.

Solo i casi rientranti in quest'ultima categoria, tuttavia, sarebbero considerati problematici secondo lo schema di regole ipotizzato dall'UE. Eppure il FAQ del Libro bianco afferma: "questa è la forma più intrusiva di riconoscimento facciale, e sarebbe in linea di principio proibita nell'Unione Europea", a meno che la sua adozione non sia materia di "sostanziale interesse pubblico".

Il documento esplicativo aggiunge anche che "consentire il riconoscimento facciale è attualmente l'eccezione", ma i risultati del nostro rapporto lo contraddicono: il riconoscimento facciale sembra piuttosto stare diventando rapidamente la norma. Una versione preliminare del Libro bianco, ottenuta dagli organi di stampa, pareva avere riconosciuto l'urgenza del problema, includendo l'idea di una moratoria dai tre ai cinque anni degli usi dal vivo del riconoscimento facciale in luoghi pubblici, fino a quando — e se — non si fosse trovato un modo per coniugarli con tutele democratiche.

Appena prima della pubblicazione della bozza ufficiale del Libro bianco, perfino la commissaria europea Margrethe Vestager aveva chiesto che tali usi del riconoscimento facciale venissero messi in "pausa".

Tuttavia, "funzionari della Commissione" hanno immediatamente aggiunto che la "pausa" non avrebbe impedito ai governi nazionali di farvi ricorso secondo le regole esistenti. E alla fine, l'ultima versione del Libro bianco non contiene alcuna menzione di una moratoria per il riconoscimento facciale, preferendo al contrario chiedere "un ampio dibat-

14 "Per fare due esempi: VioGén, un sistema di ADM per prevedere casi di violenza di genere, e Ghostwriter, un'applicazione per identificare frodi durante gli esami, sfuggirebbero alla forma più severa di regolamentazione, nonostante presentino rischi enormi" (https://algorithmwatch.org/en/response-european-commission-ai-consultation/)

15 "In totale, sono stati ricevuti 1215 contributi, dei quali 352 per conto di aziende o di organizzazioni e associazioni commerciali, 406 di cittadini (il 92% dell'UE), 152 per conto di istituti accademici e di ricerca, e 73 da autorità pubbliche. La voce della società civile è stata rappresentata da 160 rispondenti (tra cui 9 associazioni per i consumatori, 129 organizzazioni non governative e 22 sindacati); 72 partecipanti hanno risposto identificandosi nella categoria "altro". Commenti sono giunti "da ogni parte del mondo", inclusi paesi come "India, Cina, Giappone, Siria, Iraq, Brasile, Messico, Canada, Stati Uniti e Gran Bretagna". (Dal 'Summary Report' della consultazione pubblica, reperibile da qui: https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence)

tito europeo sulle specifiche condizioni, ammesso esistano, che potrebbero giustificarne" l'uso per scopi di identificazione biometrica dal vivo. Il Libro bianco vi include i principi di giustificazione e proporzionalità, l'esistenza di tutele democratiche, e il rispetto dei diritti umani.

In tutto il documento, i rischi associati alle tecnologie di AI vengono di norma definiti "potenziali", mentre i benefici vengono al contrario dipinti come realissimi e immediatamente realizzabili. Ciò ha portato molti[16], nella comunità dei diritti umani, ad affermare che la narrazione complessiva del Libro bianco contenga una preoccupante inversione di priorità nelle politiche dell'UE, finendo per privilegiare l'obiettivo di una maggiore competitività globale a quello di proteggere i diritti umani.

E ciononostante, i documenti UE sollevano anche alcune questioni fondamentali. Per esempio, l'interoperabilità dei sistemi di AI, e la creazione di una rete di centri di ricerca specializzati in applicazioni dell'AI che ambiscano all'"eccellenza" e alla creazione di competenze adeguate.

L'obiettivo è "attrarre oltre 20 miliardi l'anno di investimenti complessivi in AI a livello UE, nel corso del prossimo decennio".

16 Per esempio Access Now (https://www.accessnow.org/cms/assets/uploads/2020/05/EU-white-paper-consultation_AccessNow_May2020.pdf), AI Now (https://ainowinstitute.org/ai-now-comments-to-eu-whitepaper-on-ai.pdf), EDRi (https://edri.org/our-work/can-the-eu-make-ai-trustworthy-no-but-they-can-make-it-just/) — e AlgotithmWatch (https://algorithmwatch.org/en/response-european-commission-ai-consultation/).

Il Libro bianco sembra anche risentire di un certo determinismo tecnologico. "È essenziale", vi si legge infatti, "che le amministrazioni pubbliche, gli ospedali, i servizi di utilità e trasporti, i supervisori finanziari, e altri settori di pubblico interesse comincino rapidamente ad adottare prodotti e servizi basati sull'AI nelle loro attività. Un'attenzione specifica sarà posta nei settori della sanità e dei trasporti, dove la tecnologia è matura per un'adozione su larga scala".

Tuttavia, non è chiaro se suggerire una frettolosa adozione di soluzioni di ADM in ogni sfera dell'attività umana sia compatibile con gli sforzi ipotizzati dalla stessa Commissione UE per affrontare le sfide strutturali poste dai sistemi di ADM a diritti ed equità.

/ La risoluzione del Parlamento UE sull'ADM e la protezione dei consumatori

Una Risoluzione, approvata dal Parlamento europeo nel febbraio 2020, ha affrontato più nello specifico i sistemi di ADM nel contesto della protezione dei consumatori. La Risoluzione sottolinea correttamente che "sistemi algoritmici complessi e processi di decision-making automatizzato stanno venendo creati a passo spedito", e che "le opportunità e le sfide poste da queste tecnologie sono molteplici e riguardano virtualmente ogni settore". Il testo ricorda inoltra il bisogno di "una valutazione dello schema legislativo attualmente in vigore nell'UE", così da giudicare se "sia in grado di far fronte all'emergere di AI e ADM".

Chiedendo un "approccio comune a livello UE per lo sviluppo di processi di ADM", la Risoluzione dettaglia i diversi requisiti che ogni sistema di decisioni automatizzate dovrebbe possedere per rimanere all'interno della sfera dei valori europei. I consumatori dovrebbero essere "debitamente informati" circa i modi in cui gli algoritmi influenzano le loro vite, e dovrebbero sempre ottenere l'intervento di un essere umano con poteri decisionali, così che le scelte che coinvolgono forme di automazione possano essere controllate e, se necessario, corrette. I consumatori dovrebbero poi venire anche informati "quando i prezzi di beni o servizi sono stati personalizzati sulla base di ADM e profilazione delle abitudini di consumo".

Nel ricordare alla Commissione UE che è necessario un approccio basato sui rischi che sia attentamente costruito, la Risoluzione evidenzia che le tutele predisposte devono tenere in considerazione che i sistemi di ADM "possono evolvere e agire in modi non contemplati quando inizialmente

SE L'AI È DAVVERO UNA RIVOLUZIONE CHE RICHIEDE UN PACCHETTO LEGISLATIVO APPOSITO, I RAPPRESENTANTI DEMOCRATICAMENTE ELETTI VOGLIONO AVERE VOCE IN CAPITOLO.

introdotti sul mercato", e che le responsabilità non sono sempre facilmente attribuibili nel caso in cui la loro adozione risulti dannosa.

La Risoluzione richiama inoltre l'art. 22 del GDPR, quando nota che un essere umano deve sempre poter intervenire nel caso in cui "siano in gioco legittimi interessi pubblici", e che sia sempre un essere umano a dover essere, in ultima analisi, responsabile delle decisioni riguardanti "le professioni mediche, legali e contabili, nonché il settore bancario". In particolare, una "adeguata" valutazione del rischio dovrebbe precedere ogni forma di automazione di servizi professionali.

Infine, la Risoluzione elenca una serie dettagliata di requisiti di qualità e trasparenza per il governo dei dati: tra gli altri, "l'importanza di usare solo dati di elevata qualità e privi di bias, così da migliorare l'output dei sistemi algoritmici e aumentare fiducia e accettazione nel consumatore"; usare "algoritmi spiegabili e privi di bias"; e il bisogno di una "struttura per la revisione" che consenta ai consumatori "di ottenere revisione umana e fare appello alle decisioni automatizzate che siano definitive e permanenti".

/ Prendere l'iniziativa tramite il "diritto di iniziativa" del Parlamento UE

Nel suo discorso inaugurale, Von der Leyen ha espresso il proprio chiaro supporto per un "diritto di iniziativa" per il Parlamento europeo. "Nel caso in cui quest'Aula, agendo con il mandato della maggioranza dei suoi Membri, adotti Risoluzioni che richiedono alla Commissione di formulare proposte di legge", ha affermato Von der Leyen, "mi impegno a rispondere con un atto legislativo nel pieno rispetto dei principi di proporzionalità, sussidiarietà e di "legiferare meglio" (*better law-making*)".

Se l'AI è davvero una rivoluzione che richiede un pacchetto legislativo apposito, in arrivo — pare — nel primo trimestre del 2021, i rappresentanti democraticamente eletti vogliono avere voce in capitolo. Il che, sommato all'intento espresso da Von der Leyen di aumentarne la capacità legislativa, potrebbe perfino produrre ciò che Politico ha chiamato "il momento del Parlamento". A conferma, diverse commissione parlamentari sono già all'opera sulla stesura di diversi rapporti.

Ciascun rapporto indaga uno specifico aspetto dell'automazione nelle politiche pubbliche; e per quanto ciascuno di essi sia concepito per informare la legislazione sull'AI prossima ventura, ciò è in buona parte rilevante anche per l'ADM.

Per esempio, nel suo 'Framework of ethical aspects of artificial intelligence, robotics and related technologies', la Commissione Giuridica del Parlamento UE chiede l'istituzione di una "Agenzia europea per l'AI" e, al tempo stesso, di una rete di autorità nazionali di supervisione, in ogni stato membro, così da assicurare che decisioni automatizzate attinenti la sfera morale siano, e restino, etiche.

In 'Intellectual property rights for the development of artificial intelligence technologies', la stessa commissione illustra la sua visione per il futuro dell'automazione nella proprietà intellettuale. La bozza del rapporto, per esempio,

afferma che "i metodi matematici non sono brevettabili, a meno che non costituiscano invenzioni di natura tecnica", sostenendo però poi, in tema di trasparenza algoritmica, che "il *reverse engineering* non è che un'eccezione alla regola del segreto commerciale".

Il rapporto si spinge fino a chiedersi come proteggere "creazioni tecniche e artistiche generate da AI, così da incoraggiare questa forma di creazione", immaginando che "certe opere generate da AI possano essere considerate equivalenti a opere intellettuali, e dunque protette da copyright".

Infine, in un terzo documento ('Artificial Intelligence and civil liability'), la Commissione dettaglia un "approccio per la gestione del rischio" in termini di responsabilità civile delle tecnologie di AI. Secondo tale approccio, "la parte meglio in grado di controllare e gestire i rischi correlati alla tecnologia è considerata strettamente responsabile, e unico viatico a contenziosi".

Importanti principi riguardanti l'uso di ADM nel sistema penale possono essere reperiti nel rapporto della Commissione Libertà Civili, Giustizia e Affari Interni intitolato 'Artificial Intelligence in criminal law and its use by the police and judicial authorities in criminale matters'. Dopo un elenco dettagliato di reali usi correnti dell'"AI" — in realtà, di sistemi di ADM — da parte delle forze di polizia[17], la Commissione "considera necessario creare un insieme di regole chiaro ed equo per assegnare responsabilità legali per le potenziali conseguenze negative prodotte da queste avanzate tecnologie digitali".

Ne dettaglia poi alcune caratteristiche: nessuna decisione può essere interamente automatizzata[18], gli algoritmi devono essere spiegabili in modi che siano "intelligibili agli utenti", e deve essere condotta una "obbligatoria valutazione di impatto sui diritti fondamentali (...) di ogni sistema di AI in uso dalle forze dell'ordine o dalla giustizia", prima della sua adozione, più un altrettanto obbligatorio "auditing periodico di tutti i sistemi di AI usati dalle forze dell'ordine e dalla giustizia, per testare e valutare i sistemi algoritmici mentre sono in funzione".

Nel rapporto viene anche poi chiesta una moratoria per l'uso di tecnologie di riconoscimento facciale da parte delle forze dell'ordine, quantomeno "finché gli standard tecnici non potranno essere considerati pienamente rispettosi dei diritti fondamentali, i risultati che ne derivano non saranno non discriminatori, e non ci sarà pubblica fiducia nella necessità e proporzionalità dell'adozione di tali tecnologie".

L'obiettivo è di pervenire a una maggiore trasparenza nell'uso dei sistemi di ADM, e insieme suggerire agli stati membri di fornire una "comprensione esaustiva" dei sistemi di AI adottati dalle forze dell'ordine e nel sistema giudiziario, e — sulla scia dell'idea di un "registro pubblico" — dettagliare "le tipologie di strumenti in uso, le tipologie di crimini a cui si applicano, e le aziende i cui strumenti stiano venendo utilizzati".

La Commissione Cultura e Istruzione e quella per le politiche industriali erano a loro volta al lavoro sui rispettivi rapporti, al momento della stesura di questo capitolo.

Tutte queste iniziative hanno portato, il 18 giugno 2020, alla creazione di una Commissione Speciale sull'intelligenza artificiale nell'era digitale ('Special Committee on Artificial Intelligence in a Digital Age', o 'AIDA'). Composta di 33 membri, e con una durata iniziale di dodici mesi, intende analizzare "il futuro impatto" dell'AI sull'economia dell'Unione, e "in particolare sulle competenze, l'impiego, il fintech, l'educazione, la salute, il turismo, l'agricoltura, l'ambiente, la difesa, l'industria, l'energia e l'e-government".

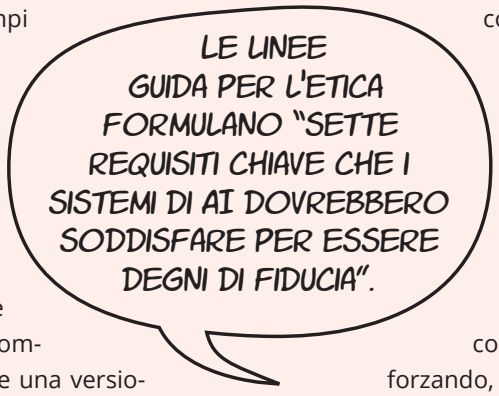
/ L'High-Level Expert Group sull'AI e la AI Alliance

Nel 2018, la Commissione Europea ha predisposto la creazione di un High-Level Expert Group (HLEG) sull'AI, un comitato composto da 52 esperti che ha l'obiettivo di contribuire all'implementazione di una strategia europea sull'intelligenza artificiale, identificare gli obiettivi che dovrebbero venire osservati per ottenere una "AI degna di fiducia" ("*trustworthy AI*") e, in qualità di comitato direttivo della più ampia "Alleanza per l'AI" (*AI Alliance*), creare una piattaforma aperta e multi-stakeholder — composta attualmente da oltre quat-

17 A pagina 5, il rapporto afferma: "Le applicazioni dell'AI usate dalle forze dell'ordine includono: le tecnologie di riconoscimento facciale, il riconoscimento automatico di numeri di targa, l'identificazione di un parlante, l'identificazione di una parlata, tecnologie per la lettura delle labbra, tecnologie di sorveglianza sonora (per esempio, algoritmi di riconoscimento di uno sparo da arma da fuoco), l'autonoma ricerca e analisi nei database identificati, tecnologie predittive (predictive policing e analisi degli hotspot criminali), strumenti di riconoscimento del comportamento, strumenti autonomi per identificare frodi finanziarie e fondi di provenienza terroristica, social media monitorino (scraping e data haversting a caccia di connessioni), IMSI (International Mobile Subscriber Identity) catcher, e strumenti di sorveglianza automatizzata che incorporano diverse funzionalità di riconoscimento (come per esempio i termoscanner e gli strumenti di riconoscimento del battito cardiaco)".

18 "In contesti giudiziari e di polizia, la decisione finale deve sempre essere presa da un essere umano." (p. 6)

tromila membri — per fornire più ampi input al lavoro del gruppo di esperti sull'AI.



Dopo la pubblicazione della prima bozza delle linee guida per l'etica in una intelligenza artificiale "degna di fiducia" ('AI Ethics Guidelines for Trustworthy AI') a dicembre 2018, che ha raccolto feedback da oltre 500 commentatori, ne è stata pubblicata anche una versione aggiornata, ad aprile 2019. Il documento propone un "approccio umano-centrico" che abbia lo scopo di ottenere soluzioni di AI legali, etiche e solide attraverso ogni fase del loro ciclo di vita. Resta, in ogni caso, uno schema volontario senza raccomandazioni concrete e applicabili in termini di operatività, implementazione ed *enforcement*.

La società civile e le organizzazioni per la tutela del consumatore e dei diritti, commentando il documento, hanno chiesto la traduzione delle linee guida in diritti tangibili. Per esempio, la no profit per i diritti digitali Access Now, membro del HLEG, ha insistito affinché il passo successivo della Commissione sia chiarire come i diversi portatori di interesse possano testare, applicare, migliorare, approvare e implementare forme di "trustworthy AI", riconoscendo al contempo il bisogno di determinare i limiti oltre il quale l'approccio europeo non consente di spingersi.

In un op-ed, altri due membri del HLEG hanno affermato che il gruppo aveva "lavorato per un anno e mezzo, solo per poi vedere le proprie proposte perlopiù ignorate o menzionate solo di passaggio" nella bozza finale stilata dalla Commissione Europea. In aggiunta, i due hanno anche scritto che, dato che il gruppo era stato creato con un primo obiettivo di identificare rischi e limiti invalicabili (*red lines*) per l'AI, alcuni suoi membri avevano indicato armi automatiche, sistemi di valutazione a punti per la cittadinanza (*citizen scoring*) e identificazione automatica tramite riconoscimento facciale tra le implementazioni dell'AI da evitare. Tuttavia, soggetti rappresentanti degli interessi dell'industria di settore, dominanti nella composizione del gruppo[19], sono riusciti a ottenere la rimozione di questi principi dalla bozza del documento prima che fosse pubblicata.

Questo sbilanciamento a favore dei potenziali benefici dell'ADM può essere osservato anche nelle pagine del secondo importante documento prodotto dal HLEG. Nel suo rapporto 'Policy and investment recommendations for trustworhy AI in Europe', reso pubblico a giugno 2019, si trovano infatti 33 raccomandazioni mirate a "indirizzare un'AI degna di fiducia nella direzione della sostenibilità, della crescita e della competitività, così come in quella dell'inclusione — rinforzando, beneficiando e proteggendo gli esseri umani". Il documento è principalmente una chiamata a incrementare utilizzo e portata dell'AI nel settore privato come in quello pubblico, investendo in strumenti e applicazioni "che aiutino le fasce più vulnerabili della popolazione", così da "non lasciare indietro nessuno".

Ciononostante, e pur date le legittime perplessità, entrambe le linee guida esprimono anche considerazioni critiche, avanzando alcune richieste per i sistemi di ADM. Per esempio, le linee guida per l'etica formulano "sette requisiti chiave che i sistemi di AI dovrebbero soddisfare per essere *trustworthy*", cioè meritare la fiducia dei cittadini. Non solo: forniscono anche una guida all'implementazione pratica di ogni requisito — autonomia e controllo umano, sicurezza e solidità tecnica, privacy e governo dei dati, trasparenza, diversità, equità e principio di non discriminazione, benessere sociale e ambientale, e responsabilità (*accountability*).

Le linee guida contengono poi anche indicazioni più concrete, in quella che definisce "Trustworthy AI assessment list", un vero e proprio elenco per pervenire all'operatività dei principi appena dettagliati. L'obiettivo è che venga adottato "nello sviluppo, nell'adozione o nell'utilizzo dei sistemi di AI", e adattato "ai casi d'uso specifici in cui il sistema sta venendo applicato".

La lista include diverse questioni associate al rischio di abusare dei diritti umani tramite sistemi di ADM. Tra di esse, figurano la mancanza di autonomia e controllo umano, problemi di natura tecnica e di sicurezza informatica, l'incapacità di evitare bias iniqui o fornire accesso universale a tali sistemi, nonché la mancanza di un reale accesso ai dati che processano.

Contestualmente, questa bozza di elenco inclusa nelle linee guida fornisce suggerimenti utili per chiunque intenda adottare sistemi di ADM. Per esempio, chiede "una valutazione di impatto sui diritti fondamentali quando può esserci un impatto negativo" su quei diritti. Chiede anche se

19 Il gruppo era composto da 24 rappresentanti del settore commerciale, 17 accademici, 5 organizzazioni della società civile, e 6 ulteriori membri, tra cui l'Agenzia dell'Unione Europea per i diritti fondamentali.

siano stati predisposti "meccanismi specifici di controllo e vigilanza" nei casi riguardanti sistemi di AI "capaci di auto-apprendimento o autonomi", e se esistano processi per "assicurare la qualità e l'integrità dei vostri dati".

Considerazioni dettagliate riguardano inoltre alcune questioni fondamentali per i sistemi di ADM, come la loro trasparenza e spiegabilità (*explainability*). Alcune delle domande incluse sono "fino a che punto le decisioni, e dunque i risultati di scelte prese da sistemi di AI, sono comprensibili?" e "fino a che punto una decisione del sistema influenza i processi decisionali dell'organizzazione?" Si tratta di questioni estremamente rilevanti per valutare i rischi posti dalla loro adozione.

In aggiunta, le raccomandazioni sulle politiche di investimento prevedono la determinazione di limiti (*red lines*) attraverso un "dialogo" istituzionalizzato "sulle politiche per l'AI con i portatori di interesse che ne devono fronteggiare l'impatto", inclusi gli esperti della società civile. Sollecitano inoltre a "mettere al bando sistemi massivi di valutazioni a punteggio (*scoring*) degli individui basati su AI, così come definiti nelle linee guida per l'etica, e [a] stabilire regole estremamente chiare e stringenti per la sorveglianza a scopi di sicurezza nazionale, e per altri scopi che siano stati dichiarati di interesse pubblico o nazionale". Il divieto includerebbe anche le tecnologie di identificazione biometrica e di profilazione.

Altro aspetto rilevante per i sistemi di decision-making automatizzato è che il documento afferma anche che "definire chiaramente se, quando e come l'AI può essere usata (...) sarà cruciale per riuscire ad avere una AI davvero degna di fiducia", ammonendo che "qualunque forma di valutazione a punteggio (*scoring*) del cittadino può portare a perdita di autonomia individuale e mettere a repentaglio il principio di non discriminazione", e "di conseguenza tali valutazioni dovrebbero essere usate solo nel caso ve ne sia chiara giustificazione, all'interno di un insieme di misure proporzionate ed eque". Si aggiunge infine che "la trasparenza non basta a prevenire discriminazioni o fare sì che sia garantita l'equità". Ciò significa che la possibilità di rifiutarsi di essere assoggettati a un meccanismo di scoring dovrebbe essere una scelta sempre disponibile, idealmente senza che ne derivi alcuno svantaggio per il cittadino.

Da un lato, il documento riconosce "che, pur arrecando sostanziali benefici all'individuo e alla società, i sistemi di AI pongono anche alcuni rischi, e possono avere impatti negativi, tra cui alcuni che sono difficili da prevedere, identificare o misurare (per esempio, concernenti la democrazia, lo stato di diritto e la giustizia distributiva, nonché la stessa mente umana)". Dall'altro, tuttavia, il gruppo afferma che "regole innecessariamente prescrittive dovrebbero essere evitate".

Nel luglio del 2020, il gruppo sull'AI ha anche introdotto la versione finale della sua lista per la valutazione dell'intelligenza artificiale meritevole di fiducia ('Assessment List for Trustworthy Artificial Intelligence', o 'ALTAI'), ottenuta al termine di un processo deliberativo che ha incluso 350 portatori di interesse.

La lista, la cui adozione è interamente volontaria e priva di ogni implicazione in termini di legge, ha l'obiettivo di tradurre i sette requisiti descritti nelle linee guida per l'etica del HLEG sull'AI in azione. L'intento è fornire uno strumento di auto-valutazione a chiunque intenda costruire soluzioni di AI che siano compatibili con i valori dell'UE — per esempio, progettisti e sviluppatori di sistemi di AI, scienziati dei dati, funzionari e specialisti del *procurement*, funzionari legali e per la *compliance*.

/ Il Consiglio d'Europa: come proteggere i diritti umani nell'ADM

In aggiunta all'Ad Hoc Committee on Artificial Intelligence (CAHAI), creato a settembre 2019, il Comitato dei Ministri del Consiglio d'Europa[20] ha a sua volta reso pubblico un framework dotato di sostanza e profondità.

Immaginata come strumento per la determinazione di standard, la sua 'Raccomandazione agli stati membri sugli impatti dei sistemi algoritmici sui diritti umani' ('Recommendation to Member States on the human rights impacts of algorithmic systems') descrive[21] le "significative sfide" che

20 Il Consiglio d'Europa (CoE) è insieme "un organo governativo in cui gli approcci nazionali ai problemi europei vengono discussi su base paritaria e un forum di risposte collettive a tali sfide". I suoi ambiti di operatività includono "gli aspetti politici dell'integrazione europea, la tutela delle istituzioni democratiche, dello stato di diritto e dei diritti umani — in altre parole, tutti i problemi che richiedano soluzioni concertate a livello europeo". Anche se le raccomandazioni ai governi degli stati membri non sono vincolanti, in alcuni casi il Comitato dei ministri del CoE può chiedere a quei governi di informarlo circa ogni azione da loro intrapresa in risposta a tali raccomandazioni (Art. 15b dello Statuto). I rapporti tra il CoE e l'UE sono dettagliati 1) nel Compendio di Testi che governa le relazioni tra CoE e UE, e 2) nel Memorandum of Understanding tra CoE e UE, cfr. https://rm.coe.int/1680306052

21 Sotto la supervisione dello Steering Committee on Media and Information Society (CDMSI) e preparato dal Committee of Experts on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence (MSI-AUT).

si presentano in congiunzione con l'emergere di tali sistemi, e con il nostro "crescente affidamento" a essi. Tali sfide sono anche definite rilevanti "per le società democratiche e lo stato di diritto".

Il documento, sottoposto a un periodo di consultazione pubblica in cui ha ricevuto commenti dettagliati da diverse organizzazioni della società civile, va ben oltre il Libro bianco della Commissione UE, in tema di salvaguardia di valori e diritti umani.

La Raccomandazione analizza attentamente gli effetti e le sempre nuove configurazioni dei sistemi algoritmici (Appendice A), concentrandosi su tutte le fasi del processo di costruzione di un algoritmo — a dire: il *procurement*, la progettazione, lo sviluppo, la sua adozione e mantenimento.

Pur seguendo in linea generale l'approccio "umano-centrico" all'AI delle linee guida HLEG, il documento delinea "obblighi" operativi "per gli Stati" (Appendice B), così come responsabilità per gli operatori del settore privato (Appendice C). In più, la Raccomandazione vi aggiunge principi quali "l'autodeterminazione informativa"[22], elenca dettagliati suggerimenti per costruire meccanismi di *accountability* e rimedi efficaci, e chiede l'istituzione di valutazioni di impatto sui diritti umani.

Pur riconoscendo chiaramente "il significativo potenziale delle tecnologie digitali di fare fronte alle sfide sociali, produrre innovazione benefica per la società, e sviluppo economico", il documento invita anche nel contempo alla cautela. L'intento è assicurare che quei sistemi non perpetuino, deliberatamente o accidentalmente, "discriminazioni razziali, di genere, e altri disequilibri sociali e nella forza lavoro che ancora non sono stati eliminati dalle nostre società".

Al contrario, i sistemi algoritmici dovrebbero essere usati in modo proattivo e attento per combattere tali disuguaglianze, e prestare "attenzione ai bisogni e alle voci dei gruppi più vulnerabili".

Anche più significativo, tuttavia, è che la Raccomandazione identifichi rischi potenzialmente più elevati per i diritti umani nel caso in cui siano gli stati membri a usare sistemi algoritmici per fornire servizi pubblici. Dato che è impossibile per un cittadino sottrarvisi, quantomeno senza che ciò comporti conseguenze negative, c'è bisogno di precauzioni e tutele per l'uso di sistemi di ADM nel governo e nell'amministrazione pubblica .

La Raccomandazione affronta anche i conflitti e le sfide derivanti dalle partnership tra pubblico e privato ("né chiaramente pubbliche né chiaramente private") in un ampio spettro di utilizzi.

Le raccomandazioni per i governi degli stati membri includono l'abbandono e il rifiuto di processi e sistemi di ADM nel caso in cui "il controllo e la supervisione umana diventino impraticabili", o quando mettano a repentaglio i diritti umani; e l'adozione di sistemi di ADM se e solo se trasparenza, responsabilità, legalità, e la protezione dei diritti umani possono essere garantite "in ogni fase del processo". Inoltre, il controllo e la valutazione di tali sistemi dovrebbero essere "costanti", "inclusive e trasparenti", ricavate da una interlocuzione con tutti i portatori di interesse rilevanti, così come da un'analisi dell'impatto ambientale e di altre potenziali esternalità su "popolazioni e ambienti".

Nell'appendice A, il Consiglio d'Europa fornisce anche una definizione di cosa intenda per sistemi algoritmici "ad alto rischio" che potrebbe essere d'ispirazione per altri organi e istituzioni. Più nello specifico, la Raccomandazione stabilisce che "il termine "ad alto rischio" si applica in riferimento all'uso di sistemi algoritmici in processi e decisioni che possono produrre serie conseguenze individuali, o in situazioni in cui la mancanza di alternative comporti una probabilità particolarmente elevata di violazione dei diritti umani, incluso tramite l'introduzione o l'amplificazione di ingiustizie distributive".

Il documento, che non ha richiesto approvazione all'unanimità, non è vincolante.

/ Regole per i contenuti terroristici online

Dopo un lungo periodo di apatia, la regolamentazione per prevenire la diffusione di contenuti terroristici ha trovato un nuovo slancio nel corso del 2020. Se dovesse ancora includere la previsione di strumenti automatizzati e proattivi per identificare e rimuovere contenuti online, questi molto probabilmente ricadrebbero tra i casi regolati dall'art. 22 del GDPR.

Come scrive il Garante europeo per la protezione dei dati personali (European Data Protection Supervisor, o 'EDPS'):

22 "Gli Stati dovrebbero assicurare che il design, lo sviluppo e il mantenimento durante l'operatività dei sistemi algoritmici forniscano agli utenti modalità per essere preventivamente informati circa il trattamento dei dati a essi correlati (inclusi gli scopi e i possibili risultati), nonché di avere il controllo dei loro dati, incluso attraverso forme di interoperabilità", si legge nella Sezione 2.1 dell'Appendice B.

"dato che gli strumenti automatizzati, così come predisposti nella Proposta, possono portare non solo alla rimozione e alla conservazione di contenuti (e relativi dati) concernenti l'uploader, ma anche, in ultima analisi, a indagini penali, tali strumenti possono avere significative conseguenze a livello personale, impattando sul diritto alla libera espressione e ponendo una rilevante sfida al godimento di diritti e libertà". Di conseguenza, per questi strumenti si applica l'art. 22(2).

Inoltre, ed è cruciale, il pacchetto normativo richiederebbe tutele ben più sostanziali di quelle attualmente previste dalla Commissione. Come spiega l'organizzazione per i diritti digitali European Digital Rigths (EDRi), "la proposta di regolamentazione per i contenuti terroristici richiede significativi aggiustamenti per essere all'altezza dei valori dell'Unione, e proteggere i diritti fondamentali e le libertà dei cittadini".

Una prima serie di forti critiche alla proposta iniziale è venuta da gruppi della società civile, commissioni del Parlamento europeo, incluse opinioni e analisi dell'Agenzia UE per i diritti fondamentali e la stessa EDRi, così come da un rapporto co-firmato da tre Special Rapporteurs delle Nazioni Unite. Insieme, ne hanno sottolineato le possibili minacce in termini di libertà di espressione e informazione, libertà e pluralismo dei media, libertà di iniziativa economica e protezione dei dati personali.

Tra gli aspetti critici, figurano una insufficiente definizione di "contenuto terroristico", la portata della regolamentazione (attualmente, sono inclusi contenuti a scopo educativo e giornalistico), le "misure proattive" sopra menzionate, una mancanza di controllo giudiziario, obblighi di trasparenza insufficienti per le forze dell'ordine, e l'assenza di tutele "in casi in cui c'è ragione di ritenere che vi sia un impatto sui diritti umani" (EDRi 2019).

L'EDPS sottolinea che tali "adeguate tutele" dovrebbero includere il diritto di ottenere un intervento umano e il diritto alla spiegazione delle decisioni ottenute tramite mezzi automatizzati (EDRi 2019).

Sebbene le tutele suggerite o richieste abbiano trovato poi spazio nella bozza di rapporto del Parlamento europeo sulla proposta, resta ancora da stabilire chi avrà l'ultima parola, in vista del voto finale. Durante i triloghi a porte chiuse tra Parlamento, Commissione e Consiglio europeo (cominciati a ottobre 2019), un documento fuoriuscito (*leaked*) sostiene che siano possibili unicamente aggiustamenti di lieve entità.

Supervisione e regolamentazione

/ Prime decisioni sul rispetto del GDPR da parte di sistemi di ADM

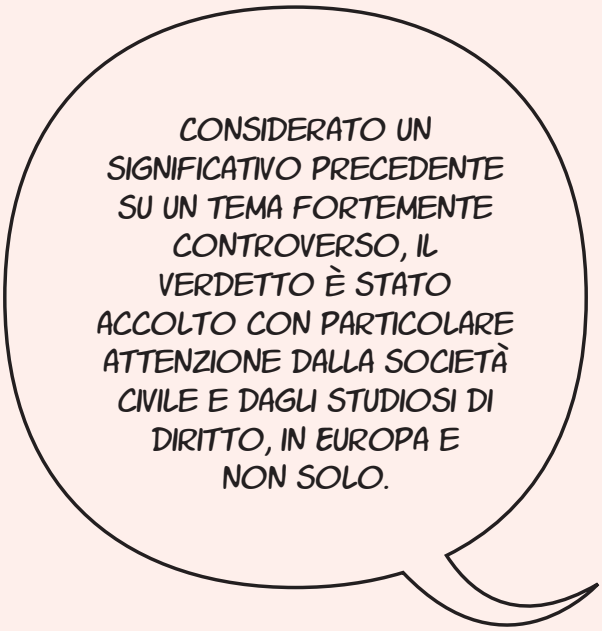
"Sebbene non ci sia stato un ampio dibattito sul riconoscimento facciale durante il passaggio delle negoziazioni sul GDPR e sulla direttiva per la protezione dei dati personali in uso dalle forze dell'ordine, le norme sono state costruite in modo tale da adattarsi nel tempo all'evolvere delle tecnologie. [...] È ora, mentre discute l'etica dell'AI e il suo bisogno di regole, che l'Europa deve determinare se — e se mai — le tecnologie di riconoscimento facciale possano essere consentite in una società democratica. Solo se la risposta è affermativa, e solo allora, potremo rivolgerci alla domanda su quali tutele e sistemi di accountability predisporre" — Wojciech Wiewiórowski, Garante europeo della protezione dei dati

"Il riconoscimento facciale è un meccanismo biometrico particolarmente intrusivo, che comporta seri rischi di abusi della privacy e delle libertà civili delle persone coinvolte" — (CNIL 2019)

Rispetto al precedente rapporto del progetto 'Automating Society', abbiamo testimoniato i primi casi di sentenze e multe correlate a violazioni delle regole emanate dalle Autorità per la protezione dei dati (DPA) nazionali sulla base del GDPR. I casi di studio che stiamo per illustrare, tuttavia, mostrano anche i limiti di applicazione pratica del GDPR in relazione all'art. 22 sui sistemi di ADM, e come stiano costringendo i Garanti nazionali a emettere giudizi caso per caso.

In Svezia, per esempio, un progetto di riconoscimento facciale in fase di sperimentazione in una classe scolastica per un limitato periodo di tempo è stato ritenuto in violazione di diversi requisiti delle norme sulla protezione dei dati (in particolare modo gli artt. 2(14) e 9(2) del GDPR). (European Data Protection Board 2019)

Un tentativo simile risulta attualmente sospeso dopo che la Commission Nationale de l'Informatique et des Libertés (CNIL) francese ha sollevato le proprie preoccupazioni circa l'idea, di due licei, di introdurre tecnologia di riconoscimento facciale in partnership con il colosso statunitense Cisco.



Il suo giudizio tuttavia non è vincolante, e l'azione legale che ne è risultata è ancora in corso[23].

Non è richiesta alcuna previa autorizzazione da parte delle Authority affinché sia possibile condurre simili sperimentazioni, dato che il consenso degli utenti è comunemente considerato sufficiente per processare dati biometrici. Eppure, nel caso svedese, non lo è stato, per via del mancato bilanciamento di poteri tra il *data controller*, titolare del trattamento dei dati, e il *data subject*, o interessato. Sono state dunque ritenute necessarie un'adeguata valutazione di impatto e una previa consultazione con la DPA nazionale.

Lo ha confermato anche l'EDPS:

"Il consenso dovrebbe essere esplicito e dato liberamente, informato e specifico. Eppure non c'è dubbio che una persona non possa esercitare né un opt out, né tantomeno un opt in, qualora necessiti di accesso a luoghi pubblici che sono monitorati da tecnologie di riconoscimento facciale. [...] Infine, la rispondenza di questa tecnologia a principi come la minimizzazione dei dati e la protezione dei dati by design è fortemente in dubbio. Il riconoscimento facciale non è mai stato completamente accurato, e ciò ha serie conseguenze per gli individui erroneamente identificati come criminali o altro. [...] Sarebbe un errore, tuttavia, concentrarsi solo sulle questioni riguardanti la privacy. Questa in una società democratica è invece, al fondamento, una questione etica." (EDPS 2019)

Access Now ha commentato:

"Ora che viene sviluppato un numero crescente di progetti di riconoscimento facciale, osserviamo che il GDPR già fornisce utili tutele in termini di diritti umani, che possono essere richiamate qualora vi siano raccolta e utilizzo illeciti di dati sensibili, per esempio biometrici. Ma le aspettative irresponsabili e spesso infondate createsi intorno all'efficienza di queste tecnologie, oltre agli interessi economici che le motivano, potrebbero condurre a tentativi, da parte di governi centrali e locali, di aggirare la legge".

/ Il riconoscimento facciale usato dalla polizia del Galles del Sud è illegale

Nel corso del 2020, la Gran Bretagna ha testimoniato una prima importante applicazione della Law Enforcement Directive (2016/80, o 'Direttiva Polizia'[24]) circa l'uso di tecnologie di riconoscimento facciale in spazi pubblici da parte delle forze dell'ordine. Considerato un significativo precedente su un tema fortemente controverso, il verdetto è stato accolto con particolare attenzione dalla società civile e dagli studiosi di diritto, in Europa e non solo [25].

Il caso è stato portato in tribunale da Ed Bridges, un 37enne di Cardiff, che aveva affermato che il suo volte fosse stato scannerizzato senza il proprio consenso sia durante lo shopping natalizio del 2017, sia a una protesta pacifica contro le armi da fuoco un anno dopo.

La corte aveva in prima battuta confermato l'uso di "tecnologie di riconoscimento facciale automatico" (*automated facial recognition*, o "AFR") da parte della polizia del Galles del Sud, dichiarandolo legale e proporzionato. Ma il verdetto è stato appellato da Liberty, un gruppo per i diritti civili, e la Corte d'Appello d'Inghilterra e Galles ha rovesciato il rigetto dell'Alta Corte, dichiarando l'uso di AFR illegale l'11 agosto 2020.

23 Si veda il capitolo sulla Francia sull'edizione integrale del rapporto, e (Kaylaki 2019).

24 La 'Direttiva polizia' ('Law Enforcement Directive'), in vigore dal maggio 2018, "concerne il trattamento di dati personali per scopi di polizia — che non ricadono all'interno del campo di applicazione del GDPR" (https://www.dataprotection.ie/en/organisations/law-enforcement-directive)

25 La sentenza è stata emessa il 4 settembre 2019 dall'Alta Corte riunita a Cardiff nel caso Bridges v. the South Wales Politce (High Court of Justice 2019)

Nel suo verdetto, contrario alla polizia del Galles del Sud per tre dei cinque capi di imputazione (*grounds*), la Corte d'Appello giudica infatti di avere trovato "fondamentali lacune" nello schema normativo vigente sull'uso di AFR, concludendo che la sua adozione non soddisfasse il principio di proporzionalità e, inoltre, che non fosse stata condotta una adeguata valutazione di impatto in termini di protezione dei dati (Data Protection Impact Assessment, DPIA), risultandone mancanti diversi passaggi.

La corte non ha, tuttavia, stabilito che il sistema stesse producendo risultati discriminatori, sulla base di genere o razza, dato che la polizia non aveva raccolto prove sufficienti a formulare un giudizio al riguardo[26]. Ciononostante, la Corte ha sentito il bisogno di aggiungere una considerazione degna di nota: "Dato che il riconoscimento facciale automatico è una tecnologia nuova e controversa, ci aspettiamo che tutte le forze di polizia che intendono usarla in futuro si assicurino di avere fatto tutto ciò che è ragionevolmente possibile per garantire che il software non contenga bias razziali o di genere".

Dopo il verdetto, Liberty ha chiesto alla polizia del Galles del Sud e ad altre forze di polizia di abbandonare l'uso di AFR.

ADM in pratica: gestione e controllo dei confini

Mentre la Commissione UE e i relativi portatori di interesse erano impegnati a dibattere se regolare o mettere al bando il riconoscimento facciale, svariate sperimentazioni ne erano già in corso in tutta Europa.

Questa sezione analizza il legame, cruciale e spesso trascurato, tra tecnologie biometriche e sistemi di gestione dei confini dell'UE, mostrando chiaramente come tecnologie che possono produrre risultati discriminatori vengano ugualmente applicate agli individui — per esempio, i migranti — che già scontano maggiormente le discriminazioni esistenti.

/ Riconoscimento facciale e uso di dati biometrici in politiche e prassi dell'UE

Nel corso dell'ultimo anno, il riconoscimento facciale e altre tipologie di tecnologie di identificazione biometrica hanno suscitato crescenti attenzioni da parte di governi, dell'UE, della società civile, e di organizzazioni a difesa dei diritti, in particolare modo riguardo al loro uso da parte delle forze dell'ordine e per la gestione dei confini.

Nel 2019, la Commissione europea ha dato a un consorzio di agenzie pubbliche il compito di "pervenire a una mappa della situazione attuale circa l'uso di tecnologie di riconoscimento facciale in indagini penali in tutti gli stati membri", così da procedere "verso un possibile scambio di dati facciali". Alla società di consulenza Deloitte è stato dunque chiesto di condurre uno studio di fattibilità per un'espansione del sistema Prüm di immagini facciali. Prüm è un sistema applicato a livello europeo che connette tra loro DNA, impronte digitali, e database di immatricolazione dei veicoli, così da consentire ricerche incrociate. La preoccupazione è che un database di volti a livello europeo possa condurre a forme di sorveglianza pervasiva, ingiustificata o illegale.

/ Sistemi di gestione dei confini senza confini

Come riportato nell'edizione precedente di questo rapporto, l'implementazione del sistema complessivo, interoperabile e "intelligente" di gestione dei confini UE, inizialmente proposto dalla Commissione già nel 2013, è in dirittura d'arrivo. Sebbene i nuovi sistemi annunciati (EES, ETIAS[27], ECRIS-TCN[28]) non saranno operativi che nel 2022, la regolamentazione sul sistema ESS (Entry/Exit System) ha già intro-

26 La polizia ha affermato di non avere avuto accesso alla composizione demografica dei dataset utilizzati per il training dell'algoritmo utilizzato, "Neoface". La Corte nota che "resta il fatto, tuttavia, che la SWP (la polizia del Galles del Sud, ndr) non ha mai tentato di verificare da sé, né direttamente né attraverso verifica indipendente, che il software adottato nel caso in esame non contenesse inaccettabili distorsioni in termini razziali o di genere".

27 ETIAS (EU Travel Information and Authorisation System), è il nuovo sistema per la gestione dei confini UE sviluppato da eu-LISA per cittadini "*visa-waiver*", cioè che non necessitano di visto all'ingresso. "Le informazioni immesse durante il processo di applicazione saranno incrociate automaticamente con i database UE esistenti (Eurodac, SIS e VIS), i sistemi futuri EES e ECRIS-TCN, e i database Interpol rilevanti. Ciò renderà possibili sistemi avanzati di verifica di rischi alla sicurezza, alla salute pubblica e di immigrazione irregolare". (ETIAS 2019)

28 L'European Criminal Records Information System — Third Country Nationals (ECRIS-TCN), che dovrà essere sviluppato da eu-LISA, sarà un sistema centralizzato per integrare l'attuale database UE di dati di rilevanza penale (ECRIS) con informazioni su cittadini non-UE condannati nell'Unione Europea.

dotto il concetto delle immagini facciali come identificativi biometrici, nonché, per la prima volta nell'impianto normativo europeo, l'uso di tecnologie di riconoscimento facciale per scopi di autenticazione (*verification*)[29].

L'Agenzia UE per i diritti fondamentali (European Fundamental Rights Agency, o "FRA") ha confermato le modifiche: "ci si attende un'introduzione più sistematica del trattamento di immagini facciali nei sistemi adottati su larga scala a livello UE in tema di asilo, migrazioni, e sicurezza [...] una volta che i necessari passaggi legali e tecnici siano stati completati".

Secondo Ana Maria Ruginis Andrei, di eu-LISA (European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice), questa nuova e ampliata infrastruttura di interoperabilità è stata "assemblata in modo da ottenere il motore adatto a combattere con successo le minacce alla sicurezza interna, controllare efficacemente i flussi migratori e superare gli attuali punti ciechi concernenti la gestione dell'identità". In pratica, ciò significa "detenere le impronte digitali, le immagini facciali, e altri dati personali di fino a 300 milioni di cittadini di paesi non appartenenti all'UE, sommando i dati da cinque sistemi diversi". (Campbell 2020)

/ ETIAS: controlli automatizzati di sicurezza al confine

L'European Travel Information and Authorization System (ETIAS), non ancora operativo mentre scriviamo, sarà condiviso da diversi database per automatizzare i controlli di sicurezza di viaggiatori non appartenenti all'UE (e che non necessitano di un visto) prima che arrivino in Europa.

Il sistema raccoglierà e analizzerà insiemi di dati per ottenere una avanzata "verifica di potenziali rischi correlati alla sicurezza o a migranti irregolari". (ETIAS 2020) Gli obiettivi sono "facilitare i controlli al confine; evitare burocrazia e ritardi per i viaggiatori che si presentano ai confini; assicurare una valutazione coordinata ed equilibrata dei rischi presentati da cittadini di paesi terzi". (ETIAS 2000)

Ann-Charlotte Nygård, a capo dell'unità di "Technical Assistance and Capacity Building" della FRA, vede due specifici rischi concernenti il sistema ETIAS: "il primo è l'uso di dati che potrebbero portare alla discriminazione involontaria di certi gruppi, per esempio se un candidato proviene da una particolare etnia ad alto rischio di migrazione; il secondo riguarda valutazioni dei rischi in termini di sicurezza che siano ricavate da condanne pregresse nel proprio paese d'origine. Alcune di queste condanne pregresse, infatti, potrebbero essere considerate irragionevoli in Europa, come per esempio quelle, in alcuni paesi, a persone di orientamento LGBT. Per evitarlo, [...] gli algoritmi devono essere sottoposti ad auditing, garantendo così che non discriminino; un auditing di questo tipo dovrebbe anche coinvolgere esperti di diverse discipline". (Nygård 2019)

/ iBorderCtrl: riconoscimento facciale e *risk scoring* ai confini

iBorderCtrl è stato un progetto che coinvolgeva le agenzie di sicurezza di Ungheria, Lettonia e Grecia, e che aveva l'obiettivo di "consentire controlli ai confini più rapidi ed esaustivi per cittadini di paesi terzi che abbiano intenzione di attraversare un confine di terra di un paese membro dell'UE". iBorderCtrl faceva ricorso a tecnologie di riconoscimento facciale, una macchina della verità, e a un sistema di valutazione a punteggio (*scoring system*) per sollecitare l'intervento di un agente in carne e ossa nel caso in cui il sistema avesse concluso che un certo individuo è pericoloso, oppure che il suo diritto di ingresso nel paese di destinazione è discutibile.

Il progetto si è concluso nell'agosto del 2019 con risultati contraddittori, nell'ottica di una potenziale implementazione in tutta l'UE.

Sebbene "resti da definirsi l'utilizzo futuro del sistema o di sue parti", la pagina dei "risultati" (*outcomes*) del progetto intravede "la possibilità di integrare funzionalità simili nel nuovo sistema ETIAS, e di estendere la possibilità di trasferire le procedure per i controlli al confine direttamente lì dove si trovano i viaggiatori (autobus, auto, treni, etc.)".

Tuttavia, i moduli a cui ciò si riferisce non sono stati specificati, e i correlati strumenti di ADM sottoposti a test non sono stati analizzati pubblicamente.

Allo stesso tempo la FAQ del progetto conferma che il sistema sperimentato non è da considerarsi "attualmente adatto all'adozione ai confini (...) a causa della sua natura di prototipo e dell'infrastruttura tecnologica esistente a livello UE". Ciò significa che "per un uso da parte delle autorità ai confini sarebbero richiesti futuri sviluppi e un'integrazione al sistema europeo vigente".

29 L'ESS entrerà in funzione nel primo trimestre del 2022, ed ETIAS vi farà seguito entro la fine del 2022. Entrambi sono concepiti come "game changer nel settore della giustizia europea e degli affari interni".

In particolare, mentre l'iBorderCtrl Consortium è stato in grado di dimostrare, in linea di principio, la funzionalità di tali tecnologie per i controlli ai confini, è altrettanto chiaro che questioni di natura etica, legale e sociale debbano ancora venire affrontate prima di qualunque reale adozione.

/ Progetti collegati a Horizon 2020

Diversi progetti, dedicati alla sperimentazione e allo sviluppo di nuovi sistemi e tecnologie per la gestione e il controllo dei confini, vi hanno fatto seguito nell'ambito del programma Horizon2020. I progetti sono elencati sul sito di CORDIS, che fornisce informazioni su tutte le attività di ricerca promosse dall'UE e a quest'ultimo collegate.

Il sito mostra che ci sono 38 progetti attualmente in corso nella sezione tematica 'H2020-EU-3.7.3. — Strengthen security through border management', cioè "rafforzare la sicurezza attraverso la gestione dei confini". Il più ampio programma di cui è parte, 'Secure societies — Protecting freedom and security of Europe and its citizens', vanta un budget complessivo di quasi 1,7 miliardi di euro, e finanzia 350 progetti. L'obiettivo è fare fronte all'"insicurezza, derivante da crimini, violenze, terrorismo, disastri naturali o causati dall'uomo, cyber-attacchi o violazioni della privacy, e altre forme di disordini economici e sociali che sempre più affliggono i cittadini" attraverso progetti che in larga parte consistono nello sviluppo di nuovi sistemi tecnologici a base di AI e ADM.

Alcuni progetti si sono già conclusi e/o loro applicazioni sono già in uso: per esempio, FastPass, ABC4EU, MOBILEPASS e EFFISEC. Ciascuno di essi ha indagato i requisiti necessari a ottenere "un controllo automatizzato dei confini (Automated Border Control, o "ABC") che sia integrato e interoperabile", e dunque applicabile ai sistemi di identificazione e ai gate "intelligenti" nei diversi contesti di attraversamento di un confine.

TRESPASS è un progetto ancora in corso, cominciato nel giugno 2018 e il cui termine è previsto a novembre 2021. La UE contribuisce a finanziarlo con quasi otto milioni di euro, e i coordinatori di iBorderCtrl (così come di FLYSEC e XPDITE) mirano a "sfruttare i risultati e i concetti implementati e sperimentati" da IborderCtrl ed "espanderli in una soluzione di sicurezza multimodale basata sul rischio per i controlli al confine, dotata di una solida base legale ed etica". (Horizon2020 2019)

Il progetto ha lo scopo dichiarato di far evolvere i controlli di sicurezza agli attraversamenti dei confini da un sistema vecchio e superato basato sulle regole (*rule-based*), a un nuovo sistema basato sui rischi (*risk-based*). Ciò significa includere l'applicazione di tecnologie biometriche e sensori, di un sistema di gestione basato sul rischio, e l'uso di modelli adeguati a valutare identità, averi, capacità, e intenti. L'obiettivo è consentire controlli attraverso "link a sistemi antiquati e database esterni come VIS/SIS/PNR", oltre a raccogliere dati da tutte le fonti sopra menzionate per scopi di sicurezza.

L'UE VI CONTRIBUISCE CON POCO MENO DI 8,2 MILIONI DI EURO, PER SVILUPPARE "METODI PIÙ PROGREDITI DI SORVEGLIANZA DEI CONFINI" E CONTRASTARE L'IMMIGRAZIONE IRREGOLARE.

Un altro progetto-pilota, FOLDOUT, è cominciato nel settembre 2018 e si concluderà a febbraio 2022. L'UE vi contribuisce con poco meno di 8,2 milioni di euro, per sviluppare "metodi più progrediti di sorveglianza dei confini" e contrastare l'immigrazione irregolare. Il progetto ha un focus specifico sul "riconoscimento di individui anche in mezzo al fogliame più denso e in condizioni climatiche estreme", combinando "diversi sensori e tecnologie, e fondendole intelligentemente in una piattaforma di identificazione intelligente e affidabile" che suggerisca possibili scenari di intervento. Sperimentazioni sono in corso in Bulgaria, con modelli dimostrativi reperibili in Grecia, Finlandia e nella Guyana Francese.

MIRROR, o 'Migration-Related Risks caused by misconceptions of Opportunities and Requirement', è invece un progetto iniziato a giugno 2019 e il cui termine è previsto a maggio 2022. L'UE vi contribuisce con poco più di cinque milioni di euro, con lo scopo di "comprendere come l'Europa sia percepita all'estero, riconoscere discrepanze tra percezione e realtà, individuare esempi di manipolazione dei media, e sviluppare le abilità necessarie a ribattere alle rappresentazioni scorrette e alle minacce in termini di sicurezza che ne derivano". Sulla base di una "analisi dei rischi specificamente dedicata alle percezioni, il progetto MIRROR affiancherà metodi per l'analisi di testi automatizzati, dei social media e di diversi media (*multimedia*) agli studi empirici", così da sviluppare "tecnologie e nozioni operative [che siano] validate attentamente insieme agli organi deputati al controllo dei confini e ai policy-maker, per esempio tramite sperimentazioni".

Tra gli ulteriori progetti che si sono già conclusi, ma che sono ugualmente menzionati, è incluso TABULA RASA ('Trusted Biometrics under Spoofing Attacks'), cominciato nel novembre 2010 e conclusosi nell'aprile 2014. TABULA RASA analizzava "i punti deboli dei software di identificazione biometrica in relazione alla loro vulnerabilità a falsificazione dell'identità (*spoofing*), che ne diminuisce l'efficacia". Tra i progetti è incluso anche Bodega, cominciato a giugno 2015 e terminato nell'ottobre 2018, che cercava di comprendere come utilizzare "l'expertise del fattore umano" nella "introduzione di sistemi più intelligenti di controlli ai confini, come gate automatici e sistemi di sicurezza self-service basati su dati biometrici".

Riferimenti bibliografici

Access Now (2019): Comments on the draft recommendation of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems https://www.accessnow.org/cms/assets/uploads/2019/10/Submission-on-CoE-recommendation-on-the-human-rights-impacts-of-algorithmic-systems-21.pdf

AlgorithmWatch (2020): Our response to the European Commission's consultation on AI https://algorithmwatch.org/en/response-european-commission-ai-consultation/

Campbell, Zach/Jones, Chris (2020): Leaked Reports Show EU Police Are Planning a Pan-European Network of Facial Recognition Databases https://theintercept.com/2020/02/21/eu-facial-recognition-database/

CNIL (2019): French privacy regulator finds facial recognition gates in schools illegal https://www.biometricupdate.com/201910/french-privacy-regulator-finds-facial-recognition-gates-in-schools-illegal

Coeckelbergh, Mark / Metzinger, Thomas(2020): Europe needs more guts when it comes to AI ethics https://background.tagesspiegel.de/digitalisierung/europe-needs-more-guts-when-it-comes-to-ai-ethics

Committee of Ministers (2020): Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154

Commissioner for Human Rights (2020): Unboxing artificial intelligence: 10 steps to protect human rights https://www.coe.int/en/web/commissioner/-/unboxing-artificial-intelligence-10-steps-to-protect-human-rights

Committee on Legal Affairs (2020): Draft Report: With recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies https://www.europarl.europa.eu/doceo/document/JURI-PR-650508_EN.pdf

Committee on Legal Affairs (2020): Artificial Intelligence and Civil Liability https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf

Committee on Legal Affairs (2020): Draft Report: On intellectual property rights for the development of artificial intelligence technologies https://www.europarl.europa.eu/doceo/document/JURI-PR-650527_EN.pdf

Committee on Civil Liberties, Justice and Home Affairs (2020): Draft Report: On artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters https://www.europarl.europa.eu/doceo/document/LIBE-PR-652625_EN.pdf

Delcker, Janosch(2020): Decoded: Drawing the battle lines — Ghost work — Parliament's moment https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-drawing-the-battle-lines-ghost-work-parliaments-moment/?utm_source=POLITICO.EU&utm_campaign=5a7d137f82-EMAIL_CAMPAIGN_2020_09_09_08_59&utm_medium=email&utm_term=0_10959edeb5-5a7d137f82-190607820

Data Protection Commission(2020): Law enforcement directive https://www.dataprotection.ie/en/organisations/law-enforcement-directive

EDRi (2019): FRA and EDPS: Terrorist Content Regulation requires improvement for fundamental rights https://edri.org/our-work/fra-edps-terrorist-content-regulation-fundamental-rights-terreg/

GDPR (Art 22): Automated individual decision-making, including profiling https://gdpr-info.eu/art-22-gdpr/

European Commission (2018): White paper: On Artificial Intelligence - A European approach to excellence and trust https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission (2020): A European data strategy https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

European Commission (2020): Shaping Europe's digital future – Questions and Answers https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_264

European Commission (2020): White Paper on Artificial Intelligence: Public consultation towards a European approach for excellence and trust https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence

European Commission (2018): Security Union: A European Travel Information and Authorisation System - Questions & Answers https://ec.europa.eu/commission/presscorner/detail/en/MEMO_18_4362

European Data Protection Board (2019): Facial recognition in school renders Sweden's first GDPR fine https://edpb.europa.eu/news/national-news/2019/facial-recognition-school-renders-swedens-first-gdpr-fine_en

European Parliament (2020): Artificial intelligence: EU must ensure a fair and safe use for consumers https://www.europarl.europa.eu/news/en/press-room/20200120IPR70622/artificial-intelligence-eu-must-ensure-a-fair-and-safe-use-for-consumers

European Parliament (2020): On automated decision-making processes: ensuring consumer protection and free movement of goods and services https://www.europarl.europa.eu/doceo/document/B-9-2020-0094_EN.pdf

European Data Protection Supervisor (2019): Facial recognition: A solution in search of a problem? https://edps.europa.eu/press-publications/press-news/blog/facial-recognition-solution-search-problem_de

ETIAS (2020): European Travel Information and Authorisation System (ETIAS) https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders/etias_en

ETIAS (2019): European Travel Information and Authorisation System (ETIAS) https://www.eulisa.europa.eu/Publications/Information%20Material/Leaflet%20ETIAS.pdf

Horizon2020 (2019): robusT Risk basEd Screening and alert System for PASSengers and luggage https://cordis.europa.eu/project/id/787120/reporting

High Court of Justice (2019): Bridges v. the South Wales Police https://www.judiciary.uk/wp-content/uploads/2019/09/bridges-swp-judgment-Final03-09-19-1.pdf

High-Level Expert Group on Artificial Intelligence (2020): Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Hunton Andrew Kurth (2020): UK Court of Appeal Finds Automated Facial Recognition Technology Unlawful in Bridges v South Wales Police https://www.huntonprivacyblog.com/2020/08/12/uk-court-of-appeal-finds-automated-facial-recognition-technology-unlawful-in-bridges-v-south-wales-police/

Kayalki, Laura (2019): French privacy watchdog says facial recognition trial in high schools is illegal https://www.politico.eu/article/french-privacy-watchdog-says-facial-recognition-trial-in-high-schools-is-illegal-privacy/

Kayser-Bril, Nicolas (2020): EU Commission publishes white paper on AI regulation 20 days before schedule, forgets regulation https://algorithmwatch.org/en/story/ai-white-paper/

Leyen, Ursula von der (2019): A Union that strives for more - My agenda for Europe https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf

Leyen, Ursula von der (2020): Paving the road to a technologically sovereign Europe https://delano.lu/d/detail/news/paving-road-technologically-sovereign-europe/209497

Leyen, Ursula von der (2020): Shaping Europe's digital future https://twitter.com/eu_commission/status/1230216379002970112?s=11

Leyen, Ursula von der (2019): Opening Statement in the European Parliament Plenary Session by Ursula von der Leyen, Candidate for President of the European Commission https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_19_4230

Nygård, (2019): The New Information Architecture as a Driver for Efficiency and Effectiveness in Internal Security https://www.eulisa.europa.eu/Publications/Reports/eu-LISA%20Annual%20Conference%20Report%202019.pdf

Sabbagh, Dan (2020): This article is more than 1 month old South Wales police lose landmark facial recognition case https://www.theguardian.com/technology/2020/aug/11/south-wales-police-lose-landmark-facial-recognition-case

South Wales Police(2020): Automated Facial Recognition https://afr.south-wales.police.uk/

Valero, Jorge (2020): Vestager: Facial recognition tech breaches EU data protection rules https://www.euractiv.com/section/digital/news/vestager-facial-recognition-tech-breaches-eu-data-protection-rules/

SVIZZERA

INCHIESTA

PAGINA 146

RICERCA

PAGINA 150

Per saperne di più, leggi “Diagnosi e cure contro il cancro” nella sezione dedicata alla “Ricerca”.

La polizia svizzera ha automatizzato la predizione del crimine, ma c'è poco da vedere

Un'analisi di tre sistemi automatici utilizzati dalla polizia e dal sistema giudiziario svizzero ne mette in luce seri problemi. Vista la mancanza di trasparenza emersa circa l'uso di tali sistemi, non è possibile valutarne gli effetti nel mondo reale.

Di Nicolas Kayser-Bril

La polizia e le autorità giudiziarie svizzere usano, secondo una stima, oltre 20 diversi sistemi automatizzati per valutare o predire comportamenti inappropriati. Polizia e giustizia sono in larga parte competenze regionali in Svizzera; ogni cantone potrebbe, di conseguenza, averne di propri.

Ne abbiamo valutati tre, basandoci su una serie di inchieste dell'emittente pubblica svizzera datate 2018.

/ Predire i furti

Il sistema Precobs è utilizzato in Svizzera fin dal 2013. Precobs viene venduto da un'azienda tedesca che non fa mistero della propria discendenza da "Minority Report", una storia di fantascienza in cui individui chiamati "precogs" sono in grado di predire alcuni crimini prima ancora che si verifichino. (La trama ruota intorno ai frequenti fallimenti dei precogs e alle conseguenti operazioni di depistaggio operate dalla polizia per occultarli)

IL SISTEMA PRECOBS È UTILIZZATO IN SVIZZERA FIN DAL 2013. PRECOBS VIENE VENDUTO DA UN'AZIENDA TEDESCA CHE NON FA MISTERO DELLA PROPRIA DISCENDENZA DA "MINORITY REPORT", UNA STORIA DI FANTASCIENZA IN CUI INDIVIDUI CHIAMATI "PRECOGS" SONO IN GRADO DI PREDIRE ALCUNI CRIMINI PRIMA ANCORA CHE SI VERIFICHINO.

Il sistema cerca di inferire i furti futuri dai dati di quelli passati, sulla base dell'assunto secondo il quale i ladri operano di norma in aree circoscritte. Se un insieme di furti viene identificato in un determinato quartiere, la teoria impone alla polizia di pattugliarlo più spesso, se vuole mettervi fine.

Tre cantoni usano Precobs: Zurigo, Argovia e Basilea Campagna, che fanno insieme quasi un terzo della popolazione svizzera. I furti sono in drastica diminuzione già dalla metà della prima decade del nuovo millennio. La polizia di Argovia si è perfino lamentata, nell'aprile 2020, che ci fossero talmente pochi furti da non consentire l'uso di Precobs.

Ma il numero di furti è diminuito in ogni cantone della Svizzera, e i tre che usano Precobs non si avvicinano nemmeno a quelli che hanno ottenuto i risultati migliori. Tra gli anni 2012-2014 (quando i furti erano al loro picco) e gli anni 2017-2019 (quando Precobs è stato utilizzato nei tre cantoni), i furti sono diminuiti in tutti i cantoni, non solo nei tre che hanno fatto uso del software. La riduzione dei furti a Zurigo e Argovia si è rivelata anzi inferiore alla media nazionale (-44%), rendendo improbabile che Precobs abbia avuto un effetto rilevante al riguardo.

Un rapporto del 2019 dell'Università di Amburgo non è riuscito a trovare prove dell'efficacia delle soluzioni di polizia predittiva (*predictive policing*), Precobs inclusa. Non è stato possibile reperire alcun documento pubblico menzionante il costo pagato dalle autorità svizzere per il sistema, ma la città di Monaco di Baviera ha speso 100,000 euro per l'installazione di Precobs — costi operativi esclusi.

/ Predire violenze contro le donne

Sei cantoni (Glarona, Lucerna, Sciaffusa, Soletta, Turgovia e Zurigo) usano il sistema Dyrias-Intimpartner per predire la probabilità che una persona venga aggredita dal proprio partner. Dyrias sta per "dynamic system for the Analysis of risk" ("sistema dinamico per l'analisi del rischio"), ed è a sua volta prodotto e venduto da un'azienda tedesca.

Secondo un'inchiesta dell'emittente pubblica svizzera SRF del 2018, Dyrias richiede al personale di polizia di rispondere a 39 domande a risposta affermativa o negativa circa ogni sospetto. Il sistema restituisce dunque un punteggio su una scala da 1 a 5, da innocuo a pericoloso. Mentre il numero totale di persone su cui è stato sperimentato Dyrias è attualmente sconosciuto, i conti di SRF mostrano che nel 2018 erano stati definiti "pericolosi" 3,000 individui (anche se l'etichetta potrebbe non essere stata applicata a seguito dell'uso di Dyrias).

L'azienda che vende Dyrias afferma che il suo software sia in grado di identificare correttamente 8 individui potenzialmente pericolosi su 10. Tuttavia, un altro studio ha analizzato i falsi positivi, gli individui cioè definiti pericolosi quando erano invece innocui, e scoperto che di 10 individui segnalati dal software, 6 avrebbero dovuto essere categorizzati al contrario tra gli innocui. In altre parole, Dyrias può vantare buoni risultati solo perché non prende rischi, e attribuisce l'etichetta "pericoloso" con facilità. (L'azienda produttrice di Dyrias contesta questi risultati.)

Perfino nel caso in cui le prestazioni di Dyrias dovessero venire migliorate, i suoi effetti resterebbero comunque impossibili da valutare. Justyna Gospodinov, co-direttrice di BIF-Frauenberatung, una organizzazione che fornisce supporto alle vittime di violenze domestiche, ha affermato ad AlgorithmWatch di non poter dire alcunché su Dyrias — nonostante la cooperazione con la polizia stia facendo progressi, e nonostante ritenga che formulare una valutazione sistematica del rischio sia una cosa positiva. Ogni qual volta prendano in carico un nuovo caso, dice infatti, i membri della sua organizzazione non sanno se il software sia stato usato o meno.

/ Predire il rischio di recidiva

Fin dal 2018, tutte le autorità giudiziarie nei cantoni della Svizzera tedesca usano ROS (acronimo per "Risikoorientierter Sanktionenvollzug", o esecuzione delle sentenze che comportano il carcere basata sul rischio). Il sistema assegna i detenuti alla categoria A in caso non presentino rischio di recidiva, alla B quando potrebbero commettere una ulteriore infrazione, e alla C quando si presenta il rischio che commettano un crimine violento. I detenuti possono essere sottoposti a test più volte, ma, in quelli successivi, possono solo passare dalla categoria A a quelle B o C, non viceversa.

Un'inchiesta di SRF ha inoltre rivelato che solo un quarto dei detenuti nella categoria C ha poi realmente commesso un ulteriore crimine dopo essere usciti di prigione (un tasso di falsi positivi del 75%), e che solo uno su cinque di quelli che hanno commesso un nuovo crimine erano nella categoria C (un tasso di falsi negativi dell'80%), sulla base di uno studio del 2013 dell'Università di Zurigo. Una nuova versione del sistema è stata rilasciata nel 2017, ma non è ancora stata analizzata.

I cantoni della Svizzera francese e italiana stanno invece lavorando a un'alternativa a ROS, che potrebbe essere pronta nel 2022. Pur se mantiene le tre categorie già viste, il loro sistema funzionerà solo integrandosi con interviste ai detenuti che stanno venendo valutati.

/ Mission: Impossible

Gli scienziati sociali sono in grado di ottenere successi considerevoli nel campo della predizione di conclusioni generali. Nel 2010, l'Ufficio federale di statistica ha predetto che la popolazione residente in Svizzera avrebbe raggiunto quota 8,5 milioni entro il 2020 (reale popolazione del 2020: 8,6 milioni). Ma nessuno scienziato potrebbe provare a predire la data di morte di un singolo individuo: la vita, semplicemente, è troppo complessa.

Da questo punto di vista, la demografia non è molto diversa dalla criminologia. Nonostante i produttori di soluzioni commerciali affermino il contrario, predire il comportamento individuale è molto probabilmente impossibile. Nel 2017, un gruppo di scienziati ha provato a fare chiarezza sulla questione una volta per tutte. Gli studiosi hanno chiesto a un gruppo di 160 ricercatori di provare a predire le prestazioni scolastiche, la probabilità di essere sfrattati, e altri quattro indicatori applicati a migliaia di teenager, sulla base di dati precisi, raccolti fin dalla nascita. Per ogni bambino erano stati resi disponibili migliaia di data point. I risultati, pubblicati ad aprile 2020, ridimensionano ogni ambizione. Non un solo team è stato in grado di predire alcun risultato con un qualche grado di accuratezza. Ma non è tutto: nemmeno i gruppi che hanno fatto ricorso a intelligenza artificiale sono riusciti a produrre predizioni migliori di chi si era invece servito solo di un pugno di variabili, utilizzando modelli statistici elementari.

Moritz Büchi, ricercatore senior all'università di Zurigo, è l'unico studioso svizzero che ha preso parte all'esperimento. In una email ad AlgorithmWatch, ha scritto che se da un lato il crimine in quanto tale non era oggetto di scrutinio, dall'altro le conoscenze ottenute attraverso l'esperimento si applicano forse alla predizione della criminalità. Questo non significa che non si debba provare a formulare predizioni, ha scritto Büchi, quanto piuttosto che trasformare le simulazioni in strumenti pronti per l'uso le ricopre di un "manto di oggettività" che potrebbe scoraggiare il pensiero critico, con conseguenze potenzialmente devastanti per le persone il cui futuro è predetto.

Precobs, che non cerca di predire il comportamento di individui specifici, non ricade nella stessa categoria, ha anche aggiunto Büchi. Più controllo potrebbe avere un effetto deterrente sui criminali. Tuttavia, l'identificazione di hotspot di criminalità si basa su serie storiche. Ciò potrebbe portare a un eccesso di controlli nelle comunità dove sono stati segnalati crimini in passato, creando così un circolo vizioso che si autoalimenta.

/ Un "effetto repressivo" sulla società svizzera

A dispetto di storie dagli esiti altalenanti, e delle prove che mostrano la quasi-impossibilità di predire comportamenti individuali, le forze dell'ordine svizzere continuano a utilizzare sistemi che affermano di poterlo, invece, fare. La loro popolarità deriva in parte dalla loro opacità. Scarsissime informazioni pubbliche esistono su Precobs, Dyrias e ROS. Le persone colpite, in stragrande maggioranza povere, hanno raramente le risorse finanziarie necessarie a opporsi ai sistemi automatizzati, dato che i loro avvocati si concentrano di norma sulla verifica dei fatti più basilari insinuati dall'accusa.

Timo Grossenbacher, il giornalista che ha investigato ROS e Dyrias per SRF nel 2018, ha detto ad AlgorithmWatch che trovare le persone affette da questi sistemi è stato "quasi impossibile". Non per mancanza di casi: ROS da solo viene usato su migliaia di detenuti ogni anno. Il problema è piuttosto l'opacità di questi sistemi, che impedisce ai watchdog della "polizia predittiva" di fare il loro lavoro, e fare dunque luce su come funzionino.

Senza una maggiore trasparenza, questi sistemi potrebbero avere un "effetto repressivo" sulla società svizzera, secondo Büchi, dell'Università di Zurigo. "Questi sistemi potrebbero scoraggiare le persone dall'esercitare i loro diritti, e portarle a modificare i loro comportamenti", ha scritto. "È una forma di obbedienza anticipata. Essendo consapevoli della possibilità di essere (ingiustamente) catturati da questi algoritmi, gli individui potrebbero tendere a diventare sempre più conformi alle norme sociali percepite. L'espressione della propria personalità e l'esplorazione di stili di vita alternativi potrebbero risultarne soppresse".

Ricerca

Di Nadja Braun Binder e Catherine Egli

Il contesto dell'ADM in Svizzera

La Svizzera è un paese spiccatamente federalista, con una pronunciata divisione dei poteri. Di conseguenza, le innovazioni tecniche nel settore pubblico sono spesso sviluppate in prima battuta nei cantoni.

L'introduzione dell'identità elettronica (eID) ne è un esempio. A livello federale, il processo legislativo richiesto per introdurre il sistema di eID non è ancora stato completato, mentre viceversa una identità elettronica ufficiale è già in funzione in un singolo cantone. Nel 2017, e all'interno della strategia cantonale sull'eGovernment in Svizzera, il canton Sciaffusa è diventato il primo a introdurre una identità digitale per i suoi residenti. Usando l'eID, i cittadini possono, tra le altre cose, fare domanda per una licenza di pesca, calcolare le imposte dovute su profitti da immobili o titoli, o richiedere una dilazione per la propria dichiarazione dei redditi.

Inoltre, ogni cittadino adulto può disporre di un profilo presso l'Autorità per la protezione dei minori e degli adulti, così che i dottori possano fare richiesta di un credito per i pazienti ospedalizzati fuori dal loro distretto. Un altro esempio, avviato a settembre 2019 all'interno di un progetto pilota nello stesso cantone, consente ai residenti di ordinare, via smartphone, estratti dal registro delle esecuzioni (Schaffhauser 2020). Anche se non è di per se stessa un processo di ADM, l'eID costituisce però un prerequisito essenziale per l'accesso a servizi pubblici digitali e, di conseguenza, potrebbe inoltre rendere più semplice l'accesso a procedure automatizzate, per esempio nel campo della tassazione. Il fatto che un singolo cantone si trovi in uno stadio di sviluppo più avanzato del progetto rispetto al governo federale è tipico, per la Svizzera.

La democrazia diretta è un altro degli elementi caratteristici dello stato svizzero. Per esempio, se il processo legislativo che porta a una eID nazionale non è stato ancora completato è a causa del referendum che dovrà tenersi sul corrispondente atto parlamentare (eID - Referendum 2020). Coloro i quali hanno chiesto il referendum non sono fondamentalmente contrari a una eID ufficiale, ma vogliono piuttosto impedire che l'emissione dell'identità digitale e la gestione di dati personali sensibili finiscano nelle mani di aziende private.

Un ulteriore elemento che va tenuto in considerazione è la buona situazione economica di cui gode la Svizzera. Ciò permette di ottenere enormi progressi in settori precisi, come le decisioni automatizzate applicate alla medicina, e in diverse aree della ricerca. Anche se la Svizzera, a causa della peculiare struttura federale e della divisione dipartimentale delle responsabilità a livello federale, non dispone di una strategia nazionale sull'AI o l'ADM, la ricerca settoriale procede a un livello che è competitivo in tutto il mondo.

Un catalogo di esempi di ADM

/ Diagnosi e cure contro il cancro

In Svizzera si sta attualmente indagando l'uso di decisioni automatizzate nella medicina, ed è questo il motivo per cui l'ADM si trova in uno stadio più avanzato di sviluppo nel campo della sanità che in altri settori. A oggi sono conosciute oltre 200 diverse tipologie di tumore, e sono disponibili quasi 120 medicinali per curarle. Ogni anno vengono fatte numerose diagnosi di cancro e, dato che ogni tumore dispone di un proprio particolare profilo di mutazioni genetiche che lo porta a crescere, ciò è fonte di problemi per gli oncologi. Anche una volta operata la diagnosi e determinata la possibile mutazione genetica, questi ultimi devono infatti ancora studiare un corpus sempre crescente di letteratu-

ra medica, per scegliere la cura più efficace. È per questo motivo che gli ospedali universitari di Ginevra sono i primi in Europa a usare lo strumento IBM Watson Health, Watson for Genomics®: per meglio vagliare le opzioni terapeutiche disponibili e per suggerire cure a pazienti affetti da cancro. I medici esaminano ancora le mutazioni genetiche, e descrivono dove e come molte di esse accadano, ma Watson for Genomics® può utilizzare quelle informazioni per compiere ricerche all'interno di un database di circa tre milioni di pubblicazioni. Da lì, il programma è poi in grado di creare un rapporto che classifica le alterazioni genetiche nel tumore di un paziente, e suggerire le più rilevanti terapie e sperimentazioni cliniche a quel tumore associate.

Fino a oggi, i medici che combattono il cancro hanno dovuto farlo da soli — con il rischio di lasciarsi sfuggire alcuni tra i possibili metodi di cura. Ora invece il software può prendere il controllo della fase di ricerca, anche se gli oncologi devono ancora valutare attentamente le liste di contenuti estrapolati dalla letteratura che il programma produce, e hanno ancora l'ultima parola su quale metodo di cura vada poi, infine, scelto. Il risultato è che Watson for Genomics® consente di risparmiare molto tempo oggi investito in analisi, e di fornire ai medici importanti informazioni aggiuntive. A Ginevra, il rapporto prodotto da questo sistema di ADM viene utilizzato durante la preparazione del "Tumor Board", una squadra multidisciplinare in cui i medici prendono nota dei trattamenti proposti da Watson for Genomics, e li discutono in una assemblea plenaria per sviluppare così insieme una strategia di cura per ciascun paziente. (Schwerzmann/ Arroyo 2019)

Sistemi di ADM sono utilizzati anche all'ospedale universitario di Zurigo, in maniera predominante — data la particolare adeguatezza a svolgere compiti ripetitivi — nella radiologia, nella patologia e, di conseguenza, per calcolare la densità mammaria. Durante una mammografia, un algoritmo analizza automaticamente le radiografie e classifica i tessuti mammari in A, B, C, o D (una classificazione internazionalmente riconosciuta per l'analisi del rischio). Analizzando il rischio sulla base della densità mammaria, l'algoritmo aiuta non poco i medici nel pervenire a una valutazione del rischio di cancro al seno, dato che la densità mammaria ne è uno dei più importanti fattori. Questo uso dell'ADM per le radiografie è già divenuto prassi comune all'ospedale universitario di Zurigo. In aggiunta, prosegue la ricerca in tema di algoritmi avanzati per l'interpretazione di immagini a ultrasuoni. (Lindner 2019)

Ciò detto, oltre un terzo dei casi di cancro al seno sfugge agli esami di screening mammografico — ed è per questo che la ricerca si occupa di indagare come i sistemi di ADM possano contribuire all'interpretazione delle immagini a ultrasuoni (US). C'è una tensione tra l'interpretazione delle immagini US mammarie e la comune mammografia digitale, che dipende in larga parte da un osservatore e richiede radiologi ben addestrati e dotati di esperienza. Per questa ragione, uno spin-off dell'ospedale universitario di Zurigo ha analizzato, a livello di ricerca, come i sistemi di ADM possano contribuire alla standardizzazione di immagini US. Per riuscirci, i ricercatori hanno simulato il processo decisionale umano tramite un sistema di breast imaging e reporting basato su dati. Questa tecnica si è rivelata molto accurata e, in futuro, l'algoritmo prodotto potrebbe venire utilizzato per imitare processi decisionali umani, al punto di diventare lo standard per il riconoscimento, l'evidenziazione e la classificazione di lesioni al seno. (Ciritisis a.o. 2019 p. 5458-5468)

/ Chatbot per l'assicurazione sociale

Alcuni cantoni fanno ricorso alle cosiddette "chatbot" per semplificare e insieme assistere la comunicazione con l'amministrazione pubblica. Una chatbot è stata sperimentata nel 2018 dall'istituto per l'assicurazione sociale del canton San Gallo ('Sozialversicherungsanstalt des Kantons St. Gallens', o 'SVA San Gallo'). La SVA San Gallo è un centro di eccellenza per ogni sorta di assicurazione sociale, inclusa la riduzione del premio per quella sanitaria.

L'assicurazione sanitaria è obbligatoria in Svizzera, e offre copertura ai residenti in caso di malattia, maternità e incidenti, offrendo a ciascuno la stessa gamma di benefici. Viene finanziata dai contributi (*premium*) dei cittadini. I premi variano a seconda dell'assicuratore, e dipendono dal luogo di residenza e dal tipo di assicurazione di cui si ha bisogno, mentre non sono collegati al reddito. Tuttavia, grazie ai sussidi distribuiti dai cantoni (riduzione del premio), i cittadini a basso reddito, i bambini e i giovani impegnati a tempo pieno nella loro istruzione o formazione pagano spesso premi ridotti. Sono i cantoni a decidere chi abbia diritto allo sconto. (FOPH 2020)

Verso la fine di ogni anno, la SVA San Gallo riceve all'incirca 80.000 domande di riduzione del premio assicurativo. Per ridurre il carico di lavoro che si accompagna a questo flusso concentrato di domande, la SVA ha sperimentato una chatbot via Facebook Messenger. L'obiettivo del progetto pilota era offrire ai clienti metodi alternativi di comunicazione. Il primo assistente digitale per l'amministrazione è stato costruito con l'idea di fornire risposte automatiche alle più ricorrenti domande circa le riduzioni del premio assicurativo. Per esempio: cosa sono le riduzioni del premio assicurativo e come posso richiederle? Ho diritto a fare domanda? Ci sono eccezioni, e come si deve procedere? Come viene calcolata, e corrisposta, una riduzione del premio? Inoltre, la chatbot era in grado di riferire i clienti ad altri servizi offerti dalla SVA San Gallo, tra cui un calcolatore per la riduzione del premio e un form di registrazione interattivo. Per quanto non sia la chatbot a prendere materialmente la decisione finale circa l'assegnazione delle riduzioni del premio assicurativo, questo strumento può comunque ridurre il numero di domande, dato che permette di informare i cittadini che non hanno diritto a sconti del fatto che la loro richiesta sarà probabilmente destinata a fallire. La chatbot ricopre anche un ruolo importante nella diffusione di contenuti informativi (Ringeisen/Bertolosi-Lehr/Demaj 2018 S.51-65).

Grazie ai feedback positivi ottenuti nella prima sperimentazione, la chatbot è stata nel 2019 integrata sul sito della SVA San Gallo, e ne è prevista anzi la graduale estensione ad altri prodotti assicurativi trattati dalla SVA stessa. È anche possibile che la chatbot sia usata per servizi collegati ai contributi versati per assicurazioni di lavoratori in pensione e familiari di lavoratori deceduti, per disabilità e per compensazioni di reddito. (IPV-Chatbot 2020)

/ ADM nel sistema penale

Il sistema svizzero di Esecuzione delle pene è basato su diversi livelli, e concede ai detenuti gradi crescenti di libertà man mano che la detenzione prosegue. Ciò lo rende un processo collaborativo tra autorità esecutive, istituzioni penali, terapeuti e servizi per la libertà vigilata. Naturalmente, i rischi di fuga e recidiva sono fattori decisivi quando si tratta di stabilire l'assegnazione di tali accresciute libertà.

Nel corso degli ultimi anni, e in riposta al fatto che molti criminali già condannati sono stati scoperti a commettere svariati ulteriori e tragici atti di violenza e abusi sessuali, è stato introdotto un sistema chiamato Risk-Oriented Sanctioning (ROS). L'obiettivo principale del sistema ROS è di

ROS SEPARA IL PERCORSO DA COMPIERSI CON I TRASGRESSORI IN QUATTRO FASI: TRIAGE, VALUTAZIONE, PIANIFICAZIONE, E PROGRESSO

prevenire le recidive armonizzando l'esecuzione delle sentenze con altre misure a vari livelli dedicate a promuovere il reinserimento nella società. ROS separa il percorso da compiersi con i trasgressori in quattro fasi: triage, valutazione, pianificazione, e progresso. Durante la fase di triage, i casi vengono classificati a seconda del bisogno di una valutazione del rischio di recidiva. Sulla base di questa classificazione viene dunque svolta una analisi, specifica a ogni caso, nella fase di valutazione. Nel corso della fase di pianificazione, i risultati ottenuti vengono trasformati in piani per comminare le sanzioni ai relativi trasgressori. I piani vengono a loro volta poi continuamente rivisti durante la fase di progresso. (ROSNET 2020)

Il triage gioca un ruolo decisivo all'inizio di questo processo, sia per il trasgressore che in termini di ADM, dato che viene svolto da un sistema di ADM chiamato Fall-Screening-Tool (Case Screening Tool, FaST). FaST categorizza automaticamente i casi analizzati nelle categorie A, B e C. Categoria A significa che non c'è alcun bisogno di valutazioni, B equivale a un generico rischio di ulteriore delinquenza, mentre la C corrisponde al rischio di crimini violenti o di natura sessuale.

La classificazione viene stabilita sulla base dei precedenti penali, e di generici fattori statistici di rischio, tra cui l'età, gli illeciti per violenze commesse prima dei 18 anni, le precedenti ammissioni presso avvocati per la gioventù, il numero di condanne pregresse, la tipologia di crimine commesso, le sentenze, la delinquenza polimorfica, il tempo trascorso senza commettere ulteriori reati dopo la scarcerazione, e la violenza domestica. Se vengono soddisfatti criteri di rischio specificamente collegati, secondo le risultanze scientifiche, a crimini violenti o di natura sessuale, allora il caso viene inserito nella categoria C. Se i criteri di rischio soddisfatti hanno una specifica connessione con più generica delinquenza, allora il caso appartiene alla categoria B. Se nessun criterio, o quasi nessuno, viene invece soddisfatto, il caso è di categoria A.

La classificazione consiste di elementi (fattori di rischio) inseriti in un formato a risposta chiusa, ciascuno dei quali è dotato di un diverso peso (punti). Se un fattore di rischio viene riscontrato, il suo punteggio è aggiunto al valore totale. Per ottenere il risultato complessivo, gli elementi pesati e confermati vengono sommati al punteggio, portando così a classificazioni A, B o C che, a loro volta, fungono da base per ulteriori valutazioni, se necessarie (fase 2).

La classificazione è svolta in modo interamente automatico dall'applicazione di ADM. Tuttavia, è importante ricordare che questa non è un'analisi di rischio, ma un modo per evidenziare i casi che maggiormente abbisognano di valutazione. (ROSNET 2020, Treuhardt/Kröger 2019 p. 76-85Treuhardt/Kröger 2018 p. 24-32)

Ciononostante, la classificazione operata in fase di triage ha un effetto su come le persone responsabili di un particolare istituto formulano le proprie decisioni, nonché su quali valutazioni debbano essere fatte. Ciò determina inoltre il cosiddetto "profilo del problema" (*problem profile*) riguardante i modi in cui sono state pianificate sentenze e misure (fase 3). Più nel dettaglio, questa forma di pianificazione definisce tutte le possibili facilitazioni per favorirne l'applicazione, come l'"open enforcement", forme di impiego in regime di "work release", e sistemi per "l'ospitalità all'esterno". Inoltre, non risulta alcuna applicazione di ADM nelle altre fasi del ROS. Il sistema FaST è, dunque, unicamente usato durante la fase di triage.

/ Predictive policing: sistemi di ADM usati dalle forze dell'ordine

In alcuni cantoni, in particolare in quelli di Basilea Campagna, Argovia e Zurigo, la polizia usa software che aiutano a prevenire crimini. Le forze dell'ordine si sono affidate alla soluzione commerciale "PRECOBS" (Pre-Crime Observation System), che viene unicamente adoperata per la previsione di furti domestici. Questa forma di criminalità relativamente comune è stata approfonditamente studiata a livello di ricerca scientifica, e le autorità di polizia detengono di norma consistenti database sulla distribuzione spaziale e temporale dei furti, nonché sulle caratteristiche del crimine commesso. Inoltre, questi crimini indicano abitualmente l'opera di un professionista, ed esibiscono di conseguenza tassi di recidiva superiori alla media. In aggiunta, i relativi modelli previsionali possono essere creati con un numero relativamente esiguo di data point. PRECOBS è, di conseguenza, basato sull'assunto per cui i ladri colpiscono svariate volte in un breve arco temporale, se hanno già avuto successo in una certa zona.

Il software è usato per andare alla ricerca di alcune regolarità nei rapporti della polizia sui furti, per esempio circa i modi in cui i colpevoli agiscono, e quando e dove colpiscono. Subito dopo, PRECOBS produce una previsione di quali saranno le aree ad accresciuto rischio di furto nelle 72 ore seguenti. Immediatamente, la polizia invia pattuglie di ricognizione nelle aree identificate. PRECOBS genera così

previsioni principalmente sulla base di scelte che devono essere inserite nel sistema, e non fa uso di metodi di machine learning. Sebbene esistano piani per ampliare l'utilizzo di PRECOBS in futuro, così da includere altri crimini (come per esempio il furto d'auto e il borseggio) e creare di conseguenza nuove funzionalità, va notato che l'uso della "polizia predittiva" (*predictive policing*) in Svizzera è attualmente limitato a un'area relativamente esigua e chiaramente delimitata del lavoro di prevenzione delle forze dell'ordine. (Blur 2017, Leese 2018 p. 57-72)

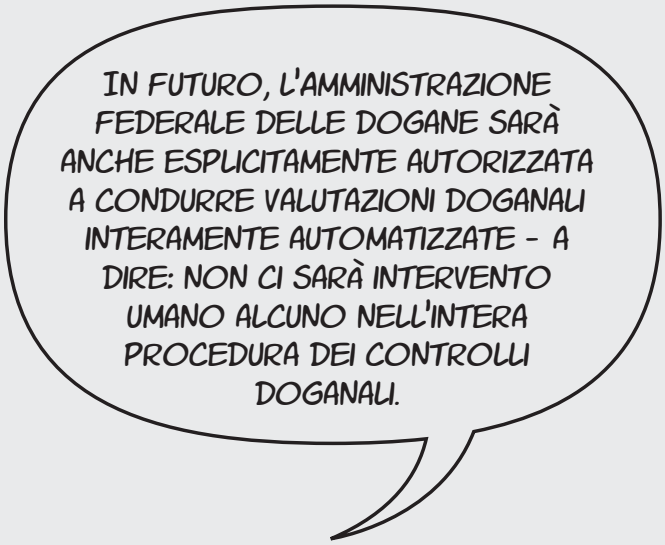
/ Operazioni doganali

A livello federale, è lecito aspettarsi un deciso ricorso all'ADM da parte dell'Amministrazione federale delle dogane, dato che il dipartimento fa già ampio uso di automazione. E infatti, la valutazione delle dichiarazioni doganali viene in larga parte eseguita per via elettronica. Il processo valutativo può essere diviso in quattro passaggi: una procedura di valutazione sommaria, l'accettazione della dichiarazione doganale, la sua verifica, e l'ispezione, a cui fa seguito una decisione finale.

La procedura di valutazione sommaria rappresenta un controllo di plausibilità e, nel caso di dichiarazioni doganali elettroniche, viene condotta direttamente dal sistema. Una volta completato il controllo, il sistema di analisi dei dati vi aggiunge automaticamente data e ora di accettazione della dichiarazione, che risulta così accettata. Fino a questo punto, la procedura funziona in assenza di alcun intervento umano da parte delle autorità.

Tuttavia l'ufficio della dogana può, in seguito, eseguire una ispezione casuale o completa, e sottoporre a verifica i beni dichiarati. A tal fine, il sistema computerizzato ne opera una selezione sulla base di una analisi del rischio. La procedura termina quando si giunge a una decisione finale sulla valutazione, e non è noto se tale decisione possa essere presa già prima di qualunque intervento umano. Tuttavia, il programma DaziT vi farà chiarezza.

Il programma DaziT è una misura federale presa per digitalizzare tutti i processi doganali entro il 2026, così da semplificare e accelerare l'attraversamento dei confini. Il rapporto tra le autorità doganali e i cittadini in relazione al movimento di beni e persone ne risulterà mutato profondamente. I cittadini ligi al dovere dovranno essere infatti in grado di completare le loro formalità per via digitale, indipendentemente da ora e luogo. Anche se la piena implementazione del programma DaziT è ancora in fase di pianificazione, la



revisione della legge che regola le dogane, collegata a DaziT, è inclusa in quella della legge federale sulla protezione dei dati.

Ciò viene spiegato più nel dettaglio qui di seguito, al fine di chiarire l'incertezza, sopra menzionata, circa la procedura di valutazione doganale automatizzata. In futuro, l'Amministrazione federale delle dogane sarà anche esplicitamente autorizzata a condurre valutazioni doganali interamente automatizzate — a dire: non ci sarà intervento umano alcuno nell'intera procedura dei controlli doganali. Così, anche la determinazione dei doveri conseguenti sarà decisa in modo completamente automatico. La possibilità di interagire con un essere umano è prevista solo nel caso di controlli su beni e persone sospette. (EZV 2000)

/ Assicurazione militare e per incidenti

Nel corso della revisione della legge sulla protezione dei dati personali (spiegata più nel dettaglio di seguito), si è deciso che le compagnie che forniscono assicurazioni militari e per incidenti avranno il diritto a processare dati personali in modo automatico. Non è ancora chiaro quali attività automatizzate tali compagnie decideranno di usare. Tuttavia potrebbero, per esempio, fare ricorso ad algoritmi per valutare i referti medici di un assicurato. Attraverso questo sistema completamente automatizzato si potrebbero calcolare premi assicurativi, e prendere decisioni circa le richieste di indennità, coordinandole con gli altri sussidi sociali. Nel piano è previsto che questi istituti siano autorizzati a emettere decisioni automatiche.

/ Riconoscimento automatico dei veicoli

Negli ultimi anni, esponenti politici e opinione pubblica hanno cominciato a mostrare preoccupazioni per l'uso di sistemi automatizzati. Un esempio è il sistema di videocamere capace di immortalare le targhe delle vetture, leggerle usando strumenti di riconoscimento ottico dei caratteri, e confrontarle con un database di riferimento. Questa tecnologia può essere utilizzata per diversi scopi, anche se al momento in Svizzera sta venendo adoperata solo in parte. A livello federale, il sistema di riconoscimento automatico dei veicoli e di monitoraggio del traffico viene usato unicamente come strumento tattico dopo attente valutazioni delle circostanze e del rischio, nonché a seguito di considerazioni di natura economica, e solo ai confini (parlament.ch 2020). Il semi-cantone di Basilea Campagna ha tuttavia dotato la registrazione automatica delle targhe delle vetture e la sua integrazione con i database corrispondenti di un fondamento di legge (EJPD 2019).

/ Assegnazione degli alunni alle scuole elementari

Un altro algoritmo già sviluppato, ma non ancora in uso, è quello progettato per l'assegnazione degli alunni alle scuole elementari. Studi internazionali hanno mostrato che il tasso di segregazione sociale ed etnica nelle scuole cittadine è in aumento. E ciò è problematico, dato che la composizione sociale ed etnica delle scuole ha un effetto dimostrabile sui risultati degli alunni, a prescindere dal loro background di provenienza. In nessun altro paese dell'OECD (Organisation for Economic Co-operation and Development) i cosiddetti "effetti composizionali" sono pronunciati quanto in Svizzera. La differente composizione delle scuole è principalmente dovuta alla segregazione tra quartieri residenziali e relativi bacini d'utenza scolastici. Di conseguenza, il Centre for Democracy di Aarau ha proposto di mischiare gli studenti non solo a seconda della loro origine sociale e linguistica, ma anche nella definizione dei bacini d'utenza scolastici stessi, così da ottenere il più alto grado di mescolanza tra istituti. Per ottimizzare questo processo, un nuovo, dettagliato algoritmo è stato addestrato a ricostruire i bacini d'utenza scolastica e sondare la composizione sociale nelle singole scuole, usando i dati a disposizione del censo per gli studenti dalla prima alla terza classe nel cantone di Zurigo. Sono stati tenuti in considerazione anche dati relativi al traffico, alle reti di marciapiede e percorsi pedonali, ai sottopassaggi e ai cavalcavia. Questi dati potrebbero di conseguenza venire usati per calcolare a quale scuola debbano essere assegnati gli alunni per massimizzare la mescolanza delle classi. Al contempo, la capienza degli edifici scolastici non verrà a questo modo superata, facendo insieme in modo che il tempo impiegato per raggiungere la scuola resti ragionevole.

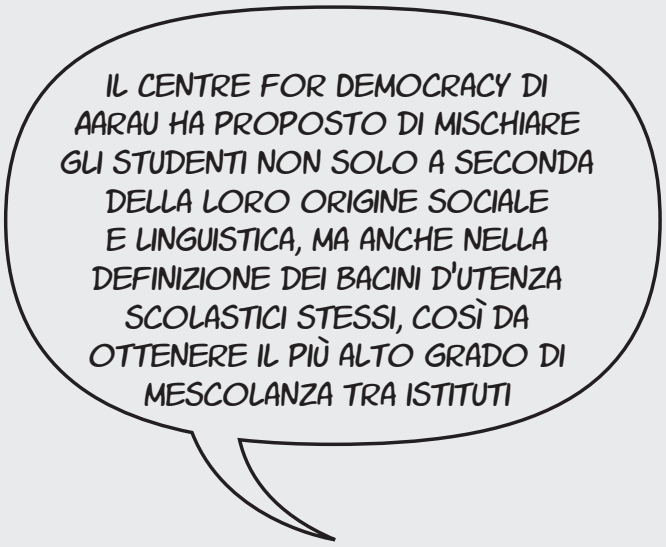
Scelte politiche, forme di controllo, e dibattiti sull'ADM

/ Il fattore chiave è la struttura federale

Per descrivere correttamente il quadro delle scelte politiche in Svizzera, è indispensabile enfatizzarne la prevalente struttura federale — una situazione di cui si è già detto anche riguardo agli esempi di ADM sopra menzionati. La Svizzera è uno stato federale, composto da 26 stati membri (cantoni) dotati di ampia autonomia. I cantoni, a loro volta, garantiscono ampi margini di azione alle autorità municipali. Il risultato è che il dibattito pubblico e politico sui sistemi di ADM dipende in larga misura dal governo preso in esame, il che significa che non se può parlare esaustivamente in questo rapporto. Inoltre questa frammentazione di scelte politiche, regole e iniziative di ricerca introduce il rischio di lavorare in parallelo su questioni che si sovrappongono, il che spiega perché la confederazione miri alle forme di coordinazione avanzata di cui stiamo per dire. Ciononostante, è il governo federale ad avere la piena responsabilità in alcuni, rilevanti campi di legge e di governance, il che vincola legalmente tutti i governi in Svizzera, con un impatto che riguarda, dunque, l'intera popolazione. Ecco, qui di seguito, gli sviluppi che riguardano l'attuale dibattito politico a livello federale.

/ Governo

Attualmente il ruolo dell'ADM nella società, genericamente trattato come "AI", è concepito all'interno del più ampio dibattito sulla digitalizzazione. Il governo federale non dispone di una strategia specifica riguardo all'AI o all'ADM, ma negli ultimi anni ha comunque lanciato una "Strategia Svizzera Digitale", in cui saranno integrati tutti gli aspetti concernenti l'AI. Più in generale, il framework legale a livello na-



zionale in tema di digitalizzazione verrà simultaneamente aggiornato all'interno della revisione della legge federale sulla protezione dei dati.

La Svizzera digitale

Nel 2018, e nel contesto di una crescente digitalizzazione che va ben oltre i soli servizi governativi, la confederazione ha lanciato la propria "Strategia Svizzera Digitale". Uno dei temi su cui il documento concentra le proprie attenzioni sono gli sviluppi dell'AI (BAKOM 2020). Responsabile della Strategia, e in particolare modo della sua coordinazione e implementazione, è il gruppo di coordinamento interdipartimentale "Svizzera digitale", insieme alla sua unità manageriale, chiamata "Direzione operativa Svizzera digitale". (Digital Switzerland 2020)

Tra le attività rientranti nella Strategia, il Consiglio Federale ha predisposto un gruppo di lavoro sul tema dell'AI, e gli ha commissionato un rapporto sulle sfide associate all'intelligenza artificiale. Il rapporto è stato ricevuto dal Consiglio federale nel dicembre 2019 (SBFI 2020). Oltre a discutere le principali problematicità collegate all'AI — cioè tracciabilità ed errori sistematici in dati o algoritmi —, il rapporto dettaglia un concreto bisogno di interventi. Viene inoltre riconosciuto che tutte le sfide, incluso il bisogno di agire, dipendono fortemente dal tema più precisamente in esame. Per questo il rapporto ne ha esaminati 17 più in profondità, tra cui l'AI nella sanità, nell'amministrazione, e nella giustizia (SBFI 2020b, SBFI 2020 c).

In linea di principio, le sfide poste dall'AI in Svizzera sono state, secondo il rapporto, già ampiamente riconosciute e affrontate in diverse aree di intervento politico. Ciononostante, il rapporto interdipartimentale identifica un certo bisogno di agire, che ha portato il Consiglio federale a stabilire quattro interventi. Nella sfera del diritto internazionale, e in tema di uso dell'AI nella formazione dell'opinione pubblica e nel processo decisionale, verranno commissionati ulteriori rapporti, per fornire chiarimenti approfonditi. In seguito, saranno poi esaminati altri modi per migliorare il coordinamento con l'amministrazione federale in tema di uso dell'AI. Sarà in particolare valutata la creazione di un "network di competenze", con un focus specifico sugli aspetti tecnici dell'applicazione dell'AI nell'amministrazione federale. Infine, le scelte politiche rilevanti per l'AI saranno considerate componenti essenziali della strategia "Svizzera Digitale". In questo contesto, il Consiglio federale ha stabilito che il lavoro tra diversi dipartimenti debba continuare, e che entro la primavera del 2020 avrebbero dovuto venire sviluppate delle linee guida per l'intera confederazione (SBFI 2020c).

In aggiunta, riunendosi il 13 maggio 2020, il Consiglio federale ha deciso di creare un "centro nazionale di competenza per la scienza dei dati". L'Ufficio federale di statistica aprirà questo centro di natura interdisciplinare in data 1 gennaio 2021, con lo scopo di supportare l'amministrazione federale nell'implementazione di progetti nel campo della scienza dei dati. A tal fine, verranno promossi tutti gli scambi di conoscenze all'interno dell'amministrazione federale, così come con i circoli scientifici, gli istituti di ricerca e gli organi responsabili della loro applicazione pratica. In particolare, il centro di eccellenza contribuirà alla produzione di informazione trasparente, tenendo al contempo in considerazione la protezione dei dati. La ratio che informa la creazione del nuovo centro trova conferma in una dichiarazione del Consiglio federale, che ha affermato che la scienza dei dati sta diventando sempre più rilevante, non ultimo nella pubblica amministrazione. Secondo il Consiglio federale la scienza dei dati, o "data science", include calcoli "intelligenti" (algoritmi), così che certi compiti complessi possano essere automatizzati. (Bundersrat 2020)

/ Forme di controllo e supervisione

Dato che, a causa dei rapidi cambiamenti tecnologici, la legge federale sulla protezione dei dati non risulta più al passo dei tempi, il Consiglio federale ha manifestato la volontà di adattarla al mutato contesto tecnologico e sociale, e in particolare di voler incrementare la trasparenza nel trattamento dei dati e rafforzare l'autodeterminazione dell'interessato (*data subject*) rispetto ai propri dati. Al contempo,

questa completa revisione dovrebbe consentire alla Svizzera di ratificare la rivista Convenzione ETS 108 del Consiglio d'Europa in tema di protezione dei dati, così come di adottare la direttiva europea 680/2016 sulla protezione dei dati nel contesto di azioni penali — come obbligatoriamente richiesto dall'accordo di Schengen. In aggiunta, la revisione dovrebbe avvicinare l'intera legislazione sulla protezione dei dati in Svizzera al 'Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (GDPR)'. La revisione è attualmente in corso di discussione al Parlamento. (EJPD 2020)

Ma se la completa revisione della legge federale attuale in tema di protezione dei dati porterà a intervenire, per l'appunto, sull'intero atto legislativo in tutti i suoi vari aspetti, una nuova disposizione è di particolare interesse in tema di ADM. Nel caso delle cosiddette "decisioni individuali automatizzate" dovrebbe esserci infatti l'obbligo di informare l'interessato, se la decisione presenta conseguenze legali o altri effetti rilevanti. L'interessato può anche richiedere che la decisione sia rivista da un essere umano, o di essere informato circa la logica a fondamento della decisione presa in modo automatico. In tal modo, una regolamentazione specifica è prevista per le decisioni degli organi federali. Pur se gli organi federali devono anche identificare ogni singola decisione automatizzata come tale, le possibilità che l'interessato possa chiedere revisione umana potrebbero essere ridotte da altre leggi federali. Diversamente da quanto prevede il GDPR nell'Unione Europea, non c'è né un divieto di decisioni automatizzate, né il diritto a chiedere di sottrarsi a una simile decisione. (SFBI 2019)

/ Società civile, mondo accademico e altre organizzazioni

Al quadro finora delineato si aggiunge un certo numero di sedi in cui la Svizzera fa ricerca, discute e lavora sulla trasformazione digitale e le sue opportunità, sfide, bisogni, ed etica. La maggior parte affronta queste questioni a livello generale, anche se alcune affrontano ADM o AI più nello specifico.

/ Istituti di ricerca

La Svizzera dispone di un certo numero di consolidati e rinomati centri di ricerca in tema di intelligenza artificiale. Spiccano, tra gli altri, lo Swiss AI Lab IDSIA (Istituto Dalle Molle di Studi sull'Intelligenza Artificiale) di Lugano (SUPSI 2020) e l'IDIAP Research Institute a Martigny (Idiap 2020), così come quelli dell'Istituto federale di tecnologia a Losanna (EPFL) (EPFL 2020) e Zurigo (ETH) (ETH 2020). Vi si aggiungono a complemento iniziative di soggetti privati, come lo Swiss Group of Artificial Intelligence and Cognitive Science (SGAICO), che mettono allo stesso tavolo ricercatori e utenti, promuovendo condivisione di conoscenze, creazione di fiducia reciproca, e interdisciplinarità. (SGAICO 2020)

Finanziamenti governativi alla ricerca

La confederazione affronta anche l'argomento dei finanziamenti alla ricerca in AI. Per esempio, il governo federale investe in due programmi di ricerca nazionali attraverso il Fondo nazionale svizzero per la ricerca scientifica (SNSF) (SNF 2020). Uno è il programma nazionale di ricerca 77 sulla "trasformazione digitale" (NRP 77) (NRP 77 2020); l'altro il programma nazionale di ricerca 75 in tema di "Big Data" (NRP 75) (NRP 75 2020). Il primo esamina le interdipendenze e gli effetti concreti della trasformazione digitale in Svizzera, e si concentra su formazione e apprendimento, etica, fiducia, governance, economia e mercato del lavoro (NFP 77 2020). Il secondo mira invece a fornire le basi scientifiche adatte a un efficace e appropriato utilizzo di vaste quantità di dati. Di conseguenza, i progetti di ricerca analizzano le questioni sollevate dall'impatto sociale delle tecnologie dell'informazione, e ne affrontano applicazioni concrete (SNF 2020).

Un altro istituto operante in questo settore è la Fondazione per la valutazione delle scelte tecnologiche (TA-Swiss). TA-Swiss è un centro di eccellenza dell'Accademia svizzera per le arti e le scienze, il cui mandato è descritto nella legge federale sulla ricerca. L'Accademia è un organo consultivo, finanziato dal settore pubblico, e ha commissionato diversi studi sull'AI. Tra di essi, il più rilevante in questa sede è uno pubblicato il 15 aprile 2020 sull'uso dell'AI in diversi settori (consumi, lavoro, educazione, ricerca, media, amministrazione pubblica e giustizia). Secondo lo studio, una legge specifica dedicata all'AI non sarebbe efficace. Ciononostante, cittadini, consumatori, e lavoratori — nei loro rapporti con lo Stato, con soggetti privati e con i datori di lavoro — dovrebbero venire informati nel modo più trasparente possibile circa l'uso dell'AI. Quando istituzioni pubbliche e aziende fanno ricorso all'intelligenza artificiale, dovrebbero farlo secondo regole chiare, in modo comprensibile e trasparente. (Christen, M. et al. 2020)

/ Digital Society Initiative

La Digital Society Initiative è stata lanciata nel 2016. Si tratta di un centro di eccellenza presso l'Università di Zurigo, che si prefigge lo scopo di condurre una riflessione critica su tutti gli aspetti della società digitale, contribuendo a informare la digitalizzazione della società, della democrazia, della scienza, della comunicazione, e dell'economia. In più, il centro intende pensare e influenzare i cambiamenti apportati oggi al pensiero dalla digitalizzazione, in un modo che sia orientato al futuro e che faccia figurare l'Università di Zurigo tra i centri di eccellenza per riflettere in modo critico su ogni aspetto della società digitale, sia a livello nazionale che internazionale. (UZH 2020)

/ Digitale Gesellschaft

Digitale Gesellschaft (Società Digitale) è una società no profit e associazione rappresentativa della protezione del cittadino e del consumatore nell'era digitale. Sin dal 2011 ha lavorato al fianco delle organizzazioni della società civile per costruire una sfera pubblica sostenibile, democratica e libera, e il suo obiettivo è difendere i diritti fondamentali in un mondo digitalmente connesso. (Digitale Gesellschaft (o.J.))

/ Altre organizzazioni

Dovrebbero essere menzionate svariate altre organizzazioni, che si concentrano sulla digitalizzazione in genere, specialmente in contesti economici. Per esempio, l'Associazione svizzera delle telecomunicazioni (asut 2020), digitalswitzerland (Castle 2020), la Swiss Data Alliance e Swiss Fintech Innovations.

Conclusioni

L'ADM è utilizzato in diversi campi del settore pubblico in Svizzera, ma non in modo onnicomprensivo o centralizzato. Solo alcuni cantoni fanno uso di ADM nelle attività di polizia, per esempio, e i sistemi utilizzati variano. Il vantaggio di un simile approccio è che i cantoni o il governo federale hanno l'opportunità di trarre beneficio dall'esperienza di altri cantoni. Il contraltare è la perdita di efficienza che ne può conseguire.

Ci sono alcune, selettive, basi legali, ma nessuna legge uniforme sull'ADM o l'e-government o alcunché di simile. Né è possibile reperire una specifica strategia per l'AI o l'ADM, anche se di recente è stata posta attenzione a una maggiore coordinazione, sia tra diversi dipartimenti a livello federale che tra governo federale e cantoni. Non sono utilizzati metodi di machine learning per attività di Stato in senso stretto, per esempio nel lavoro della polizia o nel sistema penale, per quanto se ne può sapere.

Inoltre, e similmente, l'ADM è usato o discusso in modo selettivo e parziale — anche se nel più ampio settore pubblico ciò accade più spesso. Un buon esempio ne è l'adozione nel sistema sanitario svizzero, con l'ospedale universitario di Ginevra capace di primeggiare in Europa nell'uso di sistemi di ADM per suggerire cure a pazienti malati di cancro.

Riferimenti bibliografici

Asut (o. J.): in: asut.ch, [online] https://asut.ch/asut/de/page/index.xhtml [30.01.2020]

Bundesamt für Kommunikation BAKOM (o. J.): Digitale Schweiz, in: admin.ch, [online] https://www.bakom.admin.ch/bakom/de/home/digital-und-internet/strategie-digitale-schweiz.html [30.01.2020].

Der Bundesrat (o.J.): Der Bundesrat schafft ein Kompetenzzentrum für Datenwissenschaft, In: admin.ch, [online] https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-79101.html [15.05.2020].)

Christen, M. et al. (2020): Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz, in: TA-Swiss, [online] https://www.ta-swiss.ch/themen-projekte-publikationen/informationsgesellschaft/kuenstliche-intelligenz/ [15.05.2020].

Ciritsis, Alexander / Cristina Rossi / Matthias Eberhard / Magda Marcon / Anton S. Becker / Andreas Boss (2019): Automatic classification of ultrasound breast lesions using a deep convolutional neural network mimicking human decision-making, in: European Radiology, Jg. 29, Nr. 10, S. 5458–5468, doi: 10.1007/s00330-019-06118-7.

Digitale Gesellschaft (o. J.): Über uns, in: Digitale Gesellschaft, [online] https://www.digitale-gesellschaft.ch/uber-uns/ [30.01.2020].)

digitalswitzerland (Castle, Danièle Digitalswitzerland (2019): Digitalswitzerland – Making Switzerland a Leading Digital Innovation Hub, in: digitalswitzerland, [online] https://digitalswitzerland.com [30.01.2020])

Digital Switzerland (2020): (Ofcom, Federal Office Of Communications (o. J.): Digital Switzerland Business Office, in: admin.ch, [online] https://www.bakom.admin.ch/bakom/en/homepage/ofcom/organisation/organisation-chart/information-society-business-office.html [30.01.2020].)

EPFL (o. J.): in: epfl, [online] https://www.epfl.ch/en/ [30.01.2020]

EJPD (o. J.): Stärkung des Datenschutzes, in: admin.ch, [online] https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/datenschutzstaerkung.html [30.01.2020c].

E-ID Referendum (o.J.): in: e-id-referendum.ch/, [online] https://www.e-id-referendum.ch [31.1.2020].

EJPD (o. J.): Stärkung des Datenschutzes, in: admin.ch, [online] https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/datenschutzstaerkung.html [30.01.2020c].

EJPD Eidgenössisches Justiz- und Polizeidepartement (2019): Änderung der Geschwindigkeitsmessmittel-Verordnung (SR 941.261) Automatische Erkennung von Kontrollschildern, in: admin.ch, [online] https://www.admin.ch/ch/d/gg/pc/documents/3059/Erl_Bericht_de.

EZV (2020): EZV, Eidgenössische Zollverwaltung (o. J.): Transformationsprogramm DaziT, in: admin.ch, [online] https://www.ezv.admin.ch/ezv/de/home/themen/projekte/dazit.html [30.01.2020].

ETH Zurich - Homepage (o. J.): in: ETH Zurich - Homepage | ETH Zurich, [online] https://ethz.ch/en.html [30.01.2020].

Federal office of public health FOPH (2020): (Health insurance: The Essentials in Brief (o. J.): in: admin.ch, [online] https://www.bag.admin.ch/bag/en/home/versicherungen/krankenversicherung/krankenversicherung-das-wichtigste-in-kuerze.html [13.02.2020].)

Geschäft Ansehen (o. J.): in: parlament.ch, [online] https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20143747 [30.01.2020].

Heinhold, Florian (2019): Hoffnung für Patienten?: Künstliche Intelligenz in der Medizin, in: br.ch, [online] https://www.br.de/br-fernsehen/sendungen/gesundheit/kuenstliche-intelligenz-ki-medizin-102.html [30.01.2020].

Idiap Research Institute (o. J.): in: Idiap Research Institute, Artificial Intelligence for Society, [online] https://www.idiap.ch/en [30.01.2020]

Der IPV-Chatbot – SVA St.Gallen (o. J.): in: svasg.ch, [online] https://www.svasg.ch/news/meldungen/ipv-chatbot.php [30.01.2020].

Leese, Matthias (2018): Predictive Policing in der Schweiz: Chancen, Herausforderungen Risiken, in: Bulletin zur Schweizerischen Sicherheitspolitik, Jg. 2018, S. 57–72.

Lindner, Martin (2019): KI in der Medizin: Hilfe bei einfachen und repetitiven Aufgaben, in: Neue Zürcher Zeitung, [online] https://www.nzz.ch/wissenschaft/ki-in-der-medizin-hilfe-bei-einfachen-und-repetitiven-aufgaben-ld.1497525?reduced=true [30.01.2020

Medinside (o. J.): in: Medinside, [online] https://www.medinside.ch/de/post/in-genf-schlaegt-der-computer-die-krebsbehandlung-vor [14.02.2020].

NRP 75 Big Data (o. J.): in: SNF, [online] http://www.snf.ch/en/researchinFocus/nrp/nfp-75/Pages/default.aspx [30.01.2020].

NFP [Nr.] (o. J.): in: nfp77.ch, [online] http://www.nfp77.ch/en/Pages/Home.aspx [30.01.2020]

NRP 75 Big Data (o. J.): in: SNF, [online] http://www.snf.ch/en/researchinFocus/nrp/nfp-75/Pages/default.aspx [30.01.2020].

NFP [Nr.] (o. J.): in: nfp77.ch, [online] http://www.nfp77.ch/en/Pages/Home.aspx [30.01.2020]

Ringeisen, Peter / Andrea Bertolosi-Lehr / Labinot Demaj (2018): Automatisierung und Digitalisierung in der öffentlichen Verwaltung: digitale Verwaltungsassistenten als neue Schnittstelle zwischen Bevölkerung und Gemeinwesen, in: Yearbook of Swiss Administrative Sciences, Jg. 9, Nr. 1, S. 51–65, doi: 10.5334/ssas.123.

ROSNET > ROS allgemein (o. J.): in: ROSNET, [online] https://www.rosnet.ch/de-ch/ros-allgemein [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (o. J.): Künstliche Intelligenz, in: admin.ch, [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (2019): Herausforderungen der künstlichen Intelligenz - Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, in: admin.ch, [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (o. J.): Künstliche Intelligenz, in: admin.ch, [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (2019): Herausforderungen der künstlichen Intelligenz - Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, in: admin.ch, [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

Schaffhauser eID+ - Kanton Schaffhausen (o. J.): in: sh.ch, [online] https://sh.ch/CMS/Webseite/Kanton-Schaffhausen/Beh-rde/Services/Schaffhauser-eID--2077281-DE.html [30.01.2020].

SGAICO - Swiss Group for Artificial Intelligence and Cognitive Science (2017): in: SI Hauptseite, [online] https://swissinformatics.org/de/gruppierungen/fg/sgaico/ [30.01.2020]

SNF, [online] http://www.snf.ch/en/Pages/default.aspx [30.01.2020]

Srf/Blur;Hesa (2017): Wie «Precobs» funktioniert - Die wichtigsten Fragen zur «Software gegen Einbrecher», in: Schweizer Radio und Fernsehen (SRF), [online] https://www.srf.ch/news/schweiz/wie-precobs-funktioniert-die-wichtigsten-fragen-zur-software-gegen

SUPSI - Dalle Molle Institute for Artificial Intelligence - Homepage (o. J.): in: idsia, [online] http://www.idsia.ch [30.01.2020].

Swissdataaliance (o. J.): in: swissdataalliance, [online] https://www.swissdataalliance.ch [30.01.2020].

Swiss Fintech Innovations (SFTI introduces Swiss API information platform (2019): in: Swiss Fintech Innovations – Future of Financial Services, [online] https://swissfintechinnovations.ch [30.01.2020]).

Schwerzmann, Jacqueline Amanda Arroyo (2019): Dr. Supercomputer - Mit künstlicher Intelligenz gegen den Krebs, in: Schweizer Radio und Fernsehen (SRF), [online] https://www.srf.ch/news/schweiz/dr-supercomputer-mit-kuenstlicher-intelligenz-gegen-den-krebs

Treuthardt, Daniel / Melanie Kröger (2019): Der Risikoorientierte Sanktionenvollzug (ROS) – empirische Überprüfung des Fall-Screening-Tools (FaST), in: Schweizerische Zeitschrift für Kriminologie, Jg. 2019, Nr. 1–2, S. 76–85.); (Treuthardt, Daniel / Melanie Kröger / Mirjam Loewe-Baur (2018): Der Risikoorientierte Sanktionenvollzug (ROS) – aktuelle Entwicklungen, in: Schweizerische Zeitschrift für Kriminologie, Jg. 2018, Nr. 2, S. 24–32.

ZDA (2019): Durchmischung in städtischen Schulen, in: zdaarau.ch, [online] https://www.zdaarau.ch/dokumente/SB-17-Durchmischung-Schulen-ZDA.pdf [30.01.2020].

Team

/ Beate Autering

Graphic designer e layout artist

Beate Autering è una graphic designer freelance. Si è laureata in design, e gestisce lo studio beworx. Crea progetti, grafiche e illustrazioni, e fornisce inoltre servizi di editing di immagini e post-produzione.

/ Nadja Braun Binder

Autrice del capitolo di ricerca sulla Svizzera

Nadja Braun Binder ha studiato Giurisprudenza all'Università di Berna, dove ha ottenuto il dottorato di ricerca. La sua carriera accademica l'ha portata, nel 2011, al Research Institute for Public Administration di Speyer, dove si è dedicata alla ricerca in tema di automazione delle procedure amministrative, e non solo. Nel 2017 ha ricevuto l'abilitazione dalla German University of Administrative Sciences, Speyer, prima di rispondere alla chiamata della Facoltà di Giurisprudenza dell'Università di Zurigo, dove ha lavorato come assistente universitaria fino al 2019. Dal 2019, Nadja è Docente di Diritto Pubblico all'Università di Basilea. Le sue ricerche si concentrano sulle questioni legali collegate alla digitalizzazione nel governo e nell'amministrazione pubblica. Sta attualmente conducendo uno studio sull'uso dell'intelligenza artificiale nell'amministrazione pubblica nel cantone di Zurigo.

/ Fabio Chiusi

Editor del rapporto, autore della storia giornalistica e del capitolo di ricerca sull'Italia, così come dell'introduzione e del capitolo sull'Europa

Photo: Julia Bornkessel

Fabio Chiusi lavora ad AlgorithmWatch come co-editor e project manager dell'edizione 2020 del rapporto Automating Society. Dopo un decennio nel giornalismo tecnologico, è stato consulente e ricercatore in tema di dati e politica (Tactical Tech) e AI nel giornalismo (Polis LSE). Ha coordinato il rapporto "Persuasori Social" sulla regolamentazione delle campagne politiche sui social media per il Progetto PuntoZero, e ha lavorato come responsabile politiche tecnologiche del Questore della Camera dei Deputati del Parlamento italiano, durante l'attuale legislatura. Fabio è fellow del Nexa Center for Internet & Society di Torino, e docente a contratto all'Università di San Marino, dove insegna Giornalismo e nuovi media ed Editoria e media digitali. È autore di diversi saggi su tecnologia e società, di cui il più recente è 'Io non sono qui. Visioni e inquietudini da un futuro presente' (DeA Planeta, 2018), tradotto in polacco e cinese. Scrive di politiche tecnologiche sul blog collettivo Valigia Blu.

/ Samuel Daveti

Fumettista

Samuel Daveti nasce nel 1983 a Grosseto. Socio fondatore dell'Associazione Culturale Double Shot, per il mercato francese sceneggia l'albo Akron Le guerrier (Soleil, 2009) ed è curatore del volume antologico Fascia Protetta (Double Shot, 2009). Nel 2011 è uno dei fondatori di Mam-maiuto, collettivo di autoproduzione. Sceneggiatore di Un Lungo Cammino (Mammaiuto, 2014; Shockdom, 2017), che diventerà un film per la casa di produzione cinematografica Brandon Box, nel 2018 scrive I Tre Cani per i disegni di Laura Camelli, storia vincitrice del Premio Micheluzzi al Napoli Comicon 2018 e del premio Boscarato al Treviso Comic Book Festival come Miglior Web-comic. Nel 2020 sceneggia Inerti per i disegni di Francesco Rossi, storia vincitrice del premio Bo-scarato al Treviso Comic Book Festival come miglior Webcomic.

/ Catherine Egli

Autrice del capitolo di ricerca sulla Svizzera

Catherine Egli ha di recente portato a compimento un doppio master bilingue in Giurisprudenza dalle Università di Basilea e Ginevra, con una tesi sull'automated decision-making individuale e, in particolare, sulla necessità di una regolamentazione per la legge svizzera sulla procedura amministrativa. Durante gli studi, Catherine ha lavorato anche per la cattedra della Prof.ssa Nadja Braun Binder, conducendo ricerche su questioni di legge collegate all'automated decision-making. I suoi argomenti elettivi includono la separazione dei poteri, la digitalizzazione dell'amministrazione pubblica e la democrazia digitale.

/ Sarah Fischer

Editor del rapporto

Sarah Fischer è project manager per il progetto "Ethics of Algorithms" di Bertelsmann Stiftung, per il quale è principalmente responsabile degli studi scientifici. In precedenza, ha lavorato come postdoc fellow nel programma specialistico "Trust and Communication in a Digitalized World" all'Università di Münster, dove si è concentrata sul tema della fiducia nei motori di ricerca. Nello stesso gruppo di addestramento alla ricerca, Sarah ha ottenuto il dottorato con una tesi sulla fi-ducia nei servizi sanitari online. Ha studiato scienze della comunicazione alla Friedrich Schiller University di Jena, ed è co-autrice dei paper "Where Machines can err. Sources of error and re-sponsibilities in processes of algorithmic decision making" e "What Germany know and believes about algorithms".

/ Leonard Haas

Ulteriore revisione dei testi

Leonard Haas lavora come assistente ricercatore ad AlgorithmWatch. Tra le altre cose, è stato responsabile della concezione, dell'implementazione e del mantenimento dell'AI Ethics Guidelines Global Inventory. È studente in un master nel campo delle scienze sociale alla Humboldt University di Berlino e ha ottenuto dall'Università di Lipsia due lauree di primo livello, in Digital Humanities e Scienze Politiche. La sua attività di ricerca si concentra sull'automazione del lavoro e della gover-nance. In aggiunta, Leonard si occupa di politiche dei dati di interesse pubblico e di lotte dei lavoratori del settore tecnologico.

/ Graham Holliday

Copy-editor dell'edizione completa in inglese

Graham Holliday è un editor freelance, autore, e trainer giornalistico. Ha lavorato per la BBC in diversi ruoli per quasi due decenni, ed è stato corrispondente di Reuters dalla Ruanda. Graham è anche editor per il programma della CNN Parts Unknown e per Roads & Kingdoms, la rivista internazionale di corrispondenze dall'estero. Lo scomparso Anthony Bourdain ha pubblicato i primi due libri di Graham, che sono stati recensiti da, tra gli altri, New York Times, Los Angeles Times, Wall Street Journal, Publisher's Weekly, Library Journal e NPR.

/ Nikolas Kayser-Bril

Editor, autore della storia giornalistica

Nicolas Kayser-Bril è un data journalist, e lavora ad AlgorithmWatch come reporter. Ha aperto la strada a nuove forme di giornalismo in Francia e in Europa, ed è uno dei massimi esperti di data journalism. Nicolas interviene regolarmente a conferenze di rilievo internazionale, insegna giornalismo in scuole di giornalismo francesi, e tiene sessioni formative nelle redazioni. Giornalista e sviluppatore autodidatta (oltre che laureato in Economia), ha cominciato nel 2009 sviluppando piccole applicazioni interattive basate sui dati per Le Monde, a Parigi. In seguito, nel 2010, ha messo insieme la squadra di data journalists di OWNI, prima di co-fondare e gestire Journalism++ dal 2011 al 2017. Nicolas è anche uno dei principali contributor del Data Journalism Handbook, il testo di riferimento per la popolarizzazione dell'uso dei dati nel giornalismo in tutto il mondo.

/ Anna Mätzener

Editor

Anna Mätzener è direttrice esecutiva di AlgorithmWatch Svizzera. Ha conseguito un dottorato in matematica all'Università di Zurigo, dove ha anche studiato filosofia e linguistica italiana. Prima della sua attività presso AlgorithmWatch Svizzera, è stata editor di matematica e storia delle scienze in una casa editrice scientifica internazionale e, più recentemente, insegnante di matematica in un liceo di Zurigo.

/ Lorenzo Palloni

Fumettista

Lorenzo Palloni, fumettista, classe 1987. Fra i fondatori di Mammaiuto, è autore de La lupa (Sal-dapress, 2019); Mooned (Shockdom, 2017); Scary Allan Crow (Edizioni Inkiostro, 2017) è sceneg-giatore di Desolation Club con Vittoria Macioci (Saldapress, 2019); Emma Wrong con Laura Guglielmo (Editions Akileos, 2019); Instantly Elsewhere (Shockdom, 2018) e Terranera (Feltrinelli, 2020), entrambi con Martoz. È stato ospite della "Maison Des Auteurs" di Angoulême e ha vinto due premi Boscarato come "Miglior Sceneggiatore Italiano" e, nel 2019, il premio Gran Guinigi come "Miglior Sceneggiatore". Al momento sta lavorando a libri di prossima uscita per il mercato francese e italiano, ed è docente di Sceneggiatura e Storytelling alla "Scuola Internazionale di Comics" nelle sedi di Firenze e Reggio Emilia.

/ Kristina Penner

Autrice del capitolo sull'Europa

Photo: Julia Bornkessel

Kristina Penner è executive advisor ad AlgoritmWatch. I suoi interessi di ricerca includono l'ADM nei sistemi di welfare, le forme di scoring sociale, e gli impatti dell'ADM sulla società, così come la sostenibilità delle nuove tecnologie da un punto di vista olistico. La sua analisi dei sistemi di ge-stione dei confini dell'UE è una prosecuzione della sua precedente esperienza di ricerca e consu-lenza sulle leggi in materia di asilo. Altre precedenti esperienze di Kristina includono progetti sull'uso dei media nella società civile e nel giornalismo "conflict-sensitive", così come il coinvol-gimento come stakeholder nel processo di pace nelle Filippine. Kristina ha ottenuto un master in studi internazionali/ricerca su pace e conflitto alla Goethe University di Francoforte.

/ Alessio Ravazzani

Fumettista

Alessio Ravazzani nasce a Cuggiono (MI) l'11 maggio del 1975. Si diploma al Liceo artistico Can-diani di Busto Arsizio nel 1993 e nel 2007 si diploma alla Scuola internazionale di Comics di Firen-ze. Grafico editoriale, fumettista e illustratore, si approccia al lettering come professionista nel 2010 con delle collaborazioni con le Edizioni BD e la sua divisione J-POP. Nel 2011 comincia la collaborazione con lo studio della divisione Lion della casa editrice RW, proprietaria dei diritti della DC Comic per l'Italia; quattro anni più tardi, quella con Saldapress, con la francese Kazé e l'americana Viz media. Nel 2016 si aggiungono la Symmaceo e lo studio RAM: ambedue partner della casa editrice Panini comics. Nel 2018 comincia a collaborare con Coconino-Fandango. Dal 2011 è membro del collettivo di autori Mammaiuto con cui collabora come disegnatore e grafico. Attualmente abita e lavora alla Spezia.

/ Friederike Reinhold

Editing aggiuntivo su introduzione e raccomandazioni di policy

In qualità di senior policy advisor, Friederike Reinhold è responsabile dello sviluppo degli strumenti di policy e advocacy di AlgorithmWatch. Prima di entrare in AlgorithmWatch, ha lavorato come consulente per le politiche umanitarie all'Ufficio Esteri del governo federale tedesco, con il Consiglio norvegese per i rifugiati (NRC) in Iran, con il Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) in Afghanistan, e al Centro di ricerca di Scienze Sociali WZB di Berlino.

/ Matthias Spielkamp

Editor del rapporto

Photo: Julia Bornkessel

Matthias Spielkamp è co-fondatore e direttore esecutivo di AlgorithmWatch. Ha testimoniato di fronte a diverse commissioni del Parlamento federale tedesco in tema di AI e automazione. Mat-thias siede nel board direttivo della sezione per la Germania di Reporters Without Borders e nei comitati consultivi di Stiftung Warentest e del Whistleblower Network. È stato fellow di ZEIT Stif-tung, Stiftung Mercator e dell'American Council on Germany. Matthias ha anche fondato il maga-zine online mobilsicher.de, che si occupa di sicurezza dei telefonini, e ha un pubblico di oltre 170.000 lettori mensili. Ha scritto e curato libri sul giornalismo digitale e la governance di Internet, ed è stato tra i 15 architetti del nostro futuro "data-driven" scelti da Silicon Republic. Matthias ha ottenuto un master in giornalismo all'Università del Colorado, a Boulder, e un master in filosofia all'Università Libera di Berlino.

/ Tiger Stangl

Graphic designer aggiunta e layout artist

Tiger Stangl è una designer che lavora a Berlino su progetti di design editoriale di beworx, Friedrich Ebert Stiftung, Buske Verlag, UNESCO Welterebe Deutschland e.V., Sehstern Agency, Rights Lab, e Landesspracheninstitut Bochum.

/ Marc Thümmler

Coordinatore della pubblicazione

Photo: Julia Bornkessel

Marc Thümmler è responsabile delle relazioni pubbliche e della comunicazione ad AlgorithmWatch. Ha un master in media studies, e ha lavorato come produttore ed editor in una compagnia cinematografica, oltre a gestire progetti per la Deutsche Kinemathek e per l'organizzazione della società civile Gesicht Zeigen. In aggiunta ai suoi principali incarichi, Marc è stato anche coinvolto nelle campagne di crowdfunding e crowdsourcing per OpenSCHUFA, e ha inoltre coordinato la prima edizione del rapporto Automating Society, pubblicata nel 2019.

ORGANIZZAZIONI

/ AlgorithmWatch Svizzera

AlgorithmWatch è una organizzazione no profit di ricerca e advocacy che si dedica a studiare, comprendere e analizzare i sistemi di decision-making algoritmico/automatizzato (ADM) e il loro impatto sulla società. Se un ricorso accorto ai sistemi di ADM può arrecare benefici tanto a singoli individui quanto a intere comunità, il loro uso si accompagna anche a grossi rischi. Per difendere l'autonomia umana e i diritti fondamentali, e massimizzare il bene pubblico, riteniamo cruciale sottoporre i sistemi di ADM a scrutinio democratico. L'uso di sistemi di ADM che abbiano un significativo impatto sui diritti individuali e collettivi non dovrebbe unicamente essere reso pubblico in modi chiari e accessibili, ma dovrebbe anche consentire a ogni singolo individuo di comprendere come si sia pervenuti a una decisione, e di contestarla se necessario. Per questo aiutiamo i cittadini a meglio comprendere i sistemi di ADM, così come a svilupparne nuove modalità di governance — attraverso un misto di tecnologie, regole e adeguati organi di controllo. È questo il nostro modo di dare un contributo verso una società più equa e inclusiva, oltre che per massimizare i benefici dei sistemi di ADM per la società tutta.

https://algorithmwatch.ch/it/

ALGORITHM WATCH / CH

/ Bertelsmann Stiftung

L'obiettivo di Bertelsmann Stiftung è promuovere l'inclusione sociale per tutti. L'impegno è a perseguire lo scopo attraverso programmi mirati al miglioramento dell'istruzione, a informare la democrazia, far progredire la società, promuovere la salute, rivitalizzare la cultura e rinforzare le economie. Tramite le sue attività, Bertelsmann Stiftung si propone di incoraggiare i cittadini a contribuire al bene pubblico. Fondata nel 1977 da Reinhard Mohn, la fondazione no profit detiene la maggioranza delle azioni di Bertelsmann SE & Co. KGaA. Bertelsmann Stiftung è una fondazione privata e indipendente. Con il suo progetto "Etica degli Algoritmi", sta fornendo uno sguardo approfondito sulle conseguenze sociali dei processi decisionali algoritmici, allo scopo di assicurare che tali sistemi siano utilizzati per il bene collettivo. La fondazione intende così contribuire a informare e promuovere sistemi algoritmici che facilitino l'inclusione sociale. Ciò comporta impegnarsi in ciò che è meglio per la società, piuttosto che in ciò che è tecnicamente possibile — così che le decisioni prese con le macchine siano al servizio del genere umano nel modo più proficuo.

https://www.bertelsmann-stiftung.de/en

BertelsmannStiftung

/ Engagement Migros

Il fondo di sostegno Engagement Migros rende possibili progetti pionieri di trasformazione sociale che percorrono nuove strade e sperimentano soluzioni orientate al futuro. L'approccio di sostegno orientato al risultato unisce il sostegno finanziario a prestazioni di coaching nel laboratorio pionieristico. Engagement Migros viene reso possibile dalle aziende del Gruppo Migros con circa dieci milioni di franchi all'anno e dal 2012 integra il Percento culturale Migros.

https://www.engagement-migros.ch

ENGAGEMENT
UN FONDO DI SOSTEGNO DEL GRUPPO MIGROS

Life in the automated society: How automated decision-making systems became mainstream, and what to do about it

By Fabio Chiusi

The editorial deadline for this report was 30 September 2020. Later developments could not be included.

On a cloudy August day in London, students were angry. They flocked to Parliament Square by the hundreds, in protest – their placards emblazoned with support for unusual allies: their teachers, and an even more unusual target: an algorithm.

Due to the COVID-19 pandemic, schools closed in March in the United Kingdom. With the virus still raging throughout Europe over the summer of 2020, students knew that their final exams would have to be canceled, and their assessments – somehow – changed. What they could not have imagined, however, was that thousands of them would end up with lower than expected grades as a result.

Students protesting knew what was to blame, as apparent by their signs and chants: the automated decision-making (ADM) system deployed by the Office of Qualifications and Examinations Regulation (Ofqual). It planned to produce the best data-based assessment for both General Certificates of Secondary Education and A-level results, in such a way that "the distribution of grades follows a similar pattern to that in other years, so that this year's students do not face a systemic disadvantage as a consequence of circumstances this year".

The government wanted to avoid the excess of optimism[1] that would have resulted from human judgment alone, according to its own estimates: compared to the historical series, grades would have been too high. But this attempt to be "as far as possible, fair to students who had been unable to sit their exams this summer" failed spectacularly, and, on that grey August day of protest, the students kept on coming, performing chants, and holding signs to express an urgent need for social justice. Some were desperate, some broke down and cried.

"Stop stealing our future", read one placard, echoing the Fridays for Future protests of climate activists. Others, however, were more specifically tailored to the flaws of the ADM grading system: "Grade my work, not my postcode", we're "students, not stats", they read, denouncing the discriminatory outcomes of the system[2].

Finally, a chant erupted from the crowd, one that has come to the future of protest: "Fuck the algorithm". Scared that the government was casually – and opaquely – automating their future, no matter how inconsistent with their skills and efforts, students screamed for the right not to have their life chances unduly affected by bad code. They wanted to have a say, and what they said should be heard.

Algorithms are neither "neutral" nor "objective" even though we tend to think that they are. They replicate the assumptions and beliefs of those who decide to deploy them and program them. Humans, therefore, are, or should be, responsible for both good and bad algorithmic choices, not "algorithms" or ADM systems. The machine may be scary, but the ghost within it is always human. And humans are complicated, even more so than algorithms.

The protesting students were not as naive as to believe that their woes were solely the fault of an algorithm, anyway. In fact, they were not chanting against "the algorithm" in an outburst of technological determinism; they were motivated by an urge to protect and promote social justice. In this respect, their protest more closely resembles that of the Luddites. Just as the labor movement that crushed mechanized looms and knitting frames in the 19th Century, they know that ADM systems are about power, and should not be mistaken for being an allegedly objective technology. So, they chanted "justice for the working class", asked for the resignation of the Health Secretary, portrayed the ADM system as "classism at its finest", "blatant classism".

Eventually, the students succeeded in abolishing the system which put their educational career and chances in life at risk: in a spectacular U-turn, the UK government scrapped the error-prone ADM system and utilized the grades predicted by teachers.

But there's more to this story than the fact that the protesters won in the end. This example highlights how poorly designed, implemented, and overseen systems that reproduce human bias and discrimination fail to make use of the potential that ADM systems have, such as leveraging comparability and fairness.

More clearly than many struggles in the past, this protest reveals that we're no longer just automating society. We have automated it already – and, finally, somebody noticed.

1 "The research literature suggests that, in estimating the grades students are likely to achieve, teachers tend to be optimistic (although not in all cases)", writes Ofqual, cfr. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/909035/6656-2_-_Executive_summary.pdf

2 Cfr. the UK chapter for details.

/ From Automating Society to the automated society

When launching the first edition of this report, we decided to call it "Automating Society", as ADM systems in Europe were mostly new, experimental, and unmapped – and, above all, the exception rather than the norm.

This situation has changed rapidly. As clearly shown by the many cases gathered in this report through our outstanding network of researchers, the deployment of ADM systems has vastly increased in just over a year. ADM systems now affect almost all kinds of human activities, and, most notably, the distribution of services to millions of European citizens – and their access to their rights.

The stubborn opacity surrounding the ever-increasing use of ADM systems has made it all the more urgent that we continue to increase our efforts. Therefore, we have added four countries (Estonia, Greece, Portugal, and Switzerland) to the 12 we already analyzed in the previous edition of this report, bringing the total to 16 countries. While far from exhaustive, this allows us to provide a broader picture of the ADM scenario in Europe. Considering the impact these systems may have on everyday life, and how profoundly they challenge our intuitions – if not our norms and rules – about the relationship between democratic governance and automation, we believe this is an essential endeavor.

This is especially true during the COVID-19 pandemic, a time in which we have witnessed the (mostly rushed) adoption of a plethora of ADM systems that aim to contribute to securing public health through data-based tools and automation. We deemed this development to be so important that we decided to dedicate a "preview report" to it, published[3] in August 2020 within the scope of the 'Automating Society' project.

Even in Europe, when it comes to the deployment of ADM systems, the sky is the limit. Just think of some of the cases introduced in this report, adding to the many – from welfare to education, the health system, to the judiciary – that we already reported on in the previous edition. In the following pages, and for the first time, we provide updates on the development of these cases in three ways. Firstly, through journalistic stories, then, through research-based sections cataloging different examples, and, finally, with graphic novels. We felt that these ADM systems are – and increasingly will be – so crucial in everyone's lives that we needed to try and communicate how they work, and what they actually *do to us*, in both rigorous and new ways, to reach all kinds of audiences. After all, ADM systems have an impact on all of us.

Or at least they should. We've seen, for example, how a new, automated, proactive service distributes family benefits in Estonia. Parents no longer even need to apply for benefits: from birth, the state collects all the information about each newborn and their parents and collates it in databases. As a result, the parents automatically receive benefits if they are entitled to them.

In Finland, the identification of individual risk factors related to social exclusion in young adults is automated through a tool developed by the Japanese giant, Fujitsu. In France, data from social networks can be scraped to feed machine learning algorithms that are employed to detect tax fraud.

Italy is experimenting with "predictive jurisprudence". This uses automation to help judges understand trends from previous court rulings on the subject at hand. And, in Denmark, the government tried to monitor every keyboard and mouse click on students' computers during exams, causing – again – massive student protests that led to the withdrawal of the system, for the time being.

/ Time to put ADM wrongs to right

In principle, ADM systems have the potential to benefit people's lives – by processing huge amounts of data, supporting people in decision-making processes, and providing tailored applications.

In practice, however, we found very few cases that convincingly demonstrated such a positive impact.

For example, the VioGén system, deployed in Spain since 2007 to assess risk in cases of domestic violence, while far from perfect, shows "reasonable performance indexes" and has helped protect many women from violence.

In Portugal, a centralized, automated system deployed to deter fraud associated with medical prescriptions has reportedly reduced fraud by 80% in a single year. A similar system, in Slovenia, used to combat tax fraud has proved useful for inspectors, according to tax authorities[4].

3 'Automated Decision-Making Systems in the COVID-19 Pandemic: A European Perspective', https://algorithmwatch.org/en/project/automating-society-2020-covid19/

4 Cfr. the chapter on Slovenia for details.

When looking at the current state of ADM systems in Europe, positive examples with clear benefits are rare. Throughout the report, we describe how the vast majority of uses tend to put people at risk rather than help them. But, to truly judge the actual positive and negative impact, we need more transparency about goals and more data about the workings of ADM systems that are tested and deployed.

The message for policy-makers couldn't be clearer. If we truly want to make the most of their potential, while at the same time respecting human rights and democracy, the time to step up, make those systems transparent, and put ADM wrongs right, is now,.

/ Face recognition, face recognition, everywhere

Different tools are being adopted in different countries. One technology, however, is now common to most: face recognition. This is arguably the newest, quickest, and most concerning development highlighted in this report. Face recognition, nearly absent from the 2019 edition, is being trialed and deployed at an alarming rate throughout Europe. In just over a year since our last report, face recognition is present in schools, stadiums, airports, and even in casinos. It is also used for predictive policing, to apprehend criminals, against racism, and, regarding the COVID-19 pandemic, to enforce social distancing, both in apps and through "smart" video-surveillance.

Face recognition, nearly absent from the 2019 edition, is being trialed and deployed at an alarming rate throughout Europe.

New ADM deployments continue, even in the face of mounting evidence of their lack of accuracy. And when challenges emerge, proponents of these systems simply try and find their way around them. In Belgium, a face recognition system used by the police is still "partially active", even though a temporary ban has been issued by the Oversight Body for Police Information. And, in Slovenia, the use of face recognition technology by the police was legalized five years after they first started using it.

This trend, if not challenged, risks normalizing the idea of being constantly – and opaquely – watched, thus crystallizing a new status quo of pervasive mass surveillance. This is why many from the civil liberties community would have welcomed a much more aggressive policy response by EU institutions to this[5].

Even the act of smiling is now part of an ADM system piloted in banks in Poland: the more an employee smiles, the better the reward. And it's not just faces that are being monitored. In Italy, a sound surveillance system was proposed as an anti-racism tool to be used in all football stadiums.

/ Black boxes are still black boxes

A startling finding in this report is that, while change happened rapidly regarding the deployment of ADM systems, the same is not true when it comes to the transparency of these systems. In 2015, Brooklyn Law School professor, Frank Pasquale, famously called a networked society based on opaque algorithmic systems a "black box society". Five years later, and the metaphor, unfortunately, still holds – and applies to all the countries we studied for this report, across the board: there is not enough transparency concerning ADM systems – neither in the public, nor the private sector. Poland even mandates opacity, with the law that introduced its automated system to detect bank accounts used for illegal activities ("STIR"). The law states that the disclosure of adopted algorithms and risk indicators may [illegible]in up to 5 years in jail.

[illegible]ile we firmly reject the idea that all such systems are in[illegible]erently bad – we embrace an evidence-based perspective instead – it is undoubtedly bad to be unable to assess their functioning and impact based on accurate and factual knowledge. If only because opacity severely impedes the gathering of evidence that is necessary to come to an informed judgment on the deployment of an ADM system in the first place.

5 As detailed in the chapter on Europe

When coupled with the difficulty both our researchers and journalists found in accessing any meaningful data on these systems, this paints a troubling scenario for whoever wishes to keep them in check and guarantee that their deployment is compatible with fundamental rights, the rule of law, and democracy.

/ Challenging the algorithmic status quo

What is the European Union doing about this? Even though the strategic documents produced by the EU Commission, under the guidance of Ursula Von der Leyen, refer to "artificial intelligence" rather than ADM systems directly, they do state laudable intentions: promoting and realizing a "trustworthy AI" that puts "people first"[6].

However, as described in the EU chapter, the EU's overall approach prioritizes the commercial and geopolitical imperative to lead the "AI revolution" over making sure that its products are consistent with democratic safeguards, once adopted as policy tools.

This lack of political courage, which is most apparent in the decision to ditch any suggestion of a moratorium on live face recognition technologies in public places in its AI regulation package, is surprising. Especially at a time when many Member States are witnessing an increasing number of legal challenges – and defeats – over hastily deployed ADM systems that have negatively impacted the rights of citizens.

A landmark case comes from the Netherlands, where civil rights activists took an invasive and opaque automated system, supposed to detect welfare fraud (SyRI), to court and won. Not only was the system found in violation of the European Convention on Human Rights by the court of The Hague in February, and therefore halted. The case also set a precedent: according to the ruling, governments have a "special responsibility" to safeguard human rights when implementing such ADM systems. Providing much-needed transparency is considered a crucial part of this.

Since our first report, media and civil society activists have established themselves as a driving force for accountability in ADM systems. In Sweden, for example, journalists managed to force the release of the code behind the Trelleborg system for fully automated decisions related to social benefit applications. In Berlin, the Südkreuz train station face recognition pilot project failed to lead to the implementation of the system anywhere in Germany. This was thanks to the loud opposition of activists, so loud that they managed to influence party positions and, ultimately, the government's political agenda.

Greek activists from Homo Digitalis showed that no real traveler participated in the Greek pilot trials of a system called 'iBorderCtrl', an EU-funded project that aimed to use ADM to patrol borders, thus revealing that the capabilities of many such systems are frequently oversold. Meanwhile, in Denmark, a profiling system for the early detection of risks associated with vulnerable families and children (the so-called "Gladsaxe model") was put on hold thanks to the work of academics, journalists, and the Data Protection Authority (DPA).

DPAs themselves played an important role in other countries too. In France, the national privacy authority ruled that both a sound surveillance project and one for face recognition in high schools were illegal. In Portugal, the DPA refused to approve the deployment of video surveillance systems by the police in the municipalities of Leiria and Portimão as it was deemed disproportionate and would have amounted to "large-scale systematic monitoring and tracking of people and their habits and behavior, as well as identifying people from data relating to physical characteristics". And, in the Netherlands, the Dutch DPA asked for more transparency in predictive algorithms used by government agencies.

Lastly, some countries have referred to an ombudsperson for advice. In Denmark, this advice helped to develop strategies and ethical guidelines for the use of ADM systems in the public sector. In Finland, the deputy parliamentary ombudsperson considered automated tax assessments unlawful.

And yet, given the continued deployment of such systems throughout Europe, one is left wondering: is this level of oversight enough? When the Polish ombudsperson questioned the legality of the smile detection system used in a bank (and mentioned above), the decision did not prevent a later pilot in the city of Sopot, nor did it stop several companies from showing an interest in adopting the system.

6 Cfr. the chapter on Europe, and in particular the section on the EU Commission's 'White Paper on AI'

A startling finding in this report is that, while change happened rapidly regarding the deployment of ADM systems, the same is not true when it comes to the transparency of these systems.

/ Lack of adequate auditing, enforcement, skills, and explanations

Activism is mostly a reactive endeavor. Most of the time, activists can only react if an ADM system is being trialed or if one has already been deployed. By the time citizens can organize a response, their rights may have been infringed upon unnecessarily. This can happen even with the protections that should be granted, in most cases, by EU and Member States' law. This is why proactive measures to safeguard rights – before pilots and deployments take place – are so important.

And yet, even in countries where protective legislation is in place, enforcement is just not happening. In Spain, for example, "automated administrative action" is legally codified, mandating specific requirements in terms of quality control and supervision, together with the audit of the information system and its source code. Spain also has a Freedom of Information law. However, even with these laws, only rarely, our researcher writes, do public bodies share detailed information about the ADM systems they use. Similarly, in France, a 2016 law exists that mandates algorithmic transparency, but again, to no avail.

Even bringing an algorithm to court, according to the specific provisions of an algorithmic transparency law, may not be enough to enforce and protect users' rights. As the case of the Parcoursup algorithm to sort university applicants in France shows[7], exceptions can be carved out at will to shield an administration from accountability.

This is especially troubling when coupled with the endemic lack of skills and competences around ADM systems in the public sector lamented by many researchers. How could public officials explain or provide transparency of any kind around systems they don't understand?

Recently, some countries tried to address this issue. Estonia, for example, set up a competence center suited to ADM systems to better look into how they could be used to develop public services and, more specifically, to inform the operations of the Ministry of Economic Affairs and Communications and the State Chancellery for the development of e-government. Switzerland also called for the creation of a "competence network" within the broader framing of the "Digital Switzerland" national strategy.

And yet, the lack of digital literacy is a well-known issue affecting a large proportion of the population in several European countries. Besides, it is tough to call for the enforcement of rights you don't know you have. Protests in the UK and elsewhere, together with high profile scandals based on ADM systems[8], have certainly raised awareness of both the risks and opportunities of automating society. But while on the rise, this awareness is still in its early stages in many countries.

The results from our research are clear: while ADM systems already affect all sorts of activities and judgments, they are still mainly deployed without any meaningful democratic debate. Also, it is the norm, rather than the exception, that

7 Cfr. the chapter on France

8 Think of the "Buona Scuola" algorithm debacle in Italy, cfr. the chapter on Italy.

enforcement and oversight mechanisms – if they even exist – lag behind deployment.

Even the purpose of these systems is not commonly justified or explained to affected populations, not to mention the benefits they are supposed to gain. Think of the "AuroraAI" proactive service in Finland: it is supposed to automatically identify "life events", as our Finnish researchers report, and in the minds of proponents, it should work as "a nanny" that helps citizens meet particular public service needs that may arise in conjunction with certain life circumstances, e.g., moving to a new place, changing family relations, etc. "Nudging" could be at work here, our researchers write, meaning that instead of empowering individuals, the system might end up doing just the opposite, suggesting certain decisions or limiting an individual's options through its own design and architecture.

It is then all the more important to know what it is that is being "optimized" in terms of public services: "is service usage maximized, are costs minimized, or is citizen well-being improved?", ask the researchers. "What set of criteria are these decisions based on and who choses them?" The mere fact that we don't have an answer to these fundamental questions speaks volumes about the degree of participation and transparency that is allowed, even for such a potentially invasive ADM system.

/ The techno-solutionist trap

There is an overarching ideological justification for all this. It is called "technological solutionism", and it still severely affects the way in which many of the ADM systems we studied are developed. Even if the term has been long-denounced as a flawed ideology that conceives of every social problem as a "bug" in need of a "fix" through technology[9], this rhetoric is still widely adopted – both in the media and in policy circles – to justify the uncritical adoption of automated technologies in public life.

When touted as "solutions", ADM systems immediately veer into the territory described in Arthur C. Clarke's Third Law: magic. And it is difficult, if not impossible, to regulate magic, and even more so to provide transparency and explanations around it. One can see the hand reaching inside the hat, and a bunny appears as a result, but the process is and *should remain* a "black box".

Many researchers involved in the 'Automating Society' project denounced this as the fundamental flaw in the reasoning behind many of the ADM systems they describe. This also implies, as shown in the chapter on Germany, that most critiques of such systems are framed as an all-out rejection of "innovation", portraying digital rights advocates as "neo-luddites". This not only ignores the historical reality of the Luddite movement, which dealt in labor policies and not technologies *per se*, but also, and more fundamentally, threatens the effectiveness of hypothesized oversight and enforcement mechanisms.

At a time when the "AI" industry is witnessing the emergence of a "lively" lobbying sector, most notably in the UK, this might result in "ethics-washing" guidelines and other policy responses that are ineffective and structurally inadequate to address the human rights implications of ADM systems. This view ultimately amounts to the assumption that we humans should adapt to ADM systems, much more than ADM systems should be adapted to democratic societies.

To counter this narrative, we should not refrain from foundational questions: whether ADM systems can be compatible with democracy and deployed for the benefit of society at large, and not just for parts of it. It might be the case, for example, that certain human activities – e.g., those concerning social welfare – should not be subject to automation, or that certain technologies – namely, live face recognition in public spaces – should not be promoted in an endless quest for "AI leadership", but banned altogether instead.

Even more importantly, we should reject any ideological framing that prevents us from posing such questions. On the contrary: what we need to see now is actual policies changing – in order to allow greater scrutiny of these systems. In the following section we list the key demands that result from our findings. We hope that they will be widely discussed, and ultimately implemented.

Only through an informed, inclusive, and evidence-based democratic debate can we find the right balance between the benefits that ADM systems can – and do – provide in terms of speed, efficiency, fairness, better prevention, and access to public services, and the challenges they pose to the rights of us all.

9 See Evgeny Morozov (2014), *To Save Everything, Click Here. The Folly of Technological Solutionism*, Public Affairs, https://www.publicaffairsbooks.com/titles/evgeny-morozov/to-save-everything-click-here/9781610393706/

Policy Recommendations

In light of the findings detailed in the 2020 edition of the Automating Society report, we recommend the following set of policy interventions to policymakers in the EU parliament and Member States' parliaments, the EU Commission, national governments, researchers, civil society organizations (advocacy organizations, foundations, labor unions, etc.), and the private sector (companies and business associations). The recommendations aim to better ensure that ADM systems currently being deployed and those about to be implemented throughout Europe are effectively consistent with human rights and democracy:

1. Increase the transparency of ADM systems

Without the ability to know precisely how, why, and to what end ADM systems are deployed, all other efforts for the reconciliation of fundamental rights and ADM systems are doomed to fail.

/ Establish public registers for ADM systems used within the public sector

We, therefore, ask for legislation to be enacted at the EU level to mandate that Member States establish public registers of ADM systems used by the public sector.

They should come with the legal obligation for those responsible for the ADM system to disclose and document the purpose of the system, an explanation of the model (logic involved), and information about who developed the system. This information has to be made available in an easily readable and accessible manner, including structured digital data based on a standardized protocol.

Public authorities have a particular responsibility to make the operational features of ADM systems deployed in public administration transparent. This was underlined by a recent administrative complaint in Spain, that argues that "any ADM system used by the public administration should be made public by default". If upheld, the ruling could become precedent in Europe.

Whereas disclosure schemes on ADM systems should be mandatory for the public sector in all cases, these transparency requirements should also apply to the use of ADM systems by private entities when an AI/ADM system has a significant impact on an individual, a specific group, or society at large.

/ Introduce legally-binding data access frameworks to support and enable public interest research

Increasing transparency not only requires disclosing information about a system's purpose, logic, and creator, as well as the ability to thoroughly analyze, and test a system's inputs and outputs. It also requires making training data and data results accessible to independent researchers, journalists, and civil society organizations for public interest research.

That's why we suggest the introduction of robust, legally-binding data access frameworks, focused explicitly on supporting and enabling public interest research and in full respect of data protection and privacy law.

Learning from existing best practices at the national and EU levels, such tiered frameworks should include systems of sanctions, checks and balances as well as regular reviews. As private data-sharing partnerships have illustrated, there are legitimate concerns regarding user privacy and the possible de-anonymization of certain kinds of data.

Policymakers should learn from health data sharing frameworks to facilitate privileged access to certain kinds of more granular data, while ensuring that personal data is adequately protected (e.g., through secure operating environments).

While an effective accountability framework will require transparent access to platform data, this is a requirement for many auditing approaches to be effective as well.

2. Create a meaningful accountability framework for ADM systems

As findings from Spain and France have shown, even if transparency of an ADM system is required by law and/or information has been disclosed, this does not necessarily result in accountability. Further steps are needed to ensure that laws and requirements are actually enforceable.

/ Develop and establish approaches to effectively audit algorithmic systems

To ensure that transparency is meaningful, we need to complement the first step of establishing a public register by processes that effectively audit algorithmic systems.

The term "auditing" is widely used, but there is no common understanding of the definition. We understand auditing in this context in accordance with ISO's definition as a "systematic, independent and documented process for obtaining objective evidence and evaluating it objectively to determine the extent to which the audit criteria are fulfilled."[10]

We do not have satisfying answers to the complex questions[11] raised by the auditing of algorithmic systems yet; however, our findings clearly indicate the need to find answers in a broad, stakeholder engagement process and through thorough and dedicated research.

Both audit criteria and appropriate processes of auditing should be developed, following a multi-stakeholder approach that actively takes into consideration the disproportionate effect ADM systems have on vulnerable groups and solicits their participation.

We, therefore, ask policymakers to initiate such stakeholder processes in order to clarify the outlined questions, and to make available sources of funding aimed at enabling the participation by stakeholders who have so far been inadequately represented.

We furthermore demand the provision of adequate resources to support/fund research projects on developing models to effectively audit algorithmic systems.

/ Support civil society organizations as watchdogs of ADM systems

Our findings clearly indicate that the work of civil society organizations is crucial in effectively challenging opaque ADM systems. Through research and advocacy, and, often, in cooperation with academia and journalists, they repeatedly intervened in policy debates around those systems over recent years, in several cases effectively making sure that the public interest and fundamental rights are duly considered both before and after their deployment in many European countries.

Civil society actors should, therefore, be supported as watchdogs of the "automating society". As such, they are an integral component of any effective accountability framework for ADM systems.

/ Ban face recognition that might amount to mass surveillance

Not all ADM systems are equally dangerous, and a risk-based approach to regulation, such as Germany's and the EU's, correctly reflects this. But in order to provide workable accountability for systems that are identified as risky, effective oversight and enforcement mechanisms must be put in place. This is all the more important for those deemed at "high risk" of infringing on users' rights.

A crucial example that emerged from our findings is face recognition. ADM systems that are based on biometric technologies, including face recognition, have been shown to pose a particularly serious threat to the public interest and fundamental rights, as they clear the path to indiscriminate mass surveillance – and especially as they are widely, and opaquely, deployed nonetheless.

We demand that public uses of face recognition that might amount to mass surveillance are decisively banned until further notice, and urgently, at the EU level.

Such technologies may even be considered as already illegal in the EU, at least for certain uses, if deployed without "specific consent" of the scanned subjects. This legal reading has been suggested by the authorities in Belgium, who issued a landmark fine for face recognition deployments in the country.

10 https://www.iso.org/obp/ui/#iso:std:iso:19011:ed-3:v1:en

11 Thinking of potential models of algorithmic auditing, several questions emerge. 1) Who/what (services/platforms/products) should be audited? How to customize the auditing systems to the type of platform/type of service? 2) When should an audit be undertaken by a public institution (at EU level, national level, local level), and when can it be done by private entities/experts (business, civil society, researchers)? 3) How to clarify the distinction between assessing impact ex-ante (i.e. in the design phase) and ex-post (i.e. in operation) and the respective challenges? 4) How to assess trade-offs in the different virtues and vices of auditability? (e.g., simplicity, generality, applicability, precision, flexibility, interpretability, privacy, efficacy of an auditing procedure may be in tension). 5) Which information needs to be available for an audit to be effective and reliable (e.g., source code, training data, documentation)? Do auditors need to have physical access to systems during operation in order to audit effectively? 6) What obligation to produce proof is necessary and proportionate for vendors/service providers? 7) How can we ensure the auditing is possible? Do auditing requirements need to be considered in the design of algorithmic systems ("auditable by construction")? 8) Rules for publicity: When an audit is negative, and the problems are not solved, what should be the behavior of the auditor, in what way can it be made public that a failure occurred? 9) Who audits the auditors? How to make sure the auditors are held accountable?

3. Enhance algorithmic literacy and strengthen public debate on ADM systems

More transparency of ADM systems can only be truly useful if those confronted with them, such as regulators, government, and industry bodies, can deal with those systems and their impact in a responsible and prudent manner. In addition, those affected by these systems need to be able to understand, where, why, and how these systems are deployed. This is why we need to enhance algorithmic literacy at all levels, with important stakeholders as well as the general public, and to reinforce more diverse public debates about ADM systems and their impact on society.

/ Establish independent centers of expertise on ADM

Together with our demand for algorithmic auditing and supporting research, we call for the establishment of independent centers of expertise on ADM at the national level to monitor, assess, conduct research, report on, and provide advice to government and industry in coordination with regulators, civil society, and academia about the societal and human rights implications of the use of ADM systems. The overall role of these centers is to create a meaningful accountability system and to build capacity.

The national centers of expertise should involve civil society organizations, stakeholder groups, and existing enforcement bodies such as DPAs and national human rights bodies to benefit all aspects of the ecosystem and build trust, transparency, and cooperation between all actors.

As independent statutory bodies, the centers of expertise would have a central role in coordinating policy development and national strategies relating to ADM and in helping to build the capacity (competence/skills) of existing regulators, government, and industry bodies to respond to the increased use of ADM systems.

These centers should not have regulatory powers, but provide essential expertise on how to protect individual human rights and prevent collective and societal harm. They should, for instance, support small and medium-sized enterprises (SMEs) in fulfilling their obligations under human rights due diligence, including conducting human rights assessments or algorithmic impact assessments, and by registering ADM systems in the public register discussed above.

/ Promote an inclusive and diverse democratic debate around ADM systems

Next to strengthening capacities and competencies with those deploying ADM systems, it is also vital to advance algorithmic literacy in the general public through broader debate and diverse programs.

Our findings suggest that ADM systems not only remain non-transparent to the wider public when they are in use, but that even the decision whether or not to deploy an ADM system in the first place is usually taken without either the knowledge or participation of the public.

There is, therefore, an urgent need to include the public (interest) in the decision-making on ADM systems from the very beginning.

More generally, we need a more diverse public debate about the impact of ADM. We need to move beyond exclusively addressing expert groups and make the issue more accessible to the wider public. That means speaking a language other than the techno-judicial to engage the public and spark interest.

In order to do so, detailed programs – to build and advance digital literacy – should also be put in place. If we aim at enhancing an informed public debate and creating digital autonomy for citizens in Europe, we have to start by building and advancing digital literacy, with a specific focus on the social, ethical, and political consequences of adopting ADM systems.

Setting the stage for the future of ADM in Europe

As automated decision-making systems take center stage for distributing rights and services within Europe, institutions across the region increasingly recognize their role in public life, both in terms of opportunities and challenges.

By Kristina Penner and Fabio Chiusi

Since our first report in January 2019 – and even though the EU is still mired in the broader debate around "trustworthy" artificial intelligence – several bodies, from the EU Parliament to the Council of Europe, have published documents aimed at setting the EU and Europe on a course to deal with ADM over the coming years, if not decades.

In summer 2019, newly elected Commission President, Ursula von der Leyen, a self-stated "tech optimist", pledged to put forward "legislation for a coordinated European approach on the human and ethical implications of artificial intelligence" and to "regulate Artificial Intelligence (AI)" within 100 days of taking office. Instead, in February 2020, the European Commission published a 'White Paper' on AI containing "ideas and actions" – a strategy package that aims to inform citizens and pave the way for future legislative action. It also makes the case for European "technological sovereignty": in Von der Leyen's own terms, this translates into "the capability that Europe must have to make its own choices, based on its own values, respecting its own rules", and should "help make tech optimists of us all".

A second fundamental endeavor affecting ADM in Europe is the Digital Services Act (DSA), announced in Von der Leyen's 'Agenda for Europe' and supposed to replace the E-Commerce Directive that has been in place since 2000. It aims to "upgrade our liability and safety rules for digital platforms, services and products, and complete our Digital Single Market" – thus leading to foundational debates around the role of ADM in content moderation policies, intermediary liability, and freedom of expression more generally[1].

An explicit focus on ADM systems can be found in a Resolution approved by the EU Parliament's Internal Market and Consumer Protection Committee, and in a Recommendation "on the human rights impacts of algorithmic systems" by the Council of Europe's Committee of Ministers.

The Council of Europe (CoE), in particular, was found to be playing an increasingly important role in the policy debate on AI over the last year, and even though its actual impact on regulatory efforts remains to be seen, a case can be made for it to serve as the "guardian" of human rights. This is most apparent in the Recommendation, 'Unboxing Artificial Intelligence: 10 steps to protect Human Rights', by the CoE's Commissioner on Human Rights, Dunja Mijatović, and in the work of the Ad Hoc Committee on AI (CAHAI) founded in September 2019.

Many observers see a fundamental tension between business and rights imperatives in how EU institutions, and especially the Commission, are framing their reflections and

1 Detailed remarks and recommendations around ADM systems in the context of the DSA can be found in the outputs of AlgorithmWatch's 'Governing Platforms' project.

MANY OBSERVERS SEE A FUNDAMENTAL TENSION BETWEEN BUSINESS AND RIGHTS IMPERATIVES IN HOW EU INSTITUTIONS, AND ESPECIALLY THE COMMISSION, ARE FRAMING THEIR REFLECTIONS AND PROPOSALS ON AI AND ADM.

proposals on AI and ADM. On the one hand, Europe wants "to increase the use of, and demand for, data and data-enabled products and services throughout the Single Market"; thereby becoming a "leader" in business applications of AI, and boosting the competitiveness of EU firms in the face of mounting pressure from rivals in the US and China. This is all the more important for ADM, the assumption being that, through this "data-agile" economy, the EU "can become a leading role model for a society empowered by data to make better decisions – in business and the public sector". As the White Paper on AI puts it, "Data is the lifeblood of economic development".

Whereas on the other hand, the automatic processing of data about a citizen's health, job, and welfare can form decisions with discriminatory and unfair results. This "dark side" of algorithms in decision-making processes is tackled in the EU toolbox through a series of principles. In the case of high-risk systems, rules should guarantee that automated decision-making processes are compatible with human rights and meaningful democratic checks and balances. This is an approach that EU institutions label as "human-centric" and unique, and as fundamentally opposed to those applied in the US (led by profit) and China (led by national security and mass surveillance).

However, doubts have emerged as to whether Europe can attain both goals at the same time. Face recognition is a case in point: even though, as this report shows, we now have plenty of evidence of unchecked and opaque deployments in most member countries, the EU Commission has failed to act swiftly and decisively to protect the rights of European citizens. As leaked drafts of the EC White Paper on AI revealed[2], the EU was about to ban "remote biometric identification" in public places, before shying away at the last minute and promoting a "broad debate" around the issue instead.

In the meantime, controversial applications of ADM for border controls, even including face recognition, are still being pushed in EU-funded projects.

"WE WANT TO ENCOURAGE OUR BUSINESSES, OUR RESEARCHERS, THE INNOVATORS, THE ENTREPRENEURS, TO DEVELOP ARTIFICIAL INTELLIGENCE. AND WE WANT TO ENCOURAGE OUR CITIZENS TO FEEL CONFIDENT TO USE IT. WE HAVE TO UNLEASH THIS POTENTIAL."
URSULA VON DER LEYEN

Policies and political debates

/ The European Data Strategy Package and the White Paper on AI

While the promised comprehensive legislation "for a coordinated European approach on the human and ethical implications of artificial intelligence", announced in Von der Leyen's "Agenda for Europe", has not been put forward within her "first 100 days in office", the EU Commission published a series of documents that provide a set of principles and ideas to inform it.

On February 19, 2020, a "European Strategy for Data" and a "White Paper on Artificial Intelligence" were jointly published, laying out the main principles of the EU's strategic approach to AI (including ADM systems, even though they are not explicitly mentioned). These principles include, putting "people first" ("technology that works for the people"), technological neutrality (no technology is good or bad per se; this is to be determined by its use only) and, of course, sovereignty, and optimism. As Von der Leyen puts it: "We want to encourage our businesses, our researchers, the innovators, the entrepreneurs, to develop Artificial Intelligence. And we want to encourage our citizens to feel confident to use it. We have to unleash this potential".

The underlying idea is that new technologies should not come with new values. The "new digital world" envisioned by the Von der Leyen administration should fully protect human and civil rights. "Excellence" and "trust", highlighted in the very title of the White Paper, are considered the twin pillars upon which a European model of AI can and should stand, differentiating it from the strategies of both the US and China.

However, this ambition is lacking in the detail of the White Paper. For example, the White Paper lays out a risk-based approach to AI regulation, in which regulation is proportional to the impact of "AI" systems on citizen's lives. "For high-risk cases, such as in health, policing, or transport", it reads, "AI systems

2 https://www.politico.eu/article/eu-considers-temporary-ban-on-facial-recognition-in-public-spaces/

should be transparent, traceable and guarantee human oversight". Testing and certification of adopted algorithms are also included among the safeguards that should be put in place, and should become as widespread as for "cosmetics, cars or toys". Whereas, "less risky systems" only have to follow voluntary labelling schemes instead: "The economic operators concerned would then be awarded a quality label for their AI applications".

But critics noted that the very definition of "risk" in the Paper is both circular and too vague, allowing for several impactful ADM systems to fall through the cracks of the proposed framework[3].

Comments[4] gathered in the public consultation, between February and June 2020, highlight how controversial this idea is. 42.5% of respondents agreed that "compulsory requirements" should be limited to "high-risk AI applications", while 30.6% doubted such a limitation.

Moreover, there is no description of a clear mechanism for the enforcement of such requirements. Neither is there a description of a process to move towards one.

The consequences are immediately visible for biometric technologies, and face recognition in particular. On this, the White Paper proposed a distinction between biometric "authentication", which is seen as non-controversial (e.g., face recognition to unlock a smartphone), and remote biometric "identification" (such as deployment in public squares to identify protesters), which could arouse serious human rights and privacy concerns.

Only cases in the latter category would be problematic under the scheme proposed by the EU. The FAQ in the White Paper states: "this is the most intrusive form of facial recognition and in principle prohibited in the EU", unless there is "substantial public interest" in its deployment.

The explanatory document claims that "allowing facial recognition is currently the exception", but findings in this report arguably contradict that view: face recognition seems to be rapidly becoming the norm. A leaked draft version of the White Paper seemed to recognize the urgency of the problem, by including the idea of a three-to-five-year moratorium on live uses of face recognition in public places, until – and if – a way to reconcile them with democratic checks and balances could be found.

Just before the official release of the White Paper, even EU Commissioner, Margrethe Vestager, called for a "pause" on these uses.

However, immediately after Vestager's call, Commission officials added that this "pause" would not prevent national governments from using face recognition according to the existing rules. Ultimately, the final draft of the White Paper scrapped any mention of a moratorium, and called for "a broad European debate on the specific circumstances, if any, which might justify" its use for live biometric identification purposes instead. Among them, the White Paper includes justification, proportionality, the existence of democratic safeguards, and respect for human rights.

Throughout the whole document, risks associated with AI-based technologies are more generally labeled as "potential", while the benefits are portrayed as very real and immediate. This led many[5] in the human rights community to claim that the overall narrative of the White Paper suggests a worrisome reversal of EU priorities, putting global competitiveness ahead of the protection of fundamental rights.

Some foundational issues are, however, raised in the documents. For example, the interoperability of such solutions and the creation of a network of research centers focused on applications of AI aimed at "excellence" and competence-building.

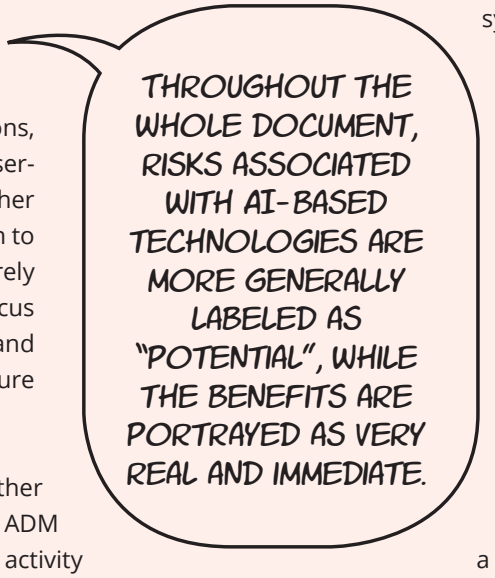
The objective is "to attract over €20 billion of total investments in the EU per year in AI over the next decade".

3 "To give two examples: VioGén, an ADM system to forecast gender-based violence, and Ghostwriter, an application to detect exam fraud, would most likely fall between the cracks of regulation, even though they come with tremendous risks" (https://algorithmwatch.org/en/response-european-commission-ai-consultation/)

4 "In total, 1215 contributions were received, of which 352 were on behalf of a company or business organisations/associations, 406 from citizens (92% EU citizens), 152 on behalf of academic/research institutions, and 73 from public authorities. Civil society voices were represented by 160 respondents (among which 9 consumer's organisations, 129 non-governmental organisations and 22 trade unions). 72 respondents contributed as "others"". Comments arrived "from all over the world", including countries such as "India, China, Japan, Syria, Iraq, Brazil, Mexico, Canada, the US and the UK". (from the Consultation's Summary Report, linked here: https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence)

5 Among them: Access Now (https://www.accessnow.org/cms/assets/uploads/2020/05/EU-white-paper-consultation_AccessNow_May2020.pdf), AI Now (https://ainowinstitute.org/ai-now-comments-to-eu-whitepaper-on-ai.pdf), EDRI (https://edri.org/can-the-eu-make-ai-trustworthy-no-but-they-can-make-it-just/) – and AlgorithmWatch (https://algorithmwatch.org/en/response-european-commission-ai-consultation/).

A certain technological determinism seems to also affect the White Paper. "It is essential", it reads, "that public administrations, hospitals, utility and transport services, financial supervisors, and other areas of public interest rapidly begin to deploy products and services that rely on AI in their activities. A specific focus will be in the areas of healthcare and transport where technology is mature for large-scale deployment."



However, it remains to be seen whether suggesting a rushed deployment of ADM solutions in all spheres of human activity is compatible with the EU Commission's efforts in addressing the structural challenges brought about by ADM systems to rights and fairness.

/ EU Parliament's Resolution on ADM and consumer protection

A Resolution, passed by the EU Parliament in February 2020, more specifically tackled ADM systems in the context of consumer protection. The Resolution correctly pointed out that "complex algorithmic-based systems and automated decision-making processes are being made at a rapid pace", and that "opportunities and challenges presented by these technologies are numerous and affect virtually all sectors". The text also highlights the need for "an examination of the current EU legal framework", to assess whether "it is able to respond to the emergence of AI and automated decision-making".

Calling for a "common EU approach to the development of automated decision-making processes", the Resolution details several conditions that any such systems should possess to remain consistent with European values. Consumers should be "properly informed" about how algorithms affect their lives, and they should have access to a human with decision-making power so that decisions can be checked and corrected if needed. They should also be informed, "when prices of goods or services have been personalized on the basis of automated decision-making and profiling of consumer behavior".

In reminding the EU Commission that a carefully drafted risk-based approach is needed, the Resolution points out that safeguards need to take into consideration that ADM systems "may evolve and act in ways not envisaged when first placed on the market", and that liability is not always easy to attribute when harm comes as a result of the deployment of an ADM system.

The Resolution echoes art. 22 of the GDPR when it notes that a human subject must always be in the loop when "legitimate public interests are at stake", and should always be ultimately responsible for decisions in the "medical, legal and accounting professions, and for the banking sector". In particular, a "proper" risk assessment should precede any automation of professional services.

Finally, the Resolution lists detailed requirements for quality and transparency in data governance: among them, "the importance of using only high-quality and unbiased data sets in order to improve the output of algorithmic systems and boost consumer trust and acceptance"; using "explainable and unbiased algorithms"; and the need for a "review structure" that allows affected consumers "to seek human review of, and redress for, automated decisions that are final and permanent".

/ Making the most of the EU Parliament's "right of initiative"

In her inaugural address, von der Leyen clearly expressed her support for a "right of initiative" for the European Parliament. "When this House, acting by majority of its Members, adopts Resolutions requesting the Commission to submit legislative proposals", she said, "I commit to responding with a legislative act in full respect of the proportionality, subsidiarity, and better law-making principles".

If "AI" is indeed a revolution requiring a dedicated legislative package, allegedly coming over the first quarter of 2021, elected representatives want to have a say about it. This, coupled with von der Leyen's stated intent of empowering their legislative capabilities, could even result in what Politico labeled a "Parliament moment", with parliamentary committees starting to draft several different reports as a result.

Each report investigates specific aspects of automation in public policy that, even though they are meant to shape upcoming "AI" legislation, are relevant for ADM.

IF "AI" IS INDEED A REVOLUTION REQUIRING A DEDICATED LEGISLATIVE PACKAGE, ELECTED REPRESENTATIVES WANT TO HAVE A SAY ABOUT IT.

For example, through its "Framework of ethical aspects of artificial intelligence, robotics and related technologies", the Committee for Legal Affairs calls for the constitution of a "European Agency for Artificial Intelligence" and, at the same time, for a network of national supervisory authorities in each Member State to make sure that ethical decisions involving automation are and remain ethical.

In "Intellectual property rights for the development of artificial intelligence technologies", the same committee lays out its view for the future of intellectual property and automation. For one, the draft report states that "mathematical methods are excluded from patentability unless they constitute inventions of a technical nature", while at the same time claiming that, regarding algorithmic transparency, "reverse engineering is an exception to the trade secrets rule".

The report goes as far as considering how to protect "technical and artistic creations generated by AI, in order to encourage this form of creation", imagining that "certain works generated by AI can be regarded as equivalent to intellectual works and could therefore be protected by copyright".

Lastly, in a third document ("Artificial Intelligence and civil liability"), the Committee details a "Risk Management Approach" for the civil liability of AI technologies. According to it, "the party who is best capable of controlling and managing a technology-related risk is held strictly liable, as a single entry point for litigation".

Important principles concerning the use of ADM in the criminal justice system can be found in the Committee of Civil Liberties, Justice and Home Affairs' report on "Artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters". After a detailed list of actual and current uses of "AI" – these are, actually, ADM systems – by police forces[6], the Committee "considers it necessary to create a clear and fair regime for assigning legal responsibility for the potential adverse consequences produced by these advanced digital technologies".

It then goes about detailing some of its features: no fully automated decisions[7], algorithmic explainability that is "intelligible to users", a "compulsory fundamental rights impact assessment (...) of any AI systems for law enforcement or judiciary" prior to its deployment or adoption, plus "periodic mandatory auditing of all AI systems used by law enforcement and the judiciary to test and evaluate algorithmic systems once they are in operation".

6 On p. 5, the report states: "AI applications in use by law enforcement include applications such as facial recognition technologies, automated number plate recognition, speaker identification, speech identification, lip-reading technologies, aural surveillance (i.e. gunshot detection algorithms), autonomous research and analysis of identified databases, forecasting (predictive policing and crime hotspot analytics), behaviour detection tools, autonomous tools to identify financial fraud and terrorist financing, social media monitoring (scraping and data harvesting for mining connections), international mobile subscriber identity (IMSI) catchers, and automated surveillance systems incorporating different detection capabilities (such as heartbeat detection and thermal cameras)".

7 "in judicial and law enforcement contexts, the final decision always needs to be taken by a human" (p. 6)

A moratorium on face recognition technologies for law enforcement is also called for in the report, "until the technical standards can be considered fully fundamental rights compliant, results derived are non-discriminatory, and there is public trust in the necessity and proportionality for the deployment of such technologies".

The aim is to eventually boost the overall transparency of such systems, and advising Member States to provide a "comprehensive understanding" of the AI systems adopted by law enforcement and the judiciary, and – along the lines of a "public register" – to detail "the type of tool in use, the types of crime they are applied to, and the companies whose tools are being used".

The Culture and Education Committee and the Industrial Policy Committee were also working on their own reports at the time of writing.

All these initiatives led to the creation of a Special Committee on "Artificial Intelligence in a Digital Age" (AIDA) on June 18, 2020. Composed of 33 members, and with an initial duration of 12 months, it will "analyse the future impact" of AI on the EU economy, and "in particular on skills, employment, fintech, education, health, transport, tourism, agriculture, environment, defence, industry, energy and e-government".

/ High-Level Expert Group on AI & AI Alliance

In 2018, the High-Level Expert Group (HLEG) on AI, an expert committee made up of 52 experts, was set up by the European Commission to support the implementation of the European strategy on AI, to identify principles that should be observed in order to achieve "trustworthy AI", and, as the steering committee of the supporting AI Alliance, to create an open multi-stakeholder platform (consisting of more than 4000 members at the time of writing) to provide broader input for the work of the AI high-level expert group.

After the publication of the first draft of the AI Ethics Guidelines for Trustworthy AI in December 2018, followed by feedback from more than 500 contributors, a revised version was published in April 2019. It puts forward a "human-centric approach" to achieve legal, ethical, and robust AI throughout the system's entire life cycle. However, it remains a voluntary framework without concrete and applicable recommendations on operationalization, implementation, and enforcement.

Civil society, consumer protection, and rights organizations commented and called for the translation of the guidelines into tangible rights for people[8]. For example, digital rights non-profit Access Now, a member of the HLEG, urged the EC to clarify how different stakeholders can test, apply, improve, endorse, and enforce "Trustworthy AI" as a next step, while at the same time recognizing the need to determine Europe's red lines.

In an op-ed, two other members of the HLEG claimed that the group had "worked for one-and-a-half years, only for its detailed proposals to be mostly ignored or mentioned only in passing" by the European Commission who drafted the final version.[9] They also argued that, because the group was initially tasked with identifying risks and "red lines" for AI, members of the group pointed to autonomous weapon systems, citizen scoring, and automated identification of individuals by using face recognition as implementations of AI that should be avoided. However, representatives of industry, who dominate the committee[10], succeeded in getting these principles deleted before the draft was published.

This imbalance towards highlighting the potentials of ADM, compared to the risks, can also be observed throughout its second deliverable. In the HLEG's "Policy and investment recommendations for trustworthy AI in Europe", made public in June 2019, there are 33 recommendations meant to "guide Trustworthy AI towards sustainability, growth and competitiveness, as well as inclusion – while empowering, benefiting and protecting human beings". The document is predominantly a call to boost the uptake and scaling of AI in the private and public sector by investing in tools and applications "to help vulnerable demographics" and "to leave no one behind".

Nevertheless, and despite all legitimate criticism, both guidelines still express critical concerns and demands regarding automated decision-making systems. For example, the ethics guidelines formulate "seven key requirements that AI systems should meet in order to be trustworthy". These guidelines go on to provide guidance for the prac-

8 E.g., the European Consumer Protection Organization (BEUC): https://www.beuc.eu/publications/beuc-x-2020-049_response_to_the_ecs_white_paper_on_artificial_intelligence.pdf

9 Mark Coeckelbergh and Thomas Metzinger: Europe needs more guts when it comes to AI ethics, https://background.tagesspiegel.de/digitalisierung/europe-needs-more-guts-when-it-comes-to-ai-ethics

10 The group was composed of 24 business representatives, 17 academics, five civil society organizations and six other members, like the European Union Agency for Fundamental Rights

tical implementation of each requirement: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability.

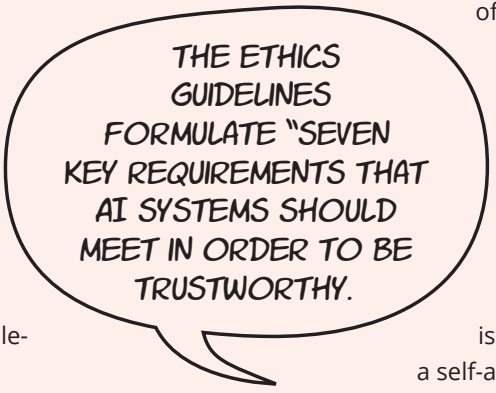
The guidelines also provide a concrete pilot, called the "Trustworthy AI assessment list", which is aimed at making those high-level principles operational. The goal is to have it adopted "when developing, deploying or using AI systems", and adapted "to the specific use case in which the system is being applied".

The list includes many issues that are associated with the risk of infringing on human rights through ADM systems. These include the lack of human agency and oversight, technical robustness and safety issues, the inability to avoid unfair bias or provide equal and universal access to such systems, and the lack of meaningful access to data fed into them.

Contextually, the pilot list included in the guidelines provides useful questions to help those who deploy ADM systems. For example, it calls for "a fundamental rights impact assessment where there could be a negative impact on fundamental rights". It also asks whether "specific mechanisms of control and oversight" have been put in place in the cases of "self-learning or autonomous AI" systems, and whether processes exist "to ensure the quality and integrity of your data".

Detailed remarks also concern foundational issues for ADM systems, such as their transparency and explainability. Questions include "to what extent the decisions and hence the outcome made by the AI system can be understood?" and "to what degree the system's decision influences the organisation's decision-making processes?" These questions are highly relevant to assess the risks posed by deploying such systems.

To avoid bias and discriminatory outcomes, the guidelines point to "oversight processes to analyze and address the system's purpose, constraints, requirements and decisions in a clear and transparent manner", while at the same time demanding stakeholder participation throughout the whole process of implementing AI systems.



Added to that, the recommendations on policy and investment foresee the determination of red lines through an institutionalized "dialogue on AI policy with affected stakeholders", including experts in civil society. Furthermore, they urge to "ban AI-enabled mass scale scoring of individuals as defined in [the] Ethics Guidelines, and [to] set very clear and strict rules for surveillance for national security purposes and other purposes claimed to be in the public or national interest". This ban would include biometric identification technologies and profiling.

Relevant to automated decision-making systems, the document also states that "clearly defin[ing] if, when and how AI can be used (...) will be crucial for the achievement of Trustworthy AI", warning that "any form of citizen scoring can lead to the loss of [the citizen's] autonomy and endanger the principle of non-discrimination", and "therefore should only be used if there is a clear justification, under proportionate and fair measures". It further stresses that "transparency cannot prevent non-discrimination or ensure fairness". This means that the possibility of opting out of a scoring mechanism should be provided, ideally without any detriment to the individual citizen.

On the one hand, the document acknowledges "that, while bringing substantial benefits to individuals and society, AI systems also pose certain risks and may have a negative impact, including impacts which may be difficult to anticipate, identify or measure (e.g. on democracy, the rule of law and distributive justice, or on the human mind itself.)". On the other, however, the group claims that "unnecessarily prescriptive regulation should be avoided".

In July 2020, the AI HLEG also presented their final Assessment List for Trustworthy Artificial Intelligence (ALTAI), compiled after a piloting process together with 350 stakeholders.

The list, which is entirely voluntary and devoid of any regulatory implications, aims to translate the seven requirements laid out in the AI HLEG's Ethics Guidelines into action. The intention is to provide whoever wants to implement AI solutions that are compatible with EU values – for example, designers and developers of AI systems, data scientists, procurement officers or specialists, and legal/compliance officers – with a self-assessment toolkit.

/ Council of Europe: how to safeguard human rights in ADM

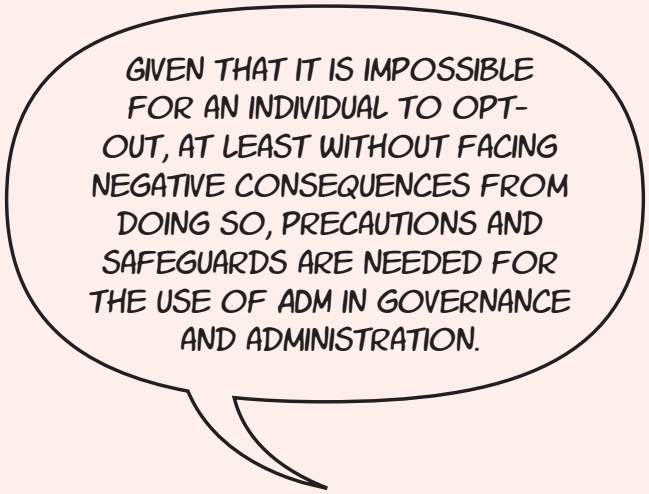
In addition to the Ad hoc Committee on Artificial Intelligence (CAHAI), set up in September 2019, the Committee of Ministers of the Council of Europe[11] has published a substantial and persuasive framework.

Envisioned as a standard-setting instrument, its "Recommendation to Member states on the human rights impacts of algorithmic systems" describes "significant challenges" [12] that arise with the emergence and our "increasing reliance" on such systems, and that are relevant "to democratic societies and the rule of law".

The framework, which underwent a public consultation period with detailed comments from civil society organizations, goes beyond the EU Commission's White Paper when it comes to safeguarding values and human rights.

The Recommendation thoroughly analyzes the effects and evolving configurations of algorithmic systems (Appendix A) by focusing on all stages of the process that go into making an algorithm, i.e., procurement, design, development, and ongoing deployment.

While generally following the 'human-centric AI approach' of the HLEG guidelines, the Recommendation outlines actionable "obligations of States" (Appendix B) as well as responsibilities for private sector actors (Appendix C). In addition, the Recommendation adds principles such as "informational self-determination"[13], lists detailed suggestions for accountability mechanisms and effective remedies, and demands human rights impact assessments.



Even though the document clearly acknowledges the "significant potential of digital technologies to address societal challenges and to socially beneficial innovation and economic development", it also urges caution. This is to ensure that those systems do not deliberately or accidentally perpetuate "racial, gender and other societal and labor force imbalances that have not yet been eliminated from our societies".

On the contrary, algorithmic systems should be used proactively and sensitively to address these imbalances, and pay "attention to the needs and voices of vulnerable groups".

Most significantly, however, the Recommendation identifies the potentially higher risk to human rights when algorithmic systems are used by Member States to deliver public services and policy. Given that it is impossible for an individual to opt-out, at least without facing negative consequences from doing so, precautions and safeguards are needed for the use of ADM in governance and administration.

11 The CoE is both "a governmental body where national approaches to European problems are discussed on an equal footing and a forum to find collective responses to these challenges." Its work includes "the political aspects of European integration, safeguarding democratic institutions and the rule of law and protecting human rights – in other words, all problems which require concerted pan-European solutions." Although the recommendations to the governments of members are not binding, in some cases the Committee may request the governments of members to inform it of the action taken by them with regard to such recommendations. (Art. 15b Statute). Relations between the Council of Europe and the European Union are set out in the (1) Compendium of Texts governing the relations between the Council of Europe and the European Union and in the (2) Memorandum of Understanding between the Council of Europe and the European Union.

12 Under the supervision of the Steering Committee on Media and Information Society (CDMSI) and prepared by the Committee of Experts on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence (MSI-AUT).

13 "States should ensure that all design, development and ongoing deployment of algorithmic systems provide an avenue for individuals to be informed in advance about the related data processing (including its purposes and possible outcomes) and to control their data, including through interoperability", reads Section 2.1 of Appendix B.

The Recommendation also addresses the conflicts and challenges arising from public-private-partnerships ("neither clearly public nor clearly private") in a wide range of uses.

Recommendations for Member State governments include abandoning processes and refusing to use ADM systems, if "human control and oversight become impractical" or human rights are put at risk; deploying ADM systems if and only if transparency, accountability, legality, and the protection of human rights can be guaranteed "at all stages of the process". Furthermore, the monitoring and evaluation of these systems should be "constant", "inclusive and transparent", comprised of a dialogue with all relevant stakeholders, as well as an analysis of the environmental impact and further potential externalities on "populations and environments".

In Appendix A, the COE also defines "high-risk" algorithms for other bodies to draw inspiration from. More specifically, the Recommendation states that "the term "high-risk" is applied when referring to the use of algorithmic systems in processes or decisions that can produce serious consequences for individuals or in situations where the lack of alternatives prompts a particularly high probability of infringement of human rights, including by introducing or amplifying distributive injustice".

The document, which did not require the unanimity of members to be adopted, is non-binding.

/ Regulation of terrorist content online

After a long period of sluggish progress, the regulation to prevent the dissemination of terrorist content gained momentum in 2020. Should the adopted regulation still include automated and proactive tools for recognizing and removing content online, these would likely fall under Art. 22 of the GDPR.

As the European Data Protection Supervisor (EDPS) puts it: "since the automated tools, as envisaged by the Proposal, could lead not only to the removal and retention of content (and related data) concerning the uploader, but also, ultimately, to criminal investigations on him or her, these tools would significantly affect this person, impacting on his or her right to freedom of expression and posing significant risks for him or her rights and freedoms", and, therefore, fall under Art. 22(2).

Also, and crucially, it would require more substantive safeguards compared to those that the Commission currently foresees. As the advocacy group, European Digital Rights (EDRi), explains: "the proposed Terrorist Content Regulation needs substantive reform to live up to the Union's values, and to safeguard the fundamental rights and freedoms of its citizens".

An early stream of strong criticism on the initial proposal from civil society groups, European Parliament (EP) committees, including opinions and analysis by the European Union Agency for Fundamental Rights (FRA), EDRi, as well as a critical joint report by three UN Special Rapporteurs, highlighted threats to the right of freedom of expression and information, the right to freedom and pluralism of the media, the freedom to conduct a business and the rights to privacy and protection of personal data.

Critical aspects include an insufficient definition of terrorist content, the scope of the regulation (at present, this includes content for educational and journalistic purposes), the aforementioned call for "proactive measures", a lack of effective judicial supervision, insufficient reporting obligations for law enforcement agencies, and missing safeguards for "cases where there are reasonable grounds to believe that fundamental rights are impacted" (EDRi 2019).

The EDPS stresses that such "suitable safeguards" should include the right to obtain human intervention and the right to an explanation of the decision reached through automated means (EDRi 2019).

Although safeguards that were suggested or demanded found their way into the EP's draft report on the proposal, it is yet to be seen who can hold their breath longer going into the last round before the final vote. During closed-door trialogues between the EP, the new EC, and the EU Council (which began in October 2019), only minor changes are still possible, according to a leaked document.

Oversight and regulation

/ First decisions on the compliance of ADM systems with the GDPR

"Although there was no great debate on facial recognition during the passage of negotiations on the GDPR and the law enforcement data protection directive, the legislation was designed so that it could adapt over time as technologies evolved. [...] Now is the moment for the EU, as it discusses the ethics of AI and the need for regulation, to determine whether – if ever – facial recognition technology can be permitted in a democratic society. If the answer is yes, only then do we turn [to] questions of how and safeguards and accountability to be put in place." – EDPS, Wojciech Wiewiórowski.

"Facial recognition processing is an especially intrusive biometric mechanism, which bears important risks of privacy or civil liberties invasions for the people affected" – (CNIL 2019).

Since the last Automating Society report, we have seen the first cases of fines and decisions related to breaches of the regulation issued by national Data Protection Authorities (DPAs) based on the GDPR. The following case studies, however, show the limits of the GDPR in practice when it comes to Art. 22 relating to ADM systems and how it is leaving the privacy regulators to make assessments on a case-by-case basis.

In Sweden, a face recognition test project, conducted in one school class for a limited period of time, was found to violate several obligations of Data Protection Regulation (esp. GDPR Art. 2(14), Art. 9 (2)). (European Data Protection Board 2019).

A similar case is on hold after the French Commission Nationale de l'Informatique et des Libertés (CNIL raised concerns when two high schools planned to introduce face recognition technology in partnership with the US tech firm, Cisco. The opinion is non-binding, and the filed suit is ongoing[14].

The ex-ante authorization by data regulators is not required to conduct such trials as the consent of the users is generally considered to be sufficient to process biometric data. And yet, in the Swedish case, it wasn't. This was due to power imbalances between the data controller and the data subjects. Instead, an adequate impact assessment and prior consultation with the DPA were deemed necessary.

The European Data Protection Supervisor (EDPS) confirmed this:

"Consent would need to be explicit as well as freely-given, informed and specific. Yet unquestionably a person cannot opt out, still less opt in, when they need access to public spaces that are covered by facial recognition surveillance. [...] Finally, the compliance of the technology with principles like data minimization and the data protection by design obligation is highly doubtful. Facial recognition technology has never been fully accurate, and this has serious consequences for individuals being falsely identified whether as criminals or otherwise. [...] It would be a mistake, however, to focus only on privacy issues. This is fundamentally an ethical question for a democratic society." (EDPS 2019)

Access Now commented:

"As more facial recognition projects develop, we already see that the GDPR provides useful human rights safeguards that can be enforced against unlawful collection and use of sensitive data such as biometrics. But the irresponsible and often unfounded hype around the efficiency of such technologies and the underlying economic interest could lead to attempts by central and local governments and private companies to circumvent the law."

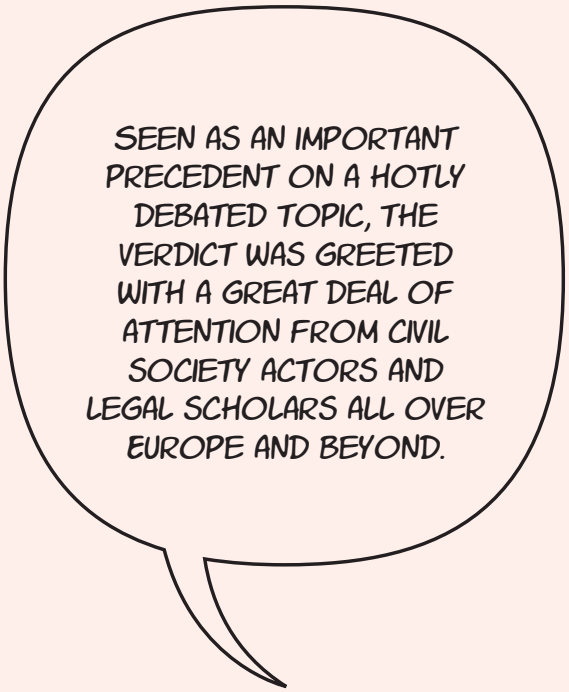
/ Automated Face Recognition in use by South Wales Police ruled unlawful

Over the course of 2020, the UK witnessed the first high profile application of the Law Enforcement Directive[15] concerning the use of face recognition technologies in public spaces by the police. Seen as an important precedent on a hotly debated topic, the verdict was greeted with a great deal of attention from civil society actors and legal scholars all over Europe and beyond [16].

14 See France chapter and (Kayalki 2019)

15 The Law Enforcement Directive, in effect from May 2018, "deals with the processing of personal data by data controllers for 'law enforcement purposes' – which falls outside of the scope of the GDPR". https://www.dataprotection.ie/en/organisations/law-enforcement-directive

16 The decision was rendered on September, 4th, 2019 by the High Court sitting in Cardiff in the case Bridges v. the South Wales Police (High Court of Justice 2019).



The case was brought to court by Ed Bridges, a 37 years old man from Cardiff, who claimed his face was scanned without his consent both during Christmas shopping in 2017, and at a peaceful anti-arms protest a year later.

The court initially upheld the use of Automated Facial Recognition technology ("AFR") by South Wales Police, declaring it lawful and proportionate. But the decision was appealed by Liberty, a civil rights group, and the Court of Appeal of England and Wales decided to overturn the High Court's dismissal and ruled it unlawful on August 11, 2020.

In ruling against South Wales Police on three out of five grounds, the Court of Appeal found "fundamental deficiencies" in the existing normative framework around the use of AFR, that its deployment did not meet the principle of "proportionality", and, also, that an adequate Data Protection Impact Assessment (DPIA) had not been performed, lacking multiple crucial steps.

The court did not, however, rule that the system was producing discriminatory results, based either on sex or race, as South Wales Police had not gathered sufficient evidence to make a judgment on that[17]. However, the court felt it was worth adding a noticeable remark: "We would hope that, as AFR is a novel and controversial technology, all police forces that intend to use it in the future would wish to satisfy themselves that everything reasonable which could be done had been done in order to make sure that the software used does not have a racial or gender bias."

After the ruling, Liberty called for South Wales Police and other police forces to withdraw the use of face recognition technologies.

ADM in practice: border management and surveillance

While the EU Commission and its stakeholders debated whether to regulate or ban face recognition technologies, extensive trials of the systems were already underway all over Europe.

This section highlights a crucial and often overlooked link between biometrics and the EU's border management systems, clearly showing how technologies that can produce discriminatory results might be applied to individuals – e.g., migrants – who already suffer the most from discrimination.

/ Face recognition and the use of biometrics data in EU policies and practice

Over the last year, face recognition and other kinds of biometrics identification technologies garnered a lot of attention from governments, the EU, civil society, and rights organizations, especially concerning law enforcement and border management.

17 The Police claimed it had no access to the demographic composition of the training dataset for the adopted algorithm, "Neoface". The Court notes that "the fact remains, however, that SWP have never sought to satisfy themselves, either directly or by way of independent verification, that the software program in this case does not have an unacceptable bias on grounds of race or sex".

Over 2019, the EC tasked a consortium of public agencies to "map the current situation of facial recognition in criminal investigations in all EU Member States," to move "towards the possible exchange of facial data". They commissioned the consultancy firm Deloitte to perform a feasibility study on the expansion of the Prüm system of face images. Prüm is an EU-wide system that connects DNA, fingerprint, and vehicle registration databases to allow mutual searching. The concern is that a pan-European, database of faces could be used for pervasive, unjustified, or illegal surveillance.

/ Border management systems without borders

As reported in the previous edition of Automating Society, the implementation of an overarching interoperable smart border management system in the EU, initially proposed by the Commission back in 2013, is on its way. Although the new systems that have been announced (EES, ETIAS[18], ECRIS-TCN[19]) will only start operating in 2022, the Entry/Exit System (EES) regulation has already introduced face images as biometric identifiers and the use of face recognition technology for verification purposes for the first time in EU law[20].

The European Fundamental Rights Agency (FRA) confirmed the changes: "the processing of facial images is expected to be introduced more systematically in large-scale EU-level IT systems used for asylum, migration and security purposes [...] once the necessary legal and technical steps are completed".

According to Ana Maria Ruginis Andrei, from the European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice (eu-LISA), this expanded new interoperability architecture was "assembled in order to forge the perfect engine to successfully fight against the threats to internal security, to effectively control migration and to overcome blind spots regarding identity management". In practice, this means to "hold the fingerprints, facial images, and other personal data of up to 300 million non-EU nationals, merging data from five separate systems." (Campbell 2020)

/ ETIAS: automated border security screenings

The European Travel Information and Authorization System (ETIAS), which is still not in operation at the time of writing, will use different databases to automate the digital security screening of non-EU travelers (who do not need a visa, or "visa-waiver") before they arrive in Europe.

This system is going to gather and analyze data for the advanced "verification of potential security or irregular migration risks" (ETIAS 2020). The aim is to "facilitate border checks; avoid bureaucracy and delays for travelers when presenting themselves at the borders; ensure a coordinated and harmonised risk assessment of third-country nationals" (ETIAS 2020).

Ann-Charlotte Nygård, head of FRA's Technical Assistance and Capacity Building unit, sees two specific risks concerning ETIAS: "first, the use of data that could lead to the unintentional discrimination of certain groups, for instance if an applicant is from a particular ethnic group with a high in-migration risk; the second relates to a security risk assessed on the basis of past convictions in the country of origin. Some such earlier convictions could be considered unreasonable by Europeans, such as LGBT convictions in certain countries. To avoid this, [...] algorithms need to be audited to ensure that they do not discriminate and this kind of auditing would involve experts from interdisciplinary areas" (Nygård 2019).

/ iBorderCtrl: face recognition and risk scoring at the borders

iBorderCtrl was a project that involved security agencies from Hungary, Latvia, and Greece that aimed to "enable faster and thorough border control for third country nationals crossing the land borders of EU Member States". iBorderCtrl used face recognition technology, a lie detector, and a scoring system to prompt a human border policeman if it deemed someone dangerous or if it deemed their right to entry was questionable.

18 ETIAS (EU Travel Information and Authorisation System), is the new "visa waiver" system for EU border management developed by eu-LISA. "The information submitted during the application will be automatically processed against existing EU databases (Eurodac, SIS and VIS), future systems EES and ECRIS-TCN, and relevant Interpol databases. This will enable advance verification of potential security, irregular migration and public health risks" (ETIAS 2019).

19 The European Criminal Records Information System – Third Country Nationals (ECRIS-TCN), to be developed by eu-LISA, will be a centralized hit/no-hit system to supplement the existing EU criminal records database (ECRIS) on non-EU nationals convicted in the European Union.

20 EES will enter into operation in the first quarter of 2022, ETIAS will follow by the end of 2022 – set out to be "game changers in the European Justice and Home Affairs (JHA) area".

The iBorderCtrl project came to an end in August 2019, and the results – for any potential EU-wide implementation of the system – were contradictory.

Although it "will have to be defined, how far the system or parts of it will be used", the project's "Outcomes" page sees "the possibility of integrating the similar functionalities of the new ETIAS system and extending the capabilities of taking the border crossing procedure to where the travellers are (bus, car, train, etc.)".

However, the modules this refers to were not specified, and the ADM-related tools that were tested have not been publicly evaluated.

At the same time, the project's FAQ page confirmed that the system that was tested is not considered to be "currently suitable for deployment at the border (...) due to its nature as a prototype and the technological infrastructure on EU level". This means that "further development and an integration in the existing EU systems would be required for a use by border authorities."

In particular, while the iBorderCtrl Consortium was able to show, in principle, the functionality of such technology for border checks, it is also clear that ethical, legal, and societal constraints need to be addressed prior to any actual deployment.

/ Related Horizon2020 projects

Several follow-up projects focused on testing and developing new systems and technologies for Border Management and Surveillance, under the Horizon2020 program. They are listed on the European Commission's CORDIS website, which provides information on all EU-supported research activities related to it.

The site shows that 38 projects are currently running under the "H2020-EU.3.7.3. – Strengthen security through border management" program/topic of the European Union. Its parent program – "Secure societies – Protecting freedom and security of Europe and its citizens", boasts an overall budget of almost 1.7 billion euros and funds 350 projects – claims to tackle "insecurity, whether from crime, violence, terrorism, natural or man-made disasters, cyber attacks or privacy abuses, and other forms of social and economic disorders increasingly affect[ing] citizens" through projects mainly developing new technological systems based on AI and ADM.

Some projects have already finished and/or their applications are already in use – for example, FastPass, ABC4EU, MOBILEPASS, and EFFISEC – all of which looked into requirements for "integrated, interoperable Automated Border Control (ABC)", identification systems, and "smart" gates at different border crossings.

TRESSPASS is an ongoing project that started in June 2018 and will finish in November 2021. The EU contributes almost eight million euros to the project, and the coordinators of iBorderCRL (as well as FLYSEC and XP-DITE) are aiming to "leverage the results and concepts implemented and tested" by iBorderCRL and "expand[ing] them into a multimodal border crossing risk-based security solution within a strong legal and ethics framework." (Horizon2020 2019)

The project has the stated goal of turning security checks at border crossing points from the old and outdated "Rule-based" to a new "Risk-based" strategy. This includes applying biometric and sensing technologies, a risk-based management system, and relevant models to assess identity, possessions, capability, and intent. It aims to enable checks through "links to legacy systems and external databases such as VIS/SIS/PNR" and is collecting data from all the above data sources for security purposes.

Another pilot project, FOLDOUT, started in September 2018 and will finish in February 2022. The EU contributes €8,199,387,75 to the project to develop "improved methods for border surveillance" to counter irregular migration with a focus on "detecting people through dense foliage in extreme climates" [...] by combining "various sensors and technologies and intelligently fuse[ing] these into an effective and robust intelligent detection platform" to suggest reaction scenarios. Pilots are underway in Bulgaria, with demonstration models in Greece, Finland, and French Guiana.

MIRROR, or Migration-Related Risks caused by misconceptions of Opportunities and Requirement, started in June 2019 and will finish in May 2022. The EU contributes just over five million euros to the project, which aims to "understand how Europe is perceived abroad, detect discrepancies between image and reality, spot instances of media manipulation, and develop their abilities for counteracting such misconceptions and the security threats resulting from them". Based on "perception-specific threat analysis, the MIRROR project will combine methods of automated text, multimedia and social network analysis for various types of media (including social media) with empirical studies" to develop "technology and actionable insights, [...]

thoroughly validated with border agencies and policy makers, e.g. via pilots".

Other projects that are already closed, but that get a mention, include Trusted Biometrics under Spoofing Attacks (TABULA RASA), which started in November 2010 and finished in April 2014. It analyzed "the weaknesses of biometric identification process software in scope of its vulnerability to spoofing, diminishing efficiency of biometric devices". Another project, Bodega, which started in June 2015 and finished in October 2018, looked into how to use "human factor expertise" when it comes to the "introduction of smarter border control systems like automated gates and self-service systems based on biometrics".

THE EU CONTRIBUTES 8,199,387,75 EUROS TO THE PROJECT TO DEVELOP "IMPROVED METHODS FOR BORDER SURVEILLANCE" TO COUNTER IRREGULAR MIGRATION.

References:

Access Now (2019): Comments on the draft recommendation of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems https://www.accessnow.org/cms/assets/uploads/2019/10/Submission-on-CoE-recommendation-on-the-human-rights-impacts-of-algorithmic-systems-21.pdf

AlgorithmWatch (2020): Our response to the European Commission's consultation on AI https://algorithmwatch.org/en/response-european-commission-ai-consultation/

Campbell, Zach/Jones, Chris (2020): Leaked Reports Show EU Police Are Planning a Pan-European Network of Facial Recognition Databases https://theintercept.com/2020/02/21/eu-facial-recognition-database/

CNIL (2019): French privacy regulator finds facial recognition gates in schools illegal https://www.biometricupdate.com/201910/french-privacy-regulator-finds-facial-recognition-gates-in-schools-illegal

Coeckelbergh, Mark/Metzinger, Thomas (2020): Europe needs more guts when it comes to AI ethics https://background.tagesspiegel.de/digitalisierung/europe-needs-more-guts-when-it-comes-to-ai-ethics

Committee of Ministers (2020): Recommendation CM/Rec (2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154

Commissioner for Human Rights (2020): Unboxing artificial intelligence: 10 steps to protect human rights https://www.coe.int/en/web/commissioner/-/unboxing-artificial-intelligence-10-steps-to-protect-human-rights

Committee on Legal Affairs (2020): Draft Report: With recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies https://www.europarl.europa.eu/doceo/document/JURI-PR-650508_EN.pdf

Committee on Legal Affairs (2020): Artificial Intelligence and Civil Liability https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf

Committee on Legal Affairs (2020): Draft Report: On intellectual property rights for the development of artificial intelligence technologies https://www.europarl.europa.eu/doceo/document/JURI-PR-650527_EN.pdf

Committee on Civil Liberties, Justice and Home Affairs (2020): Draft Report: On artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters https://www.europarl.europa.eu/doceo/document/LIBE-PR-652625_EN.pdf

Delcker, Janosch(2020): Decoded: Drawing the battle lines – Ghost work – Parliament's moment https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-drawing-the-battle-lines-ghost-work-parliaments-moment/?utm_source=POLITICO.EU&utm_campaign=5a7d137f82-EMAIL_CAMPAIGN_2020_09_09_08_59&utm_medium=email&utm_term=0_10959edeb5-5a7d137f82-190607820

Data Protection Commission(2020): Law enforcement directive https://www.dataprotection.ie/en/organisations/law-enforcement-directive

EDRi (2019): FRA and EDPS: Terrorist Content Regulation requires improvement for fundamental rights https://edri.org/our-work/fra-edps-terrorist-content-regulation-fundamental-rights-terreg/

GDPR (Art 22): Automated individual decision-making, including profiling https://gdpr-info.eu/art-22-gdpr/

European Commission (2018): White paper: On Artificial Intelligence – A European approach to excellence and trust https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission (2020): A European data strategy https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

European Commission (2020): Shaping Europe's digital future – Questions and Answers https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_264

European Commission (2020): White Paper on Artificial Intelligence: Public consultation towards a European approach for excellence and trust https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence

European Commission (2018): Security Union: A European Travel Information and Authorisation System – Questions & Answers https://ec.europa.eu/commission/presscorner/detail/en/MEMO_18_4362

European Data Protection Board (2019): Facial recognition in school renders Sweden's first GDPR fine https://edpb.europa.eu/news/national-news/2019/facial-recognition-school-renders-swedens-first-gdpr-fine_en

European Parliament (2020): Artificial intelligence: EU must ensure a fair and safe use for consumers https://www.europarl.europa.eu/news/en/press-room/20200120IPR70622/artificial-intelligence-eu-must-ensure-a-fair-and-safe-use-for-consumers

European Parliament (2020): On automated decision-making processes: ensuring consumer protection and free movement of goods and services https://www.europarl.europa.eu/doceo/document/B-9-2020-0094_EN.pdf

European Data Protection Supervisor (2019): Facial recognition: A solution in search of a problem? https://edps.europa.eu/press-publications/press-news/blog/facial-recognition-solution-search-problem_de

ETIAS (2020): European Travel Information and Authorisation System (ETIAS) https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders/etias_en

ETIAS (2019): European Travel Information and Authorisation System (ETIAS) https://www.eulisa.europa.eu/Publications/Information%20Material/Leaflet%20ETIAS.pdf

Horizon2020 (2019): robusT Risk basEd Screening and alert System for PASSengers and luggage https://cordis.europa.eu/project/id/787120/reporting

High Court of Justice (2019): Bridges v. the South Wales Police https://www.judiciary.uk/wp-content/uploads/2019/09/bridges-swp-judgment-Final03-09-19-1.pdf

High-Level Expert Group on Artificial Intelligence (2020): Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Hunton Andrew Kurth (2020): UK Court of Appeal Finds Automated Facial Recognition Technology Unlawful in Bridges v South Wales Police https://www.huntonprivacyblog.com/2020/08/12/uk-court-of-appeal-finds-automated-facial-recognition-technology-unlawful-in-bridges-v-south-wales-police/

Kayalki, Laura (2019): French privacy watchdog says facial recognition trial in high schools is illegal https://www.politico.eu/article/french-privacy-watchdog-says-facial-recognition-trial-in-high-schools-is-illegal-privacy/

Kayser-Bril, Nicolas (2020): EU Commission publishes white paper on AI regulation 20 days before schedule, forgets regulation https://algorithmwatch.org/en/story/ai-white-paper/

Leyen, Ursula von der (2019): A Union that strives for more – My agenda for Europe https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf

Leyen, Ursula von der (2020): Paving the road to a technologically sovereign Europe https://delano.lu/d/detail/news/paving-road-technologically-sovereign-europe/209497

Leyen, Ursula von der (2020): Shaping Europe's digital future https://twitter.com/eu_commission/status/1230216379002970112?s=11

Leyen, Ursula von der (2019): Opening Statement in the European Parliament Plenary Session by Ursula von der Leyen, Candidate for President of the European Commission https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_19_4230

Ann-Charlotte Nygård, (2019): The New Information Architecture as a Driver for Efficiency and Effectiveness in Internal Security https://www.eulisa.europa.eu/Publications/Reports/eu-LISA%20Annual%20Conference%20Report%202019.pdf

Sabbagh, Dan (2020): South Wales police lose landmark facial recognition case https://www.theguardian.com/technology/2020/aug/11/south-wales-police-lose-landmark-facial-recognition-case

South Wales Police (2020): Automated Facial Recognition https://afr.south-wales.police.uk/

Valero, Jorge (2020): Vestager: Facial recognition tech breaches EU data protection rules https://www.euractiv.com/section/digital/news/vestager-facial-recognition-tech-breaches-eu-data-protection-rules/

SWITZERLAND

STORY

PAGE 197

RESEARCH

PAGE 200

Find out more in the research chapter under “Cancer diagnoses and treatments”.

Swiss police automated crime predictions but have little to show for it

A review of 3 automated systems in use by the Swiss police and judiciary reveals serious issues. Real-world effects are impossible to assess due to a lack of transparency.

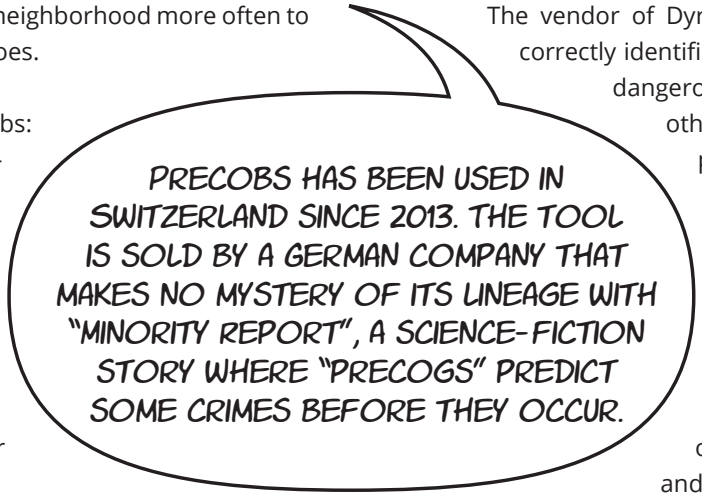
By Nicolas Kayser-Bril

/ Predicting burglaries

Precobs has been used in Switzerland since 2013. The tool is sold by a German company that makes no mystery of its lineage with "Minority Report", a science-fiction story where "precogs" predict some crimes before they occur. (The plot revolves around the frequent failures of the precogs and the subsequent cover-up by the police).

The system tries to predict burglaries from past data, based on the assumption that burglars often operate in small areas. If a cluster of burglaries is detected in a neighborhood, the police should patrol that neighborhood more often to put an end to it, the theory goes.

Three cantons use Precobs: Zürich, Aargau, and Basel-Land, totaling almost a third of the Swiss population. Burglaries have fallen dramatically since the mid-2010s. The Aargau police even complained in April 2020 that there were now too few burglaries for Precobs to use.



But burglaries fell in every Swiss canton, and the three that use Precobs are nowhere near the best performers. Between 2012-2014 (when burglaries were at their peak) and between 2017-2019 (when Precobs was in use in the three cantons), the number of burglaries decreased in all cantons, not just in the three that used the software. The decrease in Zürich and Aargau was less than the national average of -44%, making it unlikely that Precobs had much of an actual effect on burglaries.

A 2019 report by the University of Hamburg, could not find any evidence of the efficacy of predictive policing solutions, including Precobs. No public documents detail how much Swiss authorities have spent on the system, but Munich paid 100,000 euros to install Precobs (operating costs not included).

/ Predicting violence against women

Six cantons (Glarus, Luzern, Schaffhausen, Solothurn, Thurgau, and Zürich) use the Dyrias-Intimpartner system to predict the likelihood that a person will assault their intimate partner. Dyrias stands for "dynamic system for the analysis of risk" and is also built and sold by a German company.

According to a 2018 report by Swiss public-service broadcaster SRF, Dyrias requires police officers to answer 39 "yes" or "no" questions about a suspect. The tool then outputs a score on a scale from one to five, from harmless to dangerous. While the total number of persons tested by the tool is unknown, a tally by SRF showed that 3,000 individuals were labeled "dangerous" in 2018 (but the label might not be derived from using Dyrias).

The vendor of Dyrias claims that the software correctly identifies eight out of ten potentially dangerous individuals. However, another study looked at the false positives, individuals labeled dangerous who were in fact harmless, and found that six out of ten people flagged by the software should have been labeled harmless. In other words, Dyrias boasts good results only because it takes no risks and assigns the "dangerous" label liberally. (The company behind Dyrias disputes the results).

Even if the performance of the system was improved, its effects would still be impossible to assess. Justyna Gospodinov, the co-director of BIF-Frauenberatung, an organization that supports victims of domestic violence, told AlgorithmWatch that, while cooperation with the police was improving and that the systematic assessment of risk was a good thing, she could not say anything about Dyrias. "When we take in a new case, we do not know whether the software was used or not," she said.

/ Predicting recidivism

Since 2018, all justice authorities in German-speaking cantons use ROS (an acronym for "Risikoorientierter Sanktionenvollzug" or risk-oriented execution of prison sentences). The tool labels prisoners 'A' when they have no risk of recidivism, 'B' when they could commit a new offense, or 'C' when they could commit a violent crime. Prisoners can be tested several times, but subsequent tests will only allow them to move from category A to B or C and not the other way around.

A report by SRF revealed that only a quarter of the prisoners in category C committed further crimes upon being released (a false-positive rate of 75%) and that only one in five of those who committed further crimes were in category C (a false-negative rate of 80%), based on a 2013 study by the University of Zürich. A new version of the tool was released in 2017 but has yet to be reviewed.

The French and Italian-speaking cantons are working on an alternative to ROS, which should be deployed in 2022. While it keeps the same three categories, their tool will only work in conjunction with prisoner interviews that will be rated.

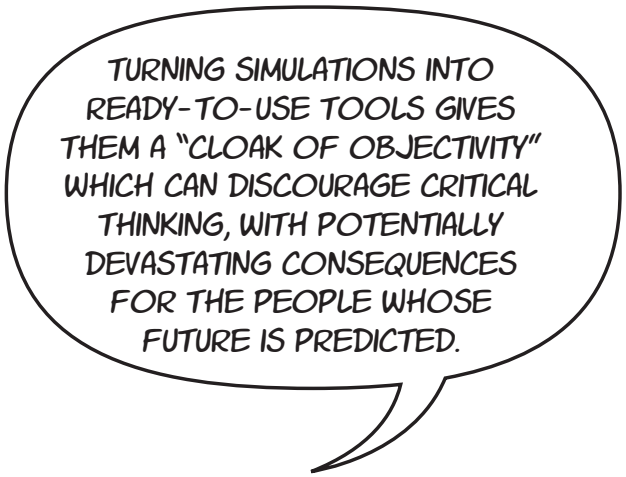
/ Mission: Impossible

Social scientists are sometimes very successful when predicting general outcomes. In 2010, the Swiss statistics office predicted that the resident population of Switzerland would reach 8.5 million by 2020 (the actual 2020 population is 8.6 million). But no scientist would try to predict the date a given individual will die: Life is simply too complicated.

In this regard, demography is no different from criminology. Despite claims to the contrary by commercial vendors, predicting individual behavior is likely to be impossible. In 2017, a group of scientists tried to settle the issue. They asked 160 teams of researchers to predict school performance, the likelihood of being evicted from home, and four other outcomes for thousands of teenagers, all based on precise data collected since birth. Thousands of data points were available for each child. The results, published in April 2020, are humbling. Not only could not a single team predict an outcome with any accuracy, but the ones who used artificial intelligence performed no better than teams who used only a few variables with basic statistical models.

Moritz Büchi, a senior researcher at the University of Zürich, is the only Swiss scholar who took part in this experiment. In an email to AlgorithmWatch, he wrote that while crime was not part of the outcomes under scrutiny, the insights gained from the experiment probably apply to predictions of criminality. This does not mean that predictions should not be attempted, Mr. Büchi wrote. But turning simulations into ready-to-use tools gives them a "cloak of objectivity" which can discourage critical thinking, with potentially devastating consequences for the people whose future is predicted.

Precobs, which does not attempt to predict the behavior of specific individuals, does not fall into the same category, he added. More policing could have a deterrent effect on criminals. However, the detection of hotspots relies on historical data. This might lead to the over-policing of communities where crime was reported in the past, in a self-reinforcing feedback loop.



/ Chilling effects

Despite their patchy track record and evidence of the near-impossibility to predict individual outcomes, Swiss law enforcement authorities keep using tools that claim to do just that. Their popularity is due in part to their opacity. Very little public information exists on Precobs, Dyrias, and ROS. The people impacted, who are overwhelmingly poor, rarely have the financial resources needed to question automated systems, as their lawyers usually focus on verifying the basic facts alleged by the prosecution.

Timo Grossenbacher, the journalist who investigated ROS and Dyrias for SRF in 2018, told AlgorithmWatch that finding people affected by these systems was "almost impossible". Not for lack of cases: ROS alone is used on thousands of inmates each year. Instead, their opacity prevents watchdogs from shedding light on algorithmic policing.

Without more transparency, these systems could have a "chilling effect" on Swiss society, according to Mr. Büchi of the University of Zürich. "These systems could deter people from exercising their rights and could lead them to modify their behavior," he wrote. "This is a form of anticipatory obedience. Being aware of the possibility of getting (unjustly) caught by these algorithms, people may tend to increase conformity with perceived societal norms. Self-expression and alternative lifestyles could be suppressed."

Research

By Nadja Braun Binder and Catherine Egli

Contextualization

Switzerland is a distinctly federalist country with a pronounced division of powers. Therefore, technical innovations in the public sector are often first developed in the cantons.

One example of this is the introduction of an electronic identity (eID). At the federal level, the legislative process required to introduce the eID has not yet been completed, whereas an officially confirmed electronic identity is already in operation in one canton. In 2017, and as part of Switzerland's cantonal eGovernment strategy, Schaffhausen canton became the first to introduce a digital identity for residents. Using this eID, citizens can apply online for a fishing permit, calculate tax liabilities of a real estate profit or a capital statement, or request a tax declaration extension, among other things.

Also, an adult account can be set up at the Child and Adult Protection Authority, and doctors can apply for credit for patients who are hospitalized outside their district. Another example, that began as part of a pilot project in the same canton in September 2019, enables residents to order extracts (via smartphone) from the debt collection register. These services are in a constant state of expansion (Schaffhauser 2020). Although the eID itself is not an ADM process, it is an essential prerequisite for access to digital government services, and therefore, could also facilitate access to automated procedures, e.g., in the tax field. The fact that a single canton has progressed further down this path than the Federal Government is typical for Switzerland.

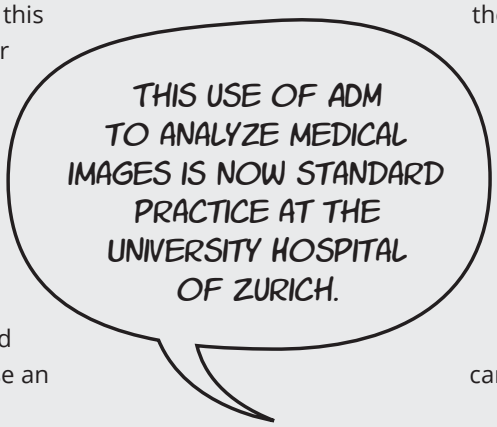
Direct democracy is another defining element of the Swiss state. For example, the legislative process for a national eID has not yet been completed because a referendum is going to be held on the corresponding parliamentary bill (eID – Referendum 2020). Those who have asked for the referendum do not fundamentally oppose an official eID, but they want to prevent private companies from issuing the eID and managing sensitive private data.

Another element that must be taken into account is the good economic situation in Switzerland. This allows great progress to be made in individual areas, such as automated decisions used in medicine, and in many areas of research. Although there is no central AI or ADM strategy in Switzerland, due to the distinct federal structure and the departmental division of responsibilities at the federal level, sectoral research is conducted at a globally competitive level.

A catalog of ADM cases

/ Cancer diagnoses and treatments

At the moment, Switzerland is investigating the use of automated decisions in medicine, which is why ADM has been developed further in the healthcare sector than in other domains. Today, more than 200 different types of cancer are known and almost 120 drugs are available to treat them. Every year, numerous cancer diagnoses are made, and, as each tumor has its own particular profile with gene mutations that help the tumor grow, this creates problems for oncologists. However, once they have made the diagnosis and determined the possible gene mutation, they have to study the ever-growing medical literature in order to select the most effective treatment.



This is why the Geneva University Hospitals are the first hospitals in Europe to use the IBM Watson Health tool, Watson for Genomics®, to better find therapeutic options and suggest treatment for cancer patients. The doctors

still examine the gene mutations and describe where and how many of them occur, but Watson for Genomics® can use this information to search a database of about three million publications. The program then creates a report classifying genetic alterations in the patient's tumor and providing associated relevant therapies and clinical trials.

Until now, cancer doctors have had to do this work themselves – with the risk that they might overlook a possible treatment method. Now the computer program can take over the research, but oncologists still have to carefully check the literature list that the program produces, and then they can decide on a treatment method. As a result, Watson for Genomics® saves a great deal of time in analysis and provides important additional information. In Geneva, the report from this ADM-tool is used during the preparation of the Tumor Board, where physicians take note of the treatments proposed by Watson for Genomics® and discuss them in plenary to jointly develop a treatment strategy for each patient (Schwerzmann/Arroyo 2019).

ADM is also in use at the University Hospital of Zurich, as it is particularly suitable for repetitive tasks, predominantly in radiology, and pathology, and, therefore, it is used to calculate breast density. During mammography, a computer algorithm automatically analyzes X-ray images and classifies the breast tissue into A, B, C or D (an internationally recognized grid for risk analysis). By analyzing risk based on breast density, the algorithm greatly assists physicians in assessing a patient's breast cancer risk, since breast density is one of the most important risk factors in breast cancer. This use of ADM to analyze medical images is now standard practice at the University Hospital of Zurich. Furthermore, research into advanced algorithms for the interpretation of ultrasound images is ongoing (Lindner 2019).

Having said this, more than one-third of breast cancers are overlooked in mammography screening examinations, which is why research is being carried out on how ADM can support the interpretation of ultrasound (US) images. The interpretation of US breast images contrasts sharply with standard digital mammography – which is largely observer-dependent and requires well-trained and experienced radiologists. For this reason, a spin-off of the University Hospital of Zurich has researched how ADM may support and standardize US imaging. In doing so, the human decision process was simulated according to the breast imaging, reporting, and data system. This technique is highly accurate, and, in the future, this algorithm may be used to mimic human decision-making, and become the standard for the detection, highlighting, and classification of US breast lesions (Ciritisis a.o. 2019 p. 5458–5468).

/ Chatbot at the Social Insurance Institution

To simplify, and support administrative communication, certain cantons also use so-called chatbots. In particular, a chatbot that was tested in 2018 at the "Sozialversicherungsanstalt des Kantons St. Gallens" (Social Insurance Institution in the canton of St. Gallen, SVA St. Gallen). The SVA St. Gallen is a center of excellence for all kinds of social insurance, including a premium reduction for health insurance. Health insurance is compulsory in Switzerland and covers every resident in the event of illness, maternity, and accidents, and offers everyone the same range of benefits. It is funded by the contributions (premiums) of citizens. The premiums vary according to the insurer, and depend on an individual's place of residence, type of insurance needed, and it is not related to income level. However, thanks to subsidies from the cantons (premium reduction), citizens on a low income, children, and young adults in full-time education or training, often pay reduced premiums. The cantons decide who is entitled to a reduction (FOPH 2020).

Towards the end of each year, the SVA St. Gallen receives approximately 80,000 applications for premium reductions. To reduce the workload associated with this concentrated flood of requests, they tested a chatbot via Facebook Messenger. The object of this pilot project was to offer customers an alternative method of communication. The first digital administrative assistant was designed to provide customers with automatic answers to the most important questions regarding premium reductions. For example: what are premium reductions and how can a claim be made? Can I claim a premium reduction? Are there any special cases, and how should I proceed? How is premium reduction calculated and paid out? Also, if it was indicated, the chatbot could refer customers to other services offered by the SVA St. Gallen, including the premium reduction calculator and the interactive registration form. While the chatbot does not make the final decision to grant premium reductions, it can still reduce the number of requests as it can inform unauthorized citizens of the likely failure of their request. It also performs a significant role in disseminating information (Ringeisen/Bertolosi-Lehr/Demaj 2018 S.51-65).

Due to the positive feedback from this first test run, the chatbot was integrated into the SVA St. Gallen's website in

2019 and there is a plan to gradually expand the chatbot to include other insurance products covered by the SVA St. Gallen. It is possible that the chatbot will also be used for services related to contributions to Old-Age and Survivor's Insurance, Disability Insurance, and Income Compensation Insurance (IPV-Chatbot 2020).

/ Penal System

The Swiss Execution of Penal Sentences and Justice is based on a system of levels. According to this system, inmates are generally granted increasing amounts of freedom as the duration of their imprisonment continues. This makes it a collaborative process between executive authorities, penal institutions, therapy providers, and probation services. Of course, the risk of escape and recidivism are decisive factors when it comes to granting these greater freedoms.

In recent years, and in response to convicted felons committing several tragic acts of violence and sex offenses, the ROS (Risk-Oriented Sanctioning) was introduced. The primary objective of ROS is to prevent recidivism by harmonizing the execution of sentences and measures across the various levels of enforcement with a consistent focus on recidivism prevention and reintegration. ROS divided the work with offenders into four stages: triage, assessment, planning, and progress. During triage, cases are classified according to their need for recidivism risk assessment. Based on this classification, a differentiated individual case analysis is carried out during the assessment stage. During the planning stage, these results are developed into an individual execution plan for the sanction of the corresponding offender, which is continuously reviewed during the progress stage (ROSNET 2020).

Triage plays a decisive role at the beginning of this process – both for the offender and in terms of ADM, as it is performed by an ADM-tool called the Fall-Screening-Tool (Case Screening Tool, FaST). FaST automatically divides all cases into classes A, B, and C. Class A signifies that there is no need for assessment, class B equates to a general risk of further delinquency, and class C corresponds to the risk of violent or sexual delinquency.

This classification is determined by using criminal records and is based on general statistical risk factors including age, violent offenses committed before the age of 18, youth attorney entries, number of previous convictions, offense category, sentences, polymorphic delinquency, offense-free time after release, and domestic violence. If risk factors

ROS DIVIDED THE WORK WITH OFFENDERS INTO FOUR STAGES: TRIAGE, ASSESSMENT, PLANNING, AND PROGRESS.

are met that, according to scientific findings, have a specific connection with violent or sexual offenses, then a C classification is made. If risk factors are met that have a specific connection with general delinquency, then a B classification is applied. If no risk factors, or hardly any, are found, then an A classification is made. Therefore, the classification consists of items (risk factors) in a closed answer format, each of which has a different weighting (points). If a risk factor is given, the relevant points are included in the total value. For the overall result, the weighted and confirmed items are added to a value, leading to the A, B or C classification that, in turn, serves as a basis to decide if further assessment is necessary (stage 2).

This classification is carried out fully automatically by the ADM-application. However, it is important to note that this is not a risk analysis, but a way of screening out the cases with increased assessment needs (Treuhardt/Kröger 2018 p. 24-32).

Nevertheless, the triage classification has an effect on how those responsible at a particular institution make decisions and which assessments are to be made. This also determines the offender's so-called 'problem profile' regarding how the execution of sentences and measures are planned (stage 3). In particular, this planning defines any possible facilitation of enforcement, such as open enforcement, day release employment, or external accommodation. Furthermore, no ADM applications are apparent in the other stages of ROS. FaST is, therefore, only used during the triage stage.

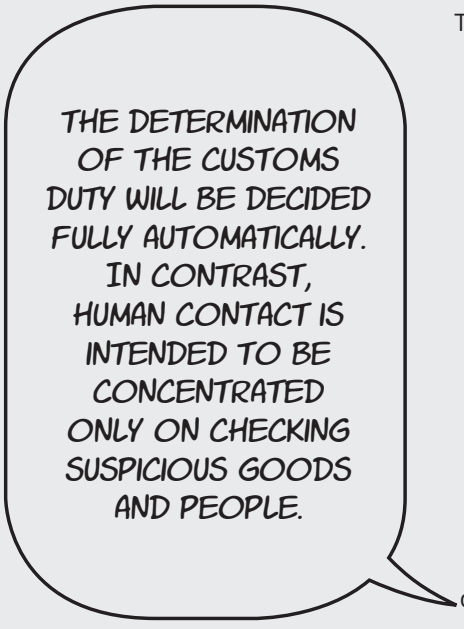
/ Predictive Policing

In some cantons, in particular in Basel-Landschaft, Aargau, and Zurich, the police use software to help prevent criminal offenses. They rely on the commercial software package "PRECOBS" (Pre-Crime Observation System), which is solely used for the prognosis of domestic burglaries. This relatively common crime has been well researched scientifically, and police authorities usually have a solid database regarding the spatial and temporal distribution of burglaries as well as crime characteristics. Furthermore, these offenses indicate a professional perpetrator and thus show an above-average probability of subsequent offenses. In addition, corresponding prognosis models can be created using relatively few data points. PRECOBS is, therefore, based on the assumption that burglars strike several times within a short period if they have already been successful in a certain location.

The software is used to search for certain patterns in the police reports on burglaries, such as how the perpetrators proceed and when and where they strike. Subsequently, PRECOBS creates a forecast for areas where there is an increased risk of burglary in the next 72 hours. Thereupon, the police send targeted patrols to the area. PRECOBS thus generates forecasts on the basis of primarily entered decisions and it does not use machine learning methods. Although there are plans to extend PRECOBS in the future to include other offenses (such as car theft or pickpocketing) and consequently create new functionalities, it should be noted that the use of predictive policing in Switzerland is currently limited to a relatively small and clearly defined area of preventive police work (Blur 2017, Leese 2018 p. 57-72).

/ Customs clearance

At the federal level, ADM is expected to be particularly present at the Federal Customs Administration (FCA), since this department is already highly automated. Accordingly, the assessment of customs declarations is largely determined electronically. The assessment procedure can be divided into four steps: summary examination procedure, acceptance of a customs declaration, verification, and inspection, followed by an assessment decision.

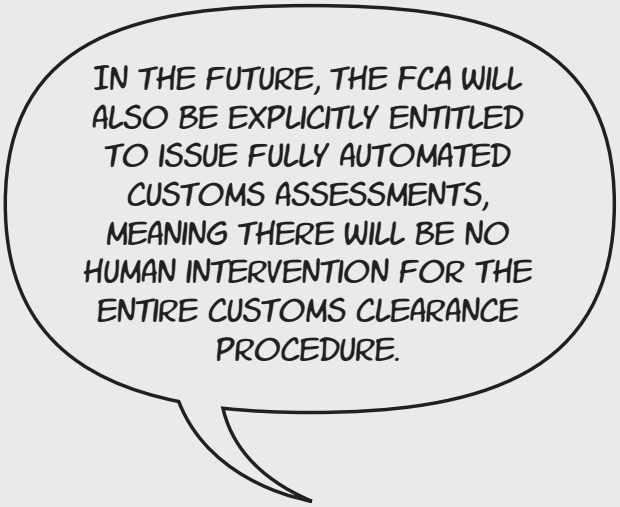


The summary examination procedure represents a plausibility check and is carried out directly by the system used in the case of electronic customs declarations. Once the electronic plausibility check has been completed, the data processing system automatically adds the acceptance date and time to the electronic customs declaration, meaning the customs declaration has been accepted. Up to this point, the procedure runs without any human intervention by the authorities.

However, the customs office may subsequently carry out a full or random inspection and verification of the

declared goods. To this end, the computerized system carries out a selection based on a risk analysis. The final stage of the procedure is when the assessment decision is issued. It is not known whether or not this assessment decision can already be issued without any human intervention. However, the DaziT program will clarify this uncertainty.

The DaziT program is a federal measure to digitize all customs processes by 2026 to simplify and accelerate border crossings. Border authorities' customer relations relating to the movement of goods and people will be fundamentally redesigned. Customers who behave correctly should be able to complete their formalities digitally, independent of time and place. While the exact implementation of the DaziT program is still at the planning stage, the revision of the Customs Act that correlates to DaziT is included in the revision of the Federal Act on Data Protection.



This is explained in more detail below, and should serve to clarify the previously mentioned uncertainty regarding the automated customs assessment procedure: In the future, the FCA will also be explicitly entitled to issue fully automated customs assessments, meaning there will be no human intervention for the entire customs clearance procedure. Thus, the determination of the customs duty will be decided fully automatically. In contrast, human contact is intended to be concentrated only on checking suspicious goods and people (EZV 2020).

/ Accident and Military Insurance

Throughout the revision of the Data Protection Act (explained in more detail below), it was decided that the accident and military insurance companies will be entitled to automatically process personal data. It is not clear what automated activities the insurance companies intend to use in the future. However, they may, for example, use algorithms to evaluate policyholder's medical reports. Through this fully automated system, premiums could be calculated, and benefit claim decisions made and coordinated with other social benefits. It is planned that these bodies will be authorized to issue automated decisions.

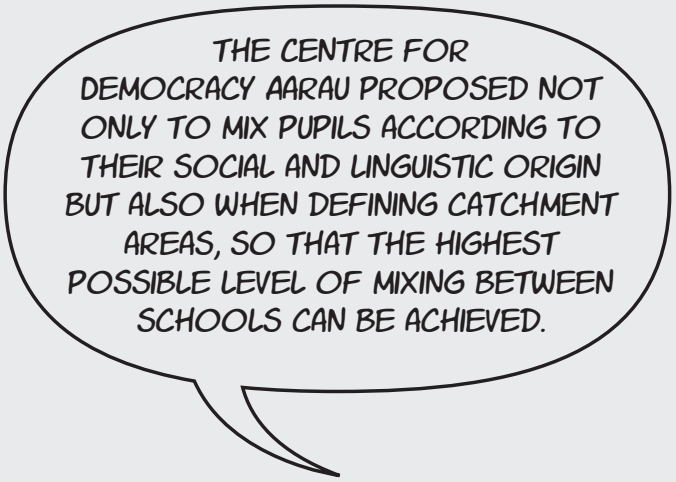
/ Automatic Vehicle Recognition

In recent years, both politicians and the public have become concerned with the use of automatic systems, such as cameras that capture vehicle license plates, read them using optical character recognition, and compare them with a database. This technology can be used for various purposes, but at the moment in Switzerland it is only used to a limited extent. At the federal level, the system for automatic vehicle detection and traffic monitoring is only used as a tactical tool depending on the situation and risk assessments, as well as economic considerations, and only at state borders (parlament.ch 2020). The half-canton of Basel-Landschaft has enacted a legal basis for the automatic recording of vehicle registration plates and alignment with corresponding databases (EJPD 2019).

/ Allocation of primary school pupils

Another algorithm that has been developed, but is not yet in use, is designed to allocate primary school pupils. International studies indicate that social and ethnic segregation between urban schools is increasing. This is problematic, as the social and ethnic composition of schools has a demonstrable effect on the performance of pupils, regardless of their background. In no other OECD country are these so-called 'compositional effects' as pronounced as in Switzerland. The different composition of schools is mainly due to segregation between residential quarters and the corresponding school catchment areas. Therefore, the Centre for Democracy Aarau proposed not only to mix pupils according to their social and linguistic origin but also when defining catchment areas, so that the highest possible level of mixing between schools can be achieved. To optimize this process, a novel, detailed algorithm was developed that could be used in the future to support school allocation and classroom planning. The algorithm was trained to reconstruct the school catchment areas and to survey the

social composition at individual schools using the census data of first to third graders in the canton of Zurich. Traffic load data, the network of pavements and footpaths, underpasses and overpasses, were also taken into account. This data could be used to calculate where pupils need to be placed to mix the classes more. At the same time, the capacity of school buildings will not be exceeded and the length of time spent traveling to school will remain reasonable (ZDA 2019).



Policy, oversight and public debate

/ Switzerland's federal structure as the prevailing circumstance

When reporting on policy in Switzerland, the prevailing federal structure must be emphasized. It is a situation that has already been reflected upon in the previously mentioned ADM examples. Switzerland is a federal state, consisting of 26 highly autonomous Member States (cantons), which in turn grant their municipalities extensive leeway. As a result, the political and public debate on ADM depends greatly on the corresponding government, which cannot be described exhaustively in this report. Furthermore, this fragmentation on policy, regulation, and in research, introduces the risk of working in parallel on overlapping issues, which is also why the confederation strives for advanced coordination as stated below. However, the Federal Government has full responsibility over certain relevant legal fields and political governance, which legally binds all the governments in Switzerland, and thus impacts the entire population. Hence, set out below are those parts of the current federal political debate.

/ Government

At the moment, the role of ADM in society, generally referred to as AI, is predominantly treated as part of a larger discussion on digitization. The Federal Government does not have a specific strategy concerning AI or ADM, but in recent years it launched a "Digital Switzerland Strategy", where all aspects regarding AI will be integrated. More generally, the national legal framework concerning digitization will be adjusted simultaneously through the revision of the Federal Act on Data Protection (FADP).

/ Digital Switzerland

In 2018, and against a background of increasing digitization beyond government services, the confederation launched the "Digital Switzerland Strategy". One focus of this is on current developments in AI (BAKOM 2020). Responsible for the strategy, especially its coordination and implementation, is the "Interdepartmental Digital Switzerland Coordination Group" (Digital Switzerland ICG) with its management unit "Digital Switzerland Business Office" (Digital Switzerland 2020).

As part of the Digital Switzerland Strategy, the Federal Council set up a working group on the subject of AI and commissioned it to submit a report to the Federal Council on the challenges associated with AI. The report was acknowledged by the Federal Council in December 2019 (SBFI 2020). Alongside discussing the most central challenges of AI – those being traceability and systematic errors in data or algorithms – the report details a concrete need for action. It is recognized that all challenges, including this need for action, depend strongly on the subject area in question, which is why the report examined 17 subject areas in greater depth, such as AI in the fields of healthcare, administration, and justice (SBFI 2020 b).

In principle, the challenges posed by AI in Switzerland have, according to the report, already been largely recognized and addressed in various policy areas. Nevertheless, the interdepartmental report identifies a certain need for action which is why the Federal Council has decided on four measures: In the area of international law and on the use of AI in public opinion-forming and decision-making, addition-

al reports will be commissioned for in-depth clarification. Further on, ways of improving coordination, relating to the use of AI in the Federal Administration, will be examined. In particular, the creation of a competence network, with a special focus on technical aspects of the application of AI in the federal administration, will be examined. Finally, AI-relevant policy will be taken into account as an essential component of the "Digital Switzerland" strategy. In this context, the Federal Council has decided that interdepartmental work should continue and that strategic guidelines for the confederation should be developed by spring 2020 (SBFI 2020 c).

In addition, at its meeting on 13 May 2020, the Federal Council decided to create a national Data Science Competence Center. The Federal Statistical Office (FSO) will establish this interdisciplinary center on 1 January 2021. The new center will support the federal administration in implementing projects in the field of data science. To this end, the transfer of knowledge within the Federal Administration as well as the exchange with scientific circles, research institutes, and the bodies responsible for practical application will be promoted. In particular, the center of excellence will contribute to the production of transparent information while taking data protection into account. The reasoning behind the new center is backed up by a statement from the Federal Council, which said that data science is becoming increasingly important, not least in public administration. According to the Federal Council data science includes "intelligent" calculations (algorithms) so that certain complex tasks can be automated (Bundesrat 2020).

/ Oversight

Since the Federal Data Protection Act is no longer up to date, due to rapid technological developments, the Federal Council intends to adapt the FADP to these changed technological and social conditions, and in particular, improve the transparency of data processing and strengthen the self-determination of data subjects over their data. At the same time, this total revision should allow Switzerland to ratify the revised Council of Europe Data Protection Convention ETS 108 and to adopt the Directive (EU) 680/2016 on data protection in the area of criminal prosecution, which it is obliged to do due to the Schengen Agreement. In addition, the revision should bring Swiss data protection legislation as a whole closer to the requirements of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of people regarding the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (GDPR). The revision is currently being discussed in parliament (EJPD 2020).

While the total revision of the existing federal law on data protection will evidently revise the entire Act with all its various aspects, one new provision is of particular interest in regard to ADM. In the case of so-called "automated individual decisions", there should be an obligation to inform the data subject if the decision has legal consequences or significant effects. The data subject may also request that the decision is reviewed by a human or that he or she should be informed of the logic on which the decision is based. Thereby, a differentiated regulation is provided for decisions by federal bodies. Correspondingly, even though federal bodies must also mark the automated individual decision as such, the possibility that the data subject can request a review by a human may be restricted by other federal laws. In contrast to the EU's GDPR, there is neither a prohibition of an automated decision, nor is there a claim not to be subject to such a decision. (SBFI 2019)

/ Civil Society, Academia and other Organizations

In addition, a number of fora in Switzerland research, discuss, and work on digital transformation and its opportunities, challenges, needs, and ethics. Most of them address this on a general level, however some address ADM or AI specifically.

/ Research Institutes

Switzerland has a number of well-known and long-established research centers regarding AI technology. These include the Swiss AI Lab IDSIA in Lugano (SUPSI 2020) and the IDIAP Research Institute in Martigny (Idiap 2020) as well as the research centers of the Swiss Federal Institute of Technology in Lausanne (EPFL 2020), and Zurich (EPFL 2020). In addition, private initiatives such as the Swiss Group of Artificial Intelligence and Cognitive Science (SGAICO) complement the academic initiatives, by bringing together researchers and users, promoting knowledge transfer, confidence building, and interdisciplinarity (SGAICO 2020).

/ Government Research Funding

The confederation also addresses the topic of AI via research funding. For example, the Federal Government invests in two national research programs via the Swiss National Science Foundation (SNSF) (SNF 2020). Firstly, the National Re-

MACHINE LEARNING METHODS ARE NOT USED IN STATE ACTIVITIES IN THE NARROWER SENSE, AS FAR AS CAN BE SEEN.

search Programme 77 "Digital Transformation" (NRP 77) (NRP77 2020) and, secondly, the National Research Programme 75 "Big Data" (NRP 75) (NRP 75 2020). The former examines the interdependencies and concrete effects of digital transformation in Switzerland, and focuses on education and learning, ethics, trustworthiness, governance, the economy, and the labor market (NFP 77 2020). The latter aims to provide the scientific basis for an effective and appropriate use of large amounts of data. Accordingly, the research projects examine questions of the social impact of information technology and address concrete applications (SNF 2020).

Another institute working in this area is the Foundation for Technology Assessment (TA-Swiss). TA-Swiss is a center of excellence of the Swiss Academies of Arts and Sciences, whose mandate is laid down in the Federal Law on Research. It is an advisory body, financed by the public sector, and it has commissioned various studies on AI. The most pertinent of these is a study published on 15 April 2020 on the use of AI in different areas (consumption, work, education, research, media, public administration, and justice). According to the study, a separate law on the use of AI is not considered to be effective. Nevertheless, citizens, consumers, and employees in their dealings with the state, companies or their employer should be informed as transparently as possible about the use of AI. When public institutions or companies use AI they should do so according to clear rules, in an understandable and transparent manner (Christen, M. et al. 2020).

/ Digital Society Initiative

The Digital Society Initiative was launched in 2016. It is a center of excellence at the University of Zurich for critical reflection on all aspects of the digital society. Its goal is to reflect on and help shape the digitization of society, democracy, science, communication, and the economy. In addition, it aims to critically reflect and shape the current change in thinking brought about by digitization in a future-oriented manner and to position the University of Zurich nationally and internationally as a center of excellence for the critical reflection of all aspects of digital society (UZH 2020).

/ Digitale Gesellschaft

The Digitale Gesellschaft (Digital Society) is a non-profit society and broad-based association for citizen and consumer protection in the digital age. Since 2011, it has been working as a civil society organization for a sustainable, democratic and free public sphere and it aims to defend fundamental rights in a digitally networked world. (Digitale Gesellschaft 2020)

/ Other organizations

Several other Swiss organizations are also worth a mention. These organizations concentrate on digitization in general, especially in an economic context, e.g., Swiss Telecommunication Association (asut 2020), digitalswitzerland (Castle 2020), Swiss Data Alliance, and Swiss Fintech Innovations.

Key takeaways

ADM is used in various parts of the public sector in Switzerland, but these tend not to be in a centralized or comprehensive manner. Only a few cantons use ADM in police work, for example, and the systems used vary. The advantage of such an approach is that cantons or the federal government can benefit from the experience of other cantons. The drawback is that efficiency losses may occur.

There are selective legal foundations, but no uniform ADM law or e-government law or anything similar. Neither is there a specific AI or ADM strategy, but recently attention has been paid to greater coordination, both between different departments at the federal level and between the Federal Government and the cantons. Machine learning methods are not used in state activities in the narrower sense, e.g., in police work or the criminal justice system, as far as can be seen.

Also, at that same level, ADM is used or discussed selectively, but not in a comprehensive manner. In the wider public sector, ADM is used more often and more widely. A good example is a deployment in the Swiss health system, where the Geneva University Hospital became the first hospital in Europe to use ADM to suggest treatments for cancer patients.

References:

Asut (o. J.): in: asut.ch, [online] https://asut.ch/asut/de/page/index.xhtml [30.01.2020]

Bundesamt für Kommunikation BAKOM (o. J.): Digitale Schweiz, in: admin.ch, [online] https://www.bakom.admin.ch/bakom/de/home/digital-und-internet/strategie-digitale-schweiz.html [30.01.2020].

Der Bundesrat (o.J.): Der Bundesrat schafft ein Kompetenzzentrum für Datenwissenschaft, In: admin.ch, [online] https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-79101.html [15.05.2020].)

Christen, M. et al. (2020): Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz, in: TA-Swiss, [online] https://www.ta-swiss.ch/themen-projekte-publikationen/informationsgesellschaft/kuenstliche-intelligenz/ [15.05.2020].

Ciritsis, Alexander / Cristina Rossi / Matthias Eberhard / Magda Marcon / Anton S. Becker / Andreas Boss (2019): Automatic classification of ultrasound breast lesions using a deep convolutional neural network mimicking human decision-making, in: European Radiology, Jg. 29, Nr. 10, S. 5458–5468, doi: 10.1007/s00330-019-06118-7.

Digitale Gesellschaft (o. J.): Über uns, in: Digitale Gesellschaft, [online] https://www.digitale-gesellschaft.ch/uber-uns/ [30.01.2020].

digitalswitzerland (Castle, Danièle Digitalswitzerland (2019): Digitalswitzerland – Making Switzerland a Leading Digital Innovation Hub, in: digitalswitzerland, [online] https://digitalswitzerland.com [30.01.2020])

Digital Switzerland (2020): (Ofcom, Federal Office Of Communications (o. J.): Digital Switzerland Business Office, in: admin.ch, [online] https://www.bakom.admin.ch/bakom/en/homepage/ofcom/organisation/organisation-chart/information-society-business-office.html [30.01.2020].)

EPFL (o. J.): in: epfl, [online] https://www.epfl.ch/en/ [30.01.2020]

EJPD (o. J.): Stärkung des Datenschutzes, in: admin.ch, [online] https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/datenschutzstaerkung.html [30.01.2020c].

E-ID Referendum (o.J.): in: e-id-referendum.ch/, [online] https://www.e-id-referendum.ch [31.1.2020].

EJPD (o. J.): Stärkung des Datenschutzes, in: admin.ch, [online] https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/datenschutzstaerkung.html [30.01.2020c].

EJPD Eidgenössisches Justiz- und Polizeidepartement (2019): Änderung der Geschwindigkeitsmessmittel-Verordnung (SR 941.261) Automatische Erkennung von Kontrollschildern, in: admin.ch, [online] https://www.admin.ch/ch/d/gg/pc/documents/3059/Erl_Bericht_de.pdf.

EZV (2020): EZV, Eidgenössische Zollverwaltung (o. J.): Transformationsprogramm DaziT, in: admin.ch, [online] https://www.ezv.admin.ch/ezv/de/home/themen/projekte/dazit.html [30.01.2020].

ETH Zurich – Homepage (o. J.): in: ETH Zurich – Homepage | ETH Zurich, [online] https://ethz.ch/en.html [30.01.2020].

Federal office of public health FOPH (2020): (Health insurance: The Essentials in Brief (o. J.): in: admin.ch, [online] https://www.bag.admin.ch/bag/de/home/versicherungen/krankenversicherung/krankenversicherung-das-wichtigste-in-kuerze.html [13.02.2020].)

Geschäft Ansehen (o. J.): in: parlament.ch, [online] https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20143747 [30.01.2020].

Heinhold, Florian (2019): Hoffnung für Patienten?: Künstliche Intelligenz in der Medizin, in: br.ch, [online] https://www.br.de/br-fernsehen/sendungen/gesundheit/kuenstliche-intelligenz-ki-medizin-102.html [30.01.2020].

Idiap Research Institute (o. J.): in: Idiap Research Institute, Artificial Intelligence for Society, [online] https://www.idiap.ch/en [30.01.2020]

Der IPV-Chatbot – SVA St.Gallen (o. J.): in: svasg.ch, [online] https://www.svasg.ch/news/meldungen/ipv-chatbot.php [30.01.2020].

Leese, Matthias (2018): Predictive Policing in der Schweiz: Chancen, Herausforderungen Risiken, in: Bulletin zur Schweizerischen Sicherheitspolitik, Jg. 2018, S. 57–72.

Lindner, Martin (2019): KI in der Medizin: Hilfe bei einfachen und repetitiven Aufgaben, in: Neue Zürcher Zeitung, [online] https://www.nzz.ch/wissenschaft/ki-in-der-medizin-hilfe-bei-einfachen-und-repetitiven-aufgaben-ld.1497525?reduced=true [30.01.2020

Medinside (o. J.): in: Medinside, [online] https://www.medinside.ch/de/post/in-genf-schlaegt-der-computer-die-krebsbehandlung-vor [14.02.2020].

NRP 75 Big Data (o. J.): in: SNF, [online] http://www.snf.ch/en/researchinFocus/nrp/nfp-75/Pages/default.aspx [30.01.2020].

NFP [Nr.] (o. J.): in: nfp77.ch, [online] https://www.nfp77.ch/en/ [30.01.2020]

NRP 75 Big Data (o. J.): in: SNF, [online] http://www.snf.ch/en/researchinFocus/nrp/nfp-75/Pages/default.aspx [30.01.2020].

NFP [Nr.] (o. J.): in: nfp77.ch, [online] http://www.nfp77.ch/en/Pages/Home.aspx [30.01.2020]

Ringeisen, Peter / Andrea Bertolosi-Lehr / Labinot Demaj (2018): Automatisierung und Digitalisierung in der öffentlichen Verwaltung: digitale Verwaltungsassistenten als neue Schnittstelle zwischen Bevölkerung und Gemeinwesen, in: Yearbook of Swiss Administrative Sciences, Jg. 9, Nr. 1, S. 51–65, doi: 10.5334/ssas.123.

ROSNET > ROS allgemein (o. J.): in: ROSNET, [online] https://www.rosnet.ch/de-ch/ros-allgemein [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (o. J.): Künstliche Intelligenz, in: admin.ch, [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (2019): Herausforderungen der künstlichen Intelligenz – Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, in: admin.ch, [online] https://www.sbfi.admin.ch/sbfi/de/home/bfi-politik/bfi-2021-2024/transversale-themen/digitalisierung-bfi/kuenstliche-intelligenz.html [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (o. J.): Künstliche Intelligenz, in: admin.ch, [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

SBFI, Staatssekretariat für Bildung, Forschung und Innovation (2019): Herausforderungen der künstlichen Intelligenz – Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, in: admin.ch, [online] https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung/kuenstliche-intelligenz.html [30.01.2020].

Schaffhauser eID+ – Kanton Schaffhausen (o. J.): in: sh.ch, [online] https://sh.ch/CMS/Webseite/Kanton-Schaffhausen/Beh-rde/Services/Schaffhauser-eID--2077281-DE.html [30.01.2020].

SGAICO – Swiss Group for Artificial Intelligence and Cognitive Science (2017): in: SI Hauptseite, [online] https://swissinformatics.org/de/gruppierungen/fg/sgaico/ [30.01.2020]

SNF, [online] http://www.snf.ch/en/Pages/default.aspx [30.01.2020]

Srf/Blur;Hesa (2017): Wie «Precobs» funktioniert – Die wichtigsten Fragen zur «Software gegen Einbrecher», in: Schweizer Radio und Fernsehen (SRF), [online] https://www.srf.ch/news/schweiz/wie-precobs-funktioniert-die-wichtigsten-fragen-zur-software-gegen-einbrecher

SUPSI – Dalle Molle Institute for Artificial Intelligence – Homepage (o. J.): in: idsia, [online] http://www.idsia.ch [30.01.2020].

Swissdataaliance (o. J.): in: swissdataalliance, [online] https://www.swissdataalliance.ch [30.01.2020].

Swiss Fintech Innovations (SFTI introduces Swiss API information platform (2019): in: Swiss Fintech Innovations – Future of Financial Services, [online] https://swissfintechinnovations.ch [30.01.2020]).

Schwerzmann, Jacqueline Amanda Arroyo (2019): Dr. Supercomputer – Mit künstlicher Intelligenz gegen den Krebs, in: Schweizer Radio und Fernsehen (SRF), [online] https://www.srf.ch/news/schweiz/dr-supercomputer-mit-kuenstlicher-intelligenz-gegen-den-krebs

Treuthardt, Daniel / Melanie Kröger (2019): Der Risikoorientierte Sanktionenvollzug (ROS) – empirische Überprüfung des Fall-Screening-Tools (FaST), in: Schweizerische Zeitschrift für Kriminologie, Jg. 2019, Nr. 1–2, S. 76–85.); (Treuthardt, Daniel / Melanie Kröger / Mirjam Loewe-Baur (2018): Der Risikoorientierte Sanktionenvollzug (ROS) – aktuelle Entwicklungen, in: Schweizerische Zeitschrift für Kriminologie, Jg. 2018, Nr. 2, S. 24–32.

ZDA (2019): Durchmischung in städtischen Schulen, in: zdaarau.ch, [online] https://www.zdaarau.ch/dokumente/SB-17-Durchmischung-Schulen-ZDA.pdf [30.01.2020].

Team

/ Beate Autering

Graphic designer and layout artist

Beate Autering is a freelance graphic designer. She graduated in design and runs the beworx studio. She create designs, graphics, and illustrations and also provide image editing and post-production services.

/ Nadja Braun Binder

Author of the research chapter on Switzerland

Nadja Braun Binder studied law at the University of Berne and received her doctorate there. Her academic career took her to the Research Institute of Public Administration in Speyer in 2011, where she conducted research on the automation of administrative procedures, among other things. In 2017, she was habilitated by the German University of Administrative Sciences, Speyer, and followed a call to the Faculty of Law at the University of Zurich, where she worked as an assistant professor until 2019. Since 2019 Nadja has been Professor of Public Law at the University of Basel. Her research focuses on legal issues related to digitization in government and administration. She is currently conducting a study on the use of artificial intelligence in public administration in the canton of Zurich.

/ Fabio Chiusi

Editor, author of the introduction and the chapter on Europe

Fabio Chiusi works at AlgorithmWatch as the co-editor and project manager for the 2020 edition of the Automating Society report. After a decade in tech reporting, he started as a consultant and assistant researcher in data and politics (Tactical Tech) and AI in journalism (Polis LSE). He coordinated the "Persuasori Social" report about the regulation of political campaigns on social media for the PuntoZero Project, and he worked as a tech-policy staffer within the Chamber of Deputies of the Italian Parliament during the current legislation. Fabio is a fellow at the Nexa Center for Internet & Society in Turin and an adjunct professor at the University of San Marino, where he teaches journalism and new media and publishing and digital media. He is the author of several essays on technology and society, the latest being "Io non sono qui. Visioni e inquietudini da un futuro presente" (DeA Planeta, 2018), which is currently being translated into Polish and Chinese. He also writes as a tech-policy reporter for the collective blog ValigiaBlu.

/ Samuel Daveti

Comic artist

Samuel Daveti is a founding member of the Cultural Association, Double Shot. He is the author of the French language graphic-novel, Akron Le guerrier (Soleil, 2009), and he is the curator of the anthological volume Fascia Protetta (Double Shot, 2009). In 2011, he became a founding member of the self-produced comics collective, Mammaiuto. Samuel also wrote Un Lungo Cammino (Mammaiuto, 2014; Shockdom, 2017), which will become a film for the media company Brandon Box. In 2018, he wrote The Three Dogs, with drawings by Laura Camelli, which won the Micheluzzi Prize at Napoli Comicon 2018 and the Boscarato award for the best webcomic at the Treviso Comic Book Festival..

/ Catherine Egli

Author of the research chapter on Switzerland

Catherine Egli recently graduated with a double bilingual master's in law degree from the Universities of Basel and Geneva. Her thesis focused on automated individual decision-making and the need for regulation in the Swiss Administrative Procedure Act. Alongside her studies, she worked for the chair of Prof. Dr. Nadja Braun Binder by conducting research on legal issues related to automated decision-making. Her favorite topics of research include the division of powers, digitization of public administration, and digital democracy.

/ Sarah Fischer

Editor

Sarah Fischer is a project manager for the "Ethics of Algorithms" project at the Bertelsmann Stiftung, in which she is primarily responsible for the scientific studies. She has previously worked as a postdoctoral fellow in the graduate program "Trust and Communication in a Digitalized World" at the University of Münster where she focused on the topic of trust in search engines. In the same research training group, she earned her doctorate with a thesis on trust in health services on the Internet. She studied communication science at the Friedrich Schiller University in Jena, and she is the co-author of the papers "Where Machines can err. Sources of error and responsibilities in processes of algorithmic decision making" and "What Germany knows and believes about algorithms".

/ Leonard Haas

Additional editing

Leonard Haas works as a research assistant at AlgorithmWatch. Among other things, he was responsible for the conception, implementation, and maintenance of the AI Ethics Guidelines Global Inventory. He is a master's student in the field of social sciences at the Humboldt University Berlin and holds two Bachelor's degrees from the University of Leipzig in Digital Humanities and Political Science. His research focuses on the automation of work and governance. In addition, he is interested in public interest data policy and labor struggles in the tech industry.

/ Graham Holliday

Copy editing

Graham Holliday is a freelance editor, author, and journalism trainer. He has worked in a number of roles for the BBC for almost two decades, and he was a correspondent for Reuters in Rwanda. He works as an editor for CNN's Parts Unknown and Roads & Kingdoms – the international journal of foreign correspondence. The late Anthony Bourdain published Graham's first two books, which were reviewed in the New York Times, Los Angeles Times, Wall Street Journal, Publisher's Weekly, Library Journal, and on NPR, among other outlets.

/ Nikolas Kayser-Bril

Editor, author of the journalistic story

Nicolas Kayser-Bril is a data journalist, and he works for AlgorithmWatch as a reporter. He pioneered new forms of journalism in France and Europe and is one of the leading experts on data journalism. He regularly speaks at international conferences, teaches journalism in French journalism schools, and gives training sessions in newsrooms. A self-taught journalist and developer (and a graduate in Economics), he started by developing small interactive, data-driven applications for Le Monde in Paris in 2009. He then built the data journalism team at OWNI in 2010 before co-founding and managing Journalism++ from 2011 to 2017. Nicolas is also one of the main contributors to the Data Journalism Handbook, the reference book for the popularization of data journalism worldwide.

/ Anna Mätzener

Editor

Anna Mätzener is managing director of AlgorithmWatch Switzerland. She holds a PhD in Mathematics from the University of Zürich, where she also minored in Philosophy and Italian Linguistics. Before joining AlgorithmWatch Switzerland, she was Editor for Mathematics and History of Science at an international scientific publisher, and most recently taught Mathematics at a high school in Zürich.

/ Lorenzo Palloni

Comic artist

Lorenzo Palloni is a cartoonist, the author of several graphic novels and webcomics, an award-winning writer, and one of the founders of comic artists collective, Mammaiuto. At the moment, he is working on forthcoming books for the French and Italian markets. Lorenzo is also a Scriptwriting and Storytelling teacher at La Scuola Internazionale di Comics di Reggio Emilia (International Comics School of Reggio Emilia).

/ Kristina Penner

Author of the chapter on Europe

Kristina Penner is the executive advisor at AlgorithmWatch. Her research interests include ADM in social welfare systems, social scoring, and the societal impacts of ADM, as well as the sustainability of new technologies through a holistic lens. Her analysis of the EU border management system builds on her previous experience in research and counseling on asylum law. Further experience includes projects on the use of media in civil society and conflict-sensitive journalism, as well as stakeholder involvement in peace processes in the Philippines. She holds a master's degree in international studies/peace and conflict research from Goethe University in Frankfurt.

/ Alessio Ravazzani

Comic artist

Alessio Ravazzani is an editorial graphic designer, cartoonist, and illustrator who collaborates with the most prestigious comics and graphic novel publishers in Italy. He is an author with the Mammaiuto collective, of which he has been a member since its foundation.

/ Friederike Reinhold

Additional editing of the introduction and policy recommendations

As a senior policy advisor, Friederike Reinhold is responsible for advancing AlgorithmWatch's policy and advocacy efforts. Prior to joining AlgorithmWatch, she worked as a Humanitarian Policy Advisor at the German Federal Foreign Office, with the Norwegian Refugee Council (NRC) in Iran, with Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) in Afghanistan, and at the WZB Berlin Social Science Center.

/ Matthias Spielkamp

Editor

Matthias Spielkamp is co-founder and executive director of AlgorithmWatch. He has testified before several committees of the German Bundestag on AI and automation. Matthias serves on the governing board of the German section of Reporters Without Borders and the advisory councils of Stiftung Warentest and the Whistleblower Network. He was a fellow of ZEIT Stiftung, Stiftung Mercator, and the American Council on Germany. Matthias founded the online magazine mobilsicher.de, reporting on the security of mobile devices, with an audience of more than 170,000 readers monthly. He has written and edited books on digital journalism and Internet governance and was named one of 15 architects building the data-driven future by Silicon Republic. He holds master's degrees in journalism from the University of Colorado in Boulder and in philosophy from the Free University of Berlin.

/ Beate Stangl

Graphic designer and layout artist

Beate Stangl is a qualified designer and she works in Berlin on editorial design projects for beworx, Friedrich Ebert Stiftung, Buske Verlag, UNESCO Welterbe Deutschland e.V., Sehstern Agency, iRights Lab, and Landesspracheninstitut Bochum.

/ Marc Thümmler

Publications coordinator

Marc Thümmler is in charge of public relations and outreach at AlgorithmWatch. He has a master's degree in media studies, has worked as a producer and editor in a film company, and managed projects for the Deutsche Kinemathek and the civil society organization Gesicht Zeigen. In addition to his core tasks at AlgorithmWatch, Marc has been involved in the crowdfunding and crowdsourcing campaign OpenSCHUFA, and he coordinated the first issue of the Automating Society report, published in 2019.

ORGANISATIONS

/ AlgorithmWatch Switzerland

AlgorithmWatch is a non-profit research and advocacy organization that is committed to watch, unpack and analyze algorithmic / automated decision-making (ADM) systems and their impact on society. While the prudent use of ADM systems can benefit individuals and communities, they come with great risks. In order to protect human autonomy and fundamental rights and maximize the public good, we consider it crucial to hold ADM systems accountable to democratic control. Use of ADM systems that significantly affect individuals' and collective rights must not only be made public in clear and accessible ways, individuals must also be able to understand how decisions are reached and to contest them if necessary. Therefore, we enable citizens to better understand ADM systems and develop ways to achieve democratic governance of these processes – with a mix of technologies, regulation, and suitable oversight institutions. With this, we strive to contribute to a fair and inclusive society and to maximize the benefit of ADM systems for society at large.

https://algorithmwatch.ch/en/

ALGORITHM WATCH / CH

/ Bertelsmann Stiftung

The Bertelsmann Stiftung works to promote social inclusion for everyone. It is committed to advancing this goal through programmes aimed at improving education, shaping democracy, advancing society, promoting health, vitalizing culture and strengthening economies. Through its activities, the Stiftung aims to encourage citizens to contribute to the common good. Founded in 1977 by Reinhard Mohn, the non-profit foundation holds the majority of shares in the Bertelsmann SE & Co. KGaA. The Bertelsmann Stiftung is a non-partisan, private operating foundation. With its "Ethics of Algorithms" project, the Bertelsmann Stiftung is taking a close look at the consequences of algorithmic decision-making in society with the goal of ensuring that these systems are used to serve society. The aim is to help inform and advance algorithmic systems that facilitate greater social inclusion. This involves committing to what is best for a society rather than what's technically possible – so that machine-informed decisions can best serve humankind.

https://www.bertelsmann-stiftung.de/en

BertelsmannStiftung

/ Engagement Migros

The Engagement Migros development fund supports pioneering projects, tackling the challenges of social change. They break new ground and test future-oriented solutions. To ensure the effectivity of this support, Engagement Migros supplements funding with coaching services provided by its Pionierlab. Engagement Migros is made possible by the companies of the Migros Group through an annual grant of approximately CHF 10 million. It has been supplementing the Migros Culture Percentage since 2012.

https://www.engagement-migros.ch

ENGAGEMENT
A DEVELOPMENT FUND OF THE MIGROS GROUP