MAPPING REGULATORY PROPOSALS FOR ARTIFICIAL INTELLIGENCE IN EUROPE

accessnow.org

accessnow

MAPPING REGULATORY PROPOSALS FOR ARTIFICIAL INTELLIGENCE IN EUROPE

Access Now would like to thank independent researcher Cori Crider for her extensive work on this report, as well as all experts from civil society, public authorities and the private sector consulted for this report. Finally, we would like to thank the Vodafone Institute for their support in the realisation of this report.

Nov 2018

With the support of the Vodafone Institute:

Vodafone Institute
for Society and
Communications

A. INTRODUCTION

> *"The main ingredients are there for the EU to become a leader in the AI revolution, in its own way and based on its values."* - European Commission AI Communication, April, 2018

The race is on to develop artificial intelligence (AI), and the EU has joined in.[1] With one eye on competitors from Silicon Valley to China, both individual member states and the European Union have announced "AI strategies," which funnel money into education, research, and development to kickstart European AI. At the same time, Europe's data protection authorities and oversight bodies are urging that AI must be subject to meaningful control. They cite headline-grabbing abuses of citizens' data—such as the use of algorithms to serve "dark ads" on Facebook and swing elections—to say that the Faustian bargain of comprehensive data-mining in exchange for "free" web services must end.

There is a tension here. The machine learning techniques that fuel AI have typically required vast quantities of training data. Europe's governments are understandably concerned not to miss the next great industrial revolution, and worry that over-regulation could fetter innovation. The stakes are high. Today in Europe, AI and algorithms may help decide whether a bank offers us a loan; whether our CV rises to the top of a pile; or even whether the police grant us bail.[2] [3] [4] In this context, the Cambridge Analytica-Facebook scandal, through which millions of users' and voters' data was unlawfully harvested and sold, was the tip of an iceberg.[5] [6] A potential crisis of trust looms between citizens, internet platforms, and governments over the risks of AI.

World leaders from Moscow, Washington to Beijing, have been engaging in a frenetic AI race and the fear of lagging behind is real. According to Russian president Vladimir Putin, the country that leads in

[1] **Definitions:** The phrase "artificial intelligence" is a wide umbrella that covers several more specific terms. In general, "**artificial general intelligence**" refers to a machine with the ability to apply intelligence to any task, rather than a pre-defined set of tasks, and does not yet exist. "**Narrow AI**," which describes current artificial intelligence applications, involves the computerised analysis of data, typically very large data sets, to analyse, model, and predict some part of the world. These can range from weather patterns, to the risk that a tumor may be malignant, to a human's credit-worthiness. It is these applications of AI--those already in use in society and being developed at pace--which are the focus of this paper. It may also be useful to define related terms that crop up in AI discussions: **"big data"** is a popular term that, in the Gartner IT glossary, refers to "high-volume, velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making." Many AI systems are programmed using a family of techniques referred to as **machine learning**. (See http://www.gartner.com/it-glossary/big-data. A useful summary of the debate about the term machine learning is available at https://www.techemergence.com/what-is-machine-learning/.) Machine learning divides broadly into two types: **supervised** learning, in which a given algorithm is developed on the basis of data which are already labelled by humans, and **unsupervised** learning, in which the software is not 'trained' by human labelling and instead left to find patterns in the data.

[2] *See* Wired, Europe's New Copyright Law Could Change the Web Worldwide, *available at* Financial Times, AI in banking: the reality behind the hype, *available at* https://www.ft.com/content/b497a134-2d21-11e8-a34a-7e7563b0b0f4

[3] The Guardian, Dehumanising, impenetrable, frustrating: the grim reality of job hunting in the age of AI, *available at* https://www.theguardian.com/inequality/2018/mar/04/dehumanising-impenetrable-frustrating-the-grim-reality-of-job-hunting-in-the-age-of-ai

[4] Wired, UK police are using AI to make custordial decisions - but it could be discriminating against the poor, *available at* https://www.wired.co.uk/article/police-ai-uk-durham-hart-checkpoint-algorithm-edit

[5] The Observer, Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach, *available at* https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

[6] The UK Information Commissioner notified its intention to fine Facebook for legal violations: See UK Information Commissioner's Office Press release, 10 July 2018, declaring decision "to fine Facebook a maximum £500,000 for two breaches of the Data Protection Act 1998" for the Cambridge Analytica breaches, *available at*
https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2018/07/findings-recommendations-and-actions-from-ico-investigation-into-data-analytics-in-political-campaigns/
£500,000 represents the maximum financial penalty available under UK data protection law until this year: this cap has been raised under the GDPR (and the UK Data Protection Act 2018) to the larger of €20 million or 4% of a company's global turnover.
In addition, several individuals from Cambridge Analytica are currently under investigation for possible criminal offenses. *See* UK Information Commissioner, Investigation into data use for political purposes update, at p. 23 *available at*
https://ico.org.uk/media/action-weve-taken/2259371/investigation-into-data-analytics-for-political-purposes-update.pdf

AI “will be the ruler of the world”.[7] While everyone seems to agree that we must all jump into the AI bandwagon, but no one seems to know where this train is going. Are countries forgetting to ask themselves the crucial question: what type of AI do we really want for our society? What impact will this technology have - or is already having - on people’s rights, on people’s life? One advantage, however, could set the EU ahead of the AI pack: the rule of law. The EU has the potential to lead the development of a human-centric AI by reaffirming its values and safeguarding rights. Regulation done right is an essential piece of this.

It will, of course, be essential to avoid knee-jerk lawmaking around AI: the controversies around the filtering and automated takedown of certain content in the EU, for example, show how ill-conceived and rushed regulation can threaten rights and freedoms.[8] A number of laws and proposals are pressing online platforms to automate the detection and speed up the suspension or removal of content. Experts we consulted for this report pointed to the German Hate Speech law and the EU’s recent debate on the Copyright Directive as some of the scenario where legislation might have been crafted with insufficient thought for the consequences.[9] At the same time, AI technologies are already being tested and used in sensitive and safety-critical areas of life (such as autonomous vehicles, cancer screening, or criminal justice) which may require intervention from the legislators.

Access Now believes that each area where AI is deployed will require a careful public and regulatory conversation: how should we meaningfully inform people about automated processes and safeguard rights such as the right to an explanation? If there are trade-offs between explainability in an AI system and accuracy, are there sectors where explainability must trump? Should it be left to individual users to interrogate and challenge algorithms that affect them, or are these collective problems that require a collective regulatory response? If it is the responsibility of government to address, for example, ethnic or gender bias in the way a given algorithm operates, which regulatory bodies are best equipped to do so? Finally, are there social areas where, for legal or democratic reasons, such as to protect human rights or the rule of law, the decision is too important or sensitive to leave to a machine at all?

Governments will shortly have to address all these questions for concrete applications of AI. They will need to decide where existing laws and enforcement bodies are equipped to address these risks, and where tweaks are required--whether regulators need more tools or regulation needs to be brought up to date. This does not necessarily mean stifling innovation, as European Data Protection Supervisor (EDPS) Giovanni Butarelli has said:

> *“In the gaps between obligations and prohibited practices, there is a vast hinterland of possibility. Good regulation steers innovation away from potentially harmful innovation and into areas of this hinterland where society can benefit.”*[10]

[7] The Verge, Putin says the nation that leads in AI ‘will be the ruler of the world’, *available at* https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world

[8] Germany recently passed a law imposing stiff penalties on internet platforms for hate speech on websites, essentially creating regulatory pressure on platforms such as YouTube and Facebook to engage in automated takedown of potential extremist content. This law, which is discussed in the section of the report on Germany, has been criticised by rights groups, see https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law
The United Nations special rapporteur on freedom of opinion and expression, David Kaye, said the draft law was at odds with international human rights standards, see https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf

[9] Wired, Europe’s New Copyright Law Could Change the Web Worldwide, *available at* https://www.wired.com/story/europes-copyright-law-could-change-the-web/

[10] Speech to the Telecommunications and Media Forum, *available at* https://edps.europa.eu/sites/edp/files/publication/18-04-24_giovanni_buttarelli_keynote_speech_telecoms_forum_en.pdf

As with automotive safety in the 20th century, creative regulation could become a mark not of European bureaucracy but of European quality.[11] The EU enjoys a robust tradition of human rights and effective regulation—from the EU Charter and European Convention on Human Rights, to world-leading data rights in the General Data Protection Regulation (GDPR), to products liability rules—and should see this as an asset. With care and foresight, Europe's legal scaffolding could support a sustainable, human-centred, and fair AI.

A word on scope: What does this report cover?

This report offers a bird's-eye survey of the major regulatory initiatives in AI in the EU and among member states. It draws on state bodies' published strategy papers and AI analyses, as well as states' consultations with experts who are helping develop regulation or assessing whether existing laws are fit for purpose. The report also canvasses differences between regulatory strategies and identifies possible risks and opportunities for human rights, transparency, and accountability.

Some of the initiatives covered here predate the explosion of interest in "AI" in terms and tend to refer to "big data". We have included these papers because their proposals are relevant as AI empowers societies to harness mass data.

This mapping report has assessed national strategies and opinions of authorities that have *explicitly* engaged with the challenges of AI or mass data regulation. The writing and thinking of the data protection authorities loom large in this space, for obvious reasons: processing of data, in particular its collection and analysis, is at the heart of AI. The laws regulating its use, and the regulators who enforce them, will be key players in this debate. The GDPR and Police Directive have been actively discussed in the context of AI, and thus are referred to in this report. The mapping has not included all general civil laws or regulations in each member state that may ultimately bear on a given AI system. It has also omitted prospective laws that may affect AI, such as legislation on cybersecurity, free flow of data and more. Ultimately, as AI becomes pervasive, it will intersect with many laws, from products liability, to patient confidentiality, to employment law. But a full assessment of AI's potential relationship to every national law is beyond the scope of this exercise.[12]

The aim of the report is to help everyone with a stake in AI—including civil society, unions, consumer groups, representatives of the private sector and legislators—participate in the development of this vital technology. Even now, AI is revolutionising our workplaces, hospitals, schools, and factories. Shortly it could touch every area of social and economic life. Much of this is positive: a properly trained worker, working *with* an AI diagnostic system (in manufacturing or medicine), may do her job far better and more efficiently than before. The data processing and analysis capabilities of AI can help alleviate some of the world's most pressing problems, from advancements in diagnosis and treatment of disease, to revolutionising transportation and urban living, to predicting and responding to natural disasters; to the benefits of workers, patients, or farmers. New high-skilled jobs will open as a result.

Yet these same capabilities can also enable monitoring and surveillance on a scale never seen before.

[11] Dating at least to the invention by Nils Bohlin of the three-point safety belt in 1959, safety innovations became a hallmark of Europe's competitiveness in the automobile industry, and involved a mix of private and public actors. See, e.g, https://www.volvocars.com/uk/about/our-company/heritage/innovations

[12] Other relevant laws and regulations at EU level may include the NIS Directive, current Cybersecurity Act, the Machinery Directive and Product Liability Directive (both being amended with AI in mind at the moment), the Radio Equipment Directive, the Free Flow of Data Regulation, as well as general principles of civil law, products liability, and public and administrative law.

They can be used to identify and discriminate against the most vulnerable. There are many areas where the social implications of AI require careful thought: should manufacturers be permitted to use haptic wristbands on workers to track and monitor their every gesture?[13] Is facial recognition software a way to make police more effective--or a recipe for encoding bias? AI cannot simply be “done to” workers, patients, or farmers *en masse* without engagement. We hope this report, by assessing where member States and the Union may be heading, will help stakeholders have their say and better understand the role that the EU can - and should - play in the AI race.

[13] The Verge, Amazon patents wristbands that track employees’ hands in real time, *available at* https://www.theverge.com/2018/2/1/16958918/amazon-patents-trackable-wristband-warehouse-employees

B. EUROPE-WIDE INITIATIVES

In one sense, the EU has a head start on developing AI law. New EU rules, particularly the General Data Protection Regulation and the Police Directive, stand to shape AI and mitigate its risks. And because many of these rules stretch beyond Europe's borders—any business who would seek to compete in Europe's vast data market must follow the GDPR—they may also contribute to set standards in Silicon Valley and beyond.

A. Existing laws: AI and the GDPR and Police Directive _

While these laws were not developed specifically for AI, they will set crucial benchmarks for the regulation of AI in Europe. By setting rules and safeguards around the processing of personal data, the GDPR and the Police Directive have the potential to directly impact the development and implementation of AI which is fueled by data.

The EU Commission has said little about how it expects these laws to apply to AI. This may be simply because the interpretation of laws is not mainly the Commission's role: it is the responsibility of the European Data Protection authorities and courts.

1. The General Data Protection Regulation

The GDPR contains seven core principles for the collection and processing of personal data:

- ➔ Lawfulness, fairness and transparency
- ➔ Purpose limitation
- ➔ Data minimisation
- ➔ Accuracy
- ➔ Storage limitation
- ➔ Integrity and confidentiality (security)[14]
- ➔ Accountability[15]

There is a broad European consensus by experts interviewed for this report that the GDPR will be relevant to AI development—but precisely how and to what extent is contested. The central debates include:

- ➔ The scope of the restrictions on fully automated processing and on profiling;
- ➔ How to respect the transparency and accountability requirement given current technical limits on explanation of some AI processes, such as deep learning and neural networks;
- ➔ How purpose limitation, minimisation and anonymisation can practically be achieved given the scale of many AI applications, the use of AI to find previously unrecognised patterns in data, and the sophistication of mass data analysis techniques;
- ➔ How to meet the GDPR's accuracy requirement in data when AI processes are inherently probabilistic; and
- ➔ How to support meaningful consent to AI processing.

[14] *See* GDPR Article 5(1), which sets the first six principles out. The full text of the GDPR is *available at* https://eur-lex.europa.eu/eli/reg/2016/679/oj

[15] GDPR Article 5(2), *supra* note 14. "The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('accountability')."

2. The Police Directive

The 2016 Police Directive will have an impact on the use of AI by law enforcement authorities in the EU.[16] The Directive aims to apply many of the rules governing personal data in the GDPR to the activities of law enforcement and investigative agencies, while still enabling these authorities to collaborate and share data when appropriate. It applies the central data protection tenets of the GDPR to police authorities in the EU, such as the requirement for a data protection officer, data protection impact assessments, and individual rights to seek amendment and correction for instance.

Among its key principles, are:

- ➔ data processing must be lawful and fair, carried out for "specified, explicit and legitimate purposes";
- ➔ subjects should be identified "for no longer than is necessary";
- ➔ there should be "periodic erasure of data,'"although this is subject to authorities' ability to carry out "archiving in the public interest…[including for] statistical or historical use";
- ➔ authorities must so far as practical distinguish between individuals suspected of an offence, convicted of an offence, and others potentially involved in the investigative or justice process, such as victims, witnesses, or associates of any of these;
- ➔ personal data based on "facts" should be distinguished from those based on "assessments";
- ➔ data that identifies sensitive personal characteristics (such as ethnicity, political affiliation, union membership) can be carried out "only where strictly necessary" subject to safeguards;
- ➔ data subjects have a (qualified) right to inspect, correct, and challenge the data processed about them for these purposes; and
- ➔ law enforcement data controllers must carry out many of the data protection activities others must do under the GDPR, such as create records of processing activities and logs, designate data protection officers, and carry out data protection impact assessments for high risk activities.[17]

The Police Directive, by its nature, requires states to pass an implementing legislation, as it is not directly applicable as a Regulation would be. This opens the door to a greater degree of local variance. The Directive also has several carve-outs for national security and public order policing that give law enforcement authorities considerably more manoeuvrability in their data processing activities than a regular data controller has.[18]

Crucial provisions of the law have yet to be tested in the context of AI in policing. These are applications that are likely to hold serious consequences for the lives of citizens. How will EU states determine if and when the use of AI by law enforcement is "necessary and proportionate in a democratic society"? Many AI applications are likely to raise questions under the Police Directive, including facial recognition, predictive policing, and others.

Among the potential questions that governments should be looking into are:

[16] *See* Police Directive *available at* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2016.119.01.0089.01.ENG

[17] *See* Articles 4 (1) and (3), 5-7,10,13,16,24,25,27, and 32-34 of the Police Directive *available at* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2016.119.01.0089.01.ENG

[18] *See* Article 15 of the Police Directive *available at* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2016.119.01.0089.01.ENG

- ➔ How can the Article 6 requirement to distinguish between specific types of people ("suspects" from "witnesses," for example) be squared with the use of mass data processes for investigative and public order purposes, such as the use of facial recognition on crowds which by definition include people that are irrelevant for specific law enforcement purposes?
- ➔ Do AI applications complicate the Article 7 requirement to distinguish between personal data based on "facts" and personal data based on "assessments"?[19]

The road ahead

Challenges remain. The adoptions of these two data protection laws were contentious: the GDPR passed in the teeth of, in the words of the European Data Protection Supervisor Giovanni Butarelli, "arguably the biggest lobbying exercise in the history of the European Union."[20] Precisely how they will impact AI applications is also likely to be contested. Some experts advising the EU have observed that AI's core functions call the very cornerstones of data protection and privacy into doubt.[21] If AI's main value is to scan mass data sets speculatively to find patterns, for example, how can this be squared with the GDPR's requirement that data must only be collected for a limited purpose? Some EU governments are concerned not to let these regulatory challenges frighten business away, but are also increasingly experiencing the benefits of having privacy and data protection laws in the wake of repeated data collection scandals.

The current European discussion reflects an effort to balance these imperatives: to attract AI talent and investments, while ensuring that AI businesses and the public sector understand and honour European law and traditions. We considered three bodies who have assessed AI regulation from a pan-European perspective: the European Commission, European data protection authorities (including the EDPS and the Article 29 Working Group, now European Data Protection Board), and the Council of Europe.

B. European Commission

Overall assessment: The EU Commission has proposed extensive funding for the development of AI technologies in Europe and their safe, equitable rollout to various sectors of the economy. The Commission's "Communication on AI" explains how the EU aims to promote AI. The Commission's *general* legislative innovations in the data space, the GDPR and the Police Directive, will be used to regulate AI, but precisely how remains an open question. The Commission is also assessing possible future amendments to the Products Liability and Machine Directives. Beyond this, the Commission's references to AI regulation at this stage tend towards soft norms.

[19] Consider, for example, the selection process used by PredPol's predictive policing algorithm: on one analysis, the dataset used to train the algorithm—arrest data in a given area—constitutes a set of "facts." A more nuanced analysis much suggest this data set is closer to one involving "personal assessment"—that is, the individual officers' decision to detain—because it does not capture whether the arrest led to convictions. This leaves open a further question: what if the data were *accurate* as to individual persons, in that they correctly captured the incidence of *e.g.,* non-violent drug offences in a given policed area, but *biased*, in that they failed to capture non-violent drug offences in other areas, owing to historical disparities in the way different communities are policed? See Kristian Lum and William Isaac, To Predict and Serve, *available at* https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x

[20] *See* Washington Post, Big tech is still violating your privacy, *avialable at* https://www.washingtonpost.com/news/theworldpost/wp/2018/08/14/gdpr/?noredirect=on&utm_term=.fab3af106226

[21] EDPS Ethics Advisory Group, Toward a Digital Ethics, Jan 2018, at 7: "The right to data protection may have so far appeared to be the key to regulating a digitised society. However, in light of recent technological developments, such a right appears insufficient to understand and address all the ethical challenges brought about by digital technologies….the tensions and frequent incompatibility of core concepts and principles of data protection with the epistemic paradigm of big data suggest limits to the GDPR even prior to its application." *Available at* https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf

Main regulatory proposals

In April 2018, the European Commission set out their plans on AI in a Communication on Artificial Intelligence.[22] The Communication:

- ➔ **Calls for new funding** to AI research and development (20 billion Euros a year by 2020), as well as the allocation of finances for retraining and other amelioration of AI's effects on the labour market.
- ➔ **Pledges investment in explainable AI "beyond 2020"**—that is, after the major infrastructure investments have been made.
- ➔ **Plans evaluation of AI regulation**. The Commission has set up working groups to consider whether existing regulations are fit for purpose. These working groups have already assessed the EU Products Liability and Machinery Directives for compatibility with AI. They aim to report back by mid-2019, at which point we can expect guidance on these two Directives. The Commission also plans a report on [*inter alia*] "the broader implications for, potential gaps in and orientations for, the liability and safety frameworks for AI" also by mid-2019.
- ➔ Indicates that the Commission will support **the use of AI in the justice system**, but offers no detail. There is no discussion of the risks of current AI applications used by police or in the criminal justice system.
- ➔ **Pledges to draft AI ethics guidelines by the end of the year.** These will address multiple rights issues and AI: "the future of work, fairness, safety, security, social inclusion and algorithmic transparency," as well as AI's impact on human rights such as "privacy, dignity, consumer protection and non-discrimination." The Commission will act on the ethical advice of a "high-level group on artificial intelligence" of 52 experts.[23] This work will be completed by the principles set out in the European Group on Ethics in Science and New Technologies (EGE)'s "Statement on AI, Robotics, and Autonomous Systems."[24]
- ➔ Proposes **dedicated retraining schemes**, diversion of resource from the European Social Fund, and widening the scope of the Globalisation Adjustment Fund to cushion redundancies from automation and mitigate AI's possible effects on inequality.
- ➔ **Calls for prompt adoption of the proposed ePrivacy Regulation** and **Cybersecurity Act** to "strengthen trust in the online world." and
- ➔ **Notes the role of the GDPR**—in particular its limitations on profiling and automated decision-making—and call on data protection authorities to "follow [GDPR's] application in the context of AI." But the Communication says little about how in practice these laws will impact AI.

Finally, under the "Digital Single Market" framework, the European Commission has begun a 16-month "algorithmic awareness-building" exercise.[25] This will study how algorithms shape public decision-making and aims to help design policy responses to the risk of bias and discrimination in AI. No findings have yet been published.

[22] *See* EU Commission Communication on AI, *available at* https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

[23] *See* EU Commission High Level Expert Group on AI, *available at* https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

[24] *See* Statement on AI and robotics, *available at* http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf
The EGE is the independent ethics and science advisory body for the Commission. This paper describes AI regulatory efforts as "a patchwork of disparate initiatives" and calls for a more centralised effort to develop and apply the law to AI. The EGE also set out a list of ethical principles it says should should guide AI development, but stops short of recommending concrete changes to law or regulation. The principles are Human dignity, Autonomy, Responsibility, Justice, equity, and solidarity, Democracy, Rule of law and accountability, Security, safety, bodily and mental integrity, Data protection and privacy, and Sustainability.

[25] *See* Algorithmic Awareness Building, *available at* https://ec.europa.eu/digital-single-market/en/algorithmic-awareness-building
Their findings will be published on a new website, algoaware.eu

C. European Data Protection Authorities

Much of the deepest and most creative thinking about AI regulation has been done by Europe's data protection authorities: in particular, the Article 29 Working Party (which became the European Data Protection Board after the GDPR came into application)[26] and the European Data Protection Supervisor.

The documents produced by data protection authorities, through the European Data Protection Board, the EDPS or at national level, are not laws but rather guidelines and opinions which contribute to the implementation of binding legislations that may impact AI such as the GDPR or the Police Directive.

1. Article 29 Working Party & European Data Protection Board

Overall assessment: The Article 29 Working Party (WP29) has produced Guidelines that will affect how the GDPR will apply to AI and help entities to comply with the law in an harmonised manner. As of yet, however, there is an unresolved tension between the speculative and probabilistic nature of many AI applications, and the limitations the GDPR tends to impose.

In what circumstances should the United Kingdom National Health Service hospitals, for example, be able to scan historical patient data to identify public health trends or new disease treatments?[27] Does the picture change if this analysis is performed by a corporate contractor? Given that medical data contain citizens' intimate details, how should states and their partners conduct important medical research while preserving patient confidentiality?

Issues like these will need to be explored further and in concrete cases by data protection authorities, legislators, and the courts.

Main regulatory proposals

I. Guidelines on automated decision making and profiling

Many economically valuable uses of AI involve to assess, categorise, profile, and predict human behaviour through data analysis. Where those predictions are used to make (or assist) consequential decisions—for example, to determine whether someone will be given a job interview or granted a loan, access to healthcare or to a school —they raise human rights concerns, including around privacy, data protection, and equality and non-discrimination.[28]

These uses will be partially regulated by the GDPR. For instance, Article 22 of the GDPR prohibits fully automated processing "where it would produce a legal or similarly significant effect", except in certain cases. What constitutes a "significant effect" on people? How will this measure be

[26] *See* Europe's new data protection rules and the EDPB: giving individuals greater control, *available at* https://edpb.europa.eu/news/news/2018/europes-new-data-protection-rules-and-edpb-giving-individuals-greater-control_en (explaining that the EDPB succeeds the Article 29 Working Group)

[27] *See, e.g.,* FAQ, Moorfields Eye Hospital - Deepmind collaboration, *available at* https://www.moorfields.nhs.uk/faq/deepmind-health-qa

[28] The rights implicated depend on the circumstances, but potentially include: the right not to be discriminated against; the right to privacy; the right to free expression and association; and data protection rights. See for instance https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals and https://www.theguardian.com/inequality/2018/mar/04/dehumanising-impenetrable-frustrating-the-grim-reality-of-job-hunting-in-the-age-of-ai

implemented and what level of information will people be given? This is an area likely to create significant legal debate; in particular as the exceptions give Member State the authority to legislate around this requirement.

The Article 29 Working Party's October 2017 Guidelines on automated decision-making and profiling anticipate these applications of AI.[29] The Guidelines distinguish a) fully automated processing, and b) profiling, where AI is used to inform or influence human decision-making.

The Guidelines assess what the GDPR requires of processors using AI for these two uses and sets out highly specific requirements:

- ➔ Telling subjects they are being profiled;[30]
- ➔ Providing "meaningful information about the logic involved" - a legal requirement likely to spur debate given current thinking on explainability and some AI processes (such as neural networks);[31]
- ➔ "Explaining the significance and envisaged consequences" of the processing; and
- ➔ Giving the subject an explanation of the decision reached and an opportunity to challenge the decision.

The other general requirements under the GDPR also apply: processing of data must be minimised, accurate, not stored for longer than necessary, and the AI processor/controller must be accountable for their use of the data.

Another crucial explanation in this paper is to treat inferred data as personal data for data protection purposes – in other words, AI companies will be violating the laws if their AI sifts people in ways that are highly correlated with a protected characteristic, such as race.

Finally, the Guidelines provides some credit examples of use of the profiling may be unfair and unlawful where it may "create discrimination, for example by denying people access to employment opportunities, credit or insurance, or targeting them with excessively risky or costly financial products."[32] In practice, these profiling restrictions are likely to mean that certain AI applications previously used in the United States (the use of AI to target low-income people in search engines for predatory loans, for example) are unlawful in the EU.

II. Guidelines on "high risk" processes requiring DPIA

Another major regulatory requirement imposed by the GDPR on data controllers, including those who use AI, is to conduct data protection impact assessments (DPIAs) for "high risk" activities. The Working Party 29 Guidelines on "high risk" processes requiring DPIA interprets these requirements and put

[29] *See* Guidelines on automated decision making and profiling http://ec.europa.eu/newsroom/article29/document.cfm?doc_id=49826

[30] *See* p. 9-10 and 16 of the Guidelines on automated decision making and profiling: "controllers must ensure they explain clearly and simply to individuals how the profiling or automated decision-making process works." *available at* http://ec.europa.eu/newsroom/article29/document.cfm?doc_id=49826

[31] The best practice recommendations deal with this by explaining that the source code or algorithm is not what is required: "Instead of providing a complex mathematical explanation about how algorithms or machine-learning work, the controller should consider using clear and comprehensive ways to deliver the information to the data subject, for example: · the categories of data that have been or will be used in the profiling or decision-making process; · why these categories are considered pertinent · how any profile used in the automated decision-making process is built, including any statistics used in the analysis; · why this profile is relevant to the automated decision-making process; and · how it is used for a decision concerning the data subject." At p. 31 of the Guidelines on automated decision making and profiling *available at* http://ec.europa.eu/newsroom/article29/document.cfm?doc_id=49826

[32] *See* p. 10 of the Guidelines on automated decision making and profiling *available at* http://ec.europa.eu/newsroom/article29/document.cfm?doc_id=49826

them in context to provide guidance to data controllers.[33]

In brief, this opinion holds that most applications of AI involving human subjects are likely to be "high risk" processing and will therefore require a DPIA**.** Many of the circumstances mentioned in the Guidelines plainly apply to AI:

- ➔ "Evaluation or scoring" - which can for instance be used when conducting automated predictions and profiling;
- ➔ "Automated decision-making with legal or similar significant effect";
- ➔ "Systematic monitoring" of individuals, including through a publicly accessible area - which can refer to the use of facial recognition of cameras for instance;
- ➔ "Sensitive data" or data of a highly personal nature - which could be included in algorithm;
- ➔ "Data processed on a large scale" - which would be the case of most AI use cases;
- ➔ "Matching or combining datasets" for a new purpose - another popular application of AI;
- ➔ "Data concerning vulnerable subjects";
- ➔ "Innovative use or applying new technical solutions"; and
- ➔ "Data transfer across borders outside the European Union" – which is relevant given the location of many of the major AI developers in the US and China.

In practice, this means that most entities seeking to use AI will have to conduct DPIA prior deployment of the technology for a specific use. To help in that process, the Guidelines set forth best practices for data protection impact assessments with detailed criteria and methodology to develop an acceptable data protection impact assessment.[34]

III. Other potentially relevant papers: Guidelines on consent and purpose limitation

The WP29 published guidelines that will likely apply in the AI context and will need to be critically assessed by regulators and courts.

The guidelines consent notes that consent is only a lawful basis for data processing if the users have control and a genuine choice about whether to accept a given term.[35] Public authorities are unlikely to be able to use consent to justify data processing because of the inherent power imbalance between state authorities and citizens. Similar concerns limit its use in an employment context. The guidelines further note that the (prevalent) practice of bundling consent to data processing—including processing unnecessary to carry out the service offered—with terms of consent is unlikely to constitute free consent and may fall afoul of the GDPR. Strictly construed, we believe this would potentially invalidate much of the current business model of many large internet platforms. The guidelines further note limits on use of consent as a legal basis to use data for different purposes as consent must be granular. In practices, this means that entities using AI would have to specifically define their objective with the use of data and could not request consent for general development of services or undefined future innovations.

The open question is how this requirement can be met when AI aims to spot emergent properties of data that are unknown at the outset. What should the privacy notices say, and how often should AI processors be required to go back to data subjects? What does this imply for the intention on the part of the EU and member states to open up historical data to AI analysis?

[33] *See* Working Party 29 Guidelines on "high risk" processes requiring DPIA http://ec.europa.eu/newsroom/document.cfm?doc_id=44137

[34] *See* Annex 2 of the Working Party 29 Guidelines on "high risk" processes requiring DPIA *available at* http://ec.europa.eu/newsroom/document.cfm?doc_id=44137

[35] *See* Working Party 29 Guidelines on consent *available at* http://ec.europa.eu/newsroom/just/document.cfm?doc_id=48849

The Guidelines on purpose limitation will also intersect with AI in a complex way. [36] Purpose limitation in data protection has two cornerstones: data must be collected for "specified, explicit and legitimate" purposes, and not be "further processed in a way incompatible" with those purposes. How will this important principles be applied when the collection and use of data is speculative – medical diagnostics, insurance, credit risk – is uncertain.

2. European Data Protection Supervisor

> *"The perverse incentive in digital markets to treat people like sources of data has to be remedied."*[37]

The European Data Protection Supervisor, Giovanni Butarelli, has published extensively on AI regulation.

Overall assessment: The EDPS is clearly alive to the challenges and trade-offs for privacy, data protection, and other human rights that AI and other mass data technologies entail. It is urging European authorities to regulate the mass internet platforms' use of data (and AI in particular) more stringently, and for the businesses developing AI to think creatively about how to innovate in ways that respect data protection law. Many of the EDPS recommendations are thorough and thoughtful. It has also added to the debate by having independent ethics experts assess the shortcomings of data protection law to AI challenges.

Main regulatory proposals

The EDPS have proposed much greater transparency around the use of mass data, an end to covert profiling and impenetrable privacy notices. The EDPS gave concrete recommendations on users' rights to give them more control by a) featurisation of data access – that companies must make it easier for users to access the data held about them, and b) data portability - tools that make it simple for users to move their data. The EPDS also press developers of AI to propose specific techniques to protect individuals' data through anonymisation.

The EDPS concur that data protection is not the only source of law that will be necessary to regulate AI, making reference to consumer protection, antitrust, and technical research and development as supplements.

In a rather provocative move for a regulator, the EDPS instructed ethics experts to assess to what extent the data protection laws remain fit for purpose in a world of big data. The EDPS has however has highlighted in many occasion the risks posed by AI to personal dignity, democratic integrity, discrimination, and the commoditisation of data.

Finally, the EDPS has repeatedly emphasised that the single most important legal and ethical driver with the power to change AI is the GDPR's principle of accountability.[38] This requires data protection authorities to have the power and resources to enforce the accountability principle.

[36] *See* Working party 29 Guidelines on purpose limitation *available at* http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

[37] April 2018 keynote speech to telecommunications forum, *available at* https://edps.europa.eu/sites/edp/files/publication/18-04-24_giovanni_buttarelli_keynote_speech_telecoms_forum_en.pdf

[38] *See* EDPS Speech to Telecommunications and Media Forum, *available at* https://edps.europa.eu/sites/edp/files/publication/18-04-24_giovanni_buttarelli_keynote_speech_telecoms_forum_en.pdf

I. 2015 opinion: meeting the challenges of big data

> *"The EU intends to maximise growth and competitiveness by exploiting big data. But the Digital Single Market cannot uncritically import the data-driven technologies and business models which have become economic mainstream in other areas of the world. Instead it needs to show leadership in developing accountable personal data processing. "*

The EDPS' 2015 opinion on big data sets out four main requirements for the lawful and sustainable use of big data.[39] It says organisations must :

- ➔ be much more transparent about how they process personal data; (i.e.,"end covert profiling");
- ➔ afford users a higher degree of control over how their data is used;
- ➔ design user friendly data protection into their products and services; and
- ➔ become more accountable for what they do.

The EDPS proposes that the GDPR transparency requirement should include "the disclosure of the 'logic of decision making', the data itself, as well as its source."[40] This is the case even where the personal data processed is (as often, with AI) inferred—for example, where predictions are made about us based on tracking our activity online. Disclosure of information to individuals must be clear and tailored, with layered (step-by-step) privacy notices. When complying with the GDPR, organisations seeking to obtain users' consent, the opinion notes that it "requires a clear understanding [by the user] of what one agrees to" and contain the right to object and opt-out.[41]

Design solutions proposed to GDPR compliance include "functional separation" - where only the amount of data needed for a given process is collected and used. It also means limits on onward transfer of data: data collected for "research" shouldn't then inform consequential decisions about people without their knowledge or consent. The EDPS adds that anonymisation techniques are difficult but still important in an AI era.

The EDPS further sets out practical steps AI companies can take to process data responsibly and demonstrate compliance, including internal controls, impact assessments, and audit trails.[42] Finally, the EDPS notes that DPAs need both powers and resources to enforce the law.

II. AI discussion paper

An AI-specific 2016 discussion paper prepared for the International Conference of Data Protection and Privacy Commissioners sets out the EDPS' understanding of the core principles that will underpin AI regulation and notes particular challenges AI poses for DPAs.[43] It expresses concern about the prevalence of profiling:

[39] *See* EDPS opinion on meeting the challenge of big data, *available at* https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf

[40] *See* p. 10. of EDPS opinion on meeting the challenge of big data, *available at* https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf

[41] At p.11, noting that consent ""has never meant long and impenetrable privacy policies, written by lawyers for lawyers, which users must 'consent' to unless they wish to abandon the use of the desired service altogether. Instead, it means a genuine, freely-given choice with the alternative, without any detriment, to say 'yes'." *See* EDPS opinion on meeting the challenge of big data, *available at* https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf

[42] *See* p. 15-16 of EDPS opinion on meeting the challenge of big data, *available at* https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf

[43] The 2016 ICDPPC Marrakesh paper on AI, *available at* https://edps.europa.eu/sites/edp/files/publication/16-10-19_marrakesh_ai_paper_en.pdf

> *"The use of artificial intelligence to predict people's behaviour risks stigmatisation, reinforcing existing stereotypes, social and cultural segregation and exclusion, subverting individual choice and equal opportunities.*[44]*"*

The AI paper points to two major obstacles to making AI-led decisions more transparent:

- ➔ the use of "trade secrets" to keep AI-led processes private; and
- ➔ the technical limitations on explainability that underpin much of machine learning.

Opening the "black box" is not going to be enough for DPAs to assess an AI system, in the EDPS' view, because "the analysis needs to be done on the machine learning process itself."[45]

The paper reflects on several contentious uses of AI that regulators will need to address.This includes the privacy and surveillance implications of AI-driven facial recognition technology, used, for example, by border control; the EDPS observes that this will need to be supervised even where used for security or intelligence purposes. Natural language processing, and the impetus to scan old documents to improve it, may erode the data protection principle of purpose limitation. In an important general point, the EDPS says AI developers must ask themselves: how much data is really necessary to make an AI system work? Finally, the paper queries who ultimately will be held accountable for autonomous machines, such as autonomous weapons systems or self-driving cars.

III. AI ethics paper

The EDPS has looked to the horizon by commissioning an expert paper—"Towards Digital Ethics"—which sets out major ethical and legal issues that AI regulators are likely to grapple with.[46] [47]

This paper makes no regulatory recommendations, but is a subtle and sophisticated canvassing of how AI challenges existing norms. It offers an extended discussion on the risks to democracy, fair trial, and other collective European values from AI. It also notes the trend toward commodification of individuals' data (in policymaking terms, the shift in treating people from individuals to treating them as data). It is critical of this growing tendency, saying it conflicts with the spirit of dignity in the EU charter:

> *"When individuals are treated not as persons but as mere temporary aggregates of data processed at an industrial scale…they are arguably, not fully respected, neither in their dignity nor in their humanity."*[48]

Yet resolving this challenge through traditional data protection may prove difficult. In an era of mass data, the paper notes, many of the traditional principles of data protection will come under serious challenge. Portability and norms of individual control, too, only go so far, when citizens are strongly identified with their data: "My in 'my data' is not the same as in 'my car' but rather the same as in 'my hands'."[49]

The paper sets out five ethical principles and risks that EU regulation should address:

[44] *See* p. 16 of the 2016 ICDPPC Marrakesh paper on AI, *available at* https://edps.europa.eu/sites/edp/files/publication/16-10-19_marrakesh_ai_paper_en.pdf

[45] *See* p. 4 of the 2016 ICDPPC Marrakesh paper on AI.

[46] *See* Towards Digital Ethics - EDPS Ethics Advisory Group, *available at* https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf

[47] Digital ethics is the theme of the forthcoming 40th ICDPPC in Brussels, and the DPAs are expected to adopt a resolution on this point. https://www.privacyconference2018.org/en/40th-international-conference-data-protection-privacy-commissioners

[48] *See* p.17 of Towards Digital Ethics - EDPS Ethics Advisory Group, *available at* https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf

[49] *See* p.25 of Towards Digital Ethics - EDPS Ethics Advisory Group, *available at* https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf

1. The dignity of the person remains inviolable in the digital age;
2. Personhood and personal data are inseparable from one another;
3. Digital technologies risk weakening the foundation of democratic governance;
4. Digitised data processing risks fostering new forms of discrimination; and
5. Data commoditisation risks shifting value from persons to personal data.

D. Council of Europe

Overall assessment: As Europe's principal human rights monitor, the Council of Europe is well-placed to offer a bird's eye assessment of the legal requirements and risks of AI. Their analysis fills some of the gaps of the legal analysis of the data protection authorities and of the EU Commission proposals, working as it does from the cornerstone of Europe's human rights laws. The Council of Europe express the general view that the state of the conversation about AI technology is still too nascent for knee-jerk regulatory responses, but helpfully identify many of the crucial risks.

Main regulatory proposals

I. Paper on algorithmic decisions and human rights

In March 2018, the Council of Europe weighed in on AI regulation. A group of internet experts, chaired by Prof. Wolfgang Schulz, published a "Study on the Human Rights Dimensions of Automated Data Processing Techniques (in particular algorithms) and possible regulatory implications."[50] The paper covers the implications of AI for most of the major human rights.[51]

Because the technology is new and poorly understood by policymakers, they conclude that in the first instance most regulatory responses should focus more on "greater transparency and accountability surrounding the use of algorithms," as well as new ethical frameworks and risk assessments, rather than "direct regulation."[52]

On transparency, the Council echoes other bodies' concerns about the technical limits of explainability and the proprietary nature of much of the algorithmic information, but proposes partial disclosure:

> *"key subsets of information about the algorithms [should] be provided to the public, for example which variables are in use, which goals the algorithms are being optimised for, the training data and average values and standard deviations of the results produced, or the amount and type of data being processed by the algorithm.*[53]

The Council notes that it is primarily states' responsibility to develop accountability frameworks - that it cannot be left to the private sector and technical innovation to protect human rights. Like the EU Commission, they believe product liability standards require updating in the AI context, and ask whether the developer or the user will typically be liable.[54] Public authorities in particular must be

[50] *See* Study on the Human Rights Dimensions of Automated Data Processing Techniques and possible regulatory implications, *available at* https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10

[51] In 2018-19, the Council of Europe will also publish an expert study on the human rights dimensions of automated data processing, *"mapping legal and ethical considerations within the existing human rights framework".*

[52] The Council of Europe note that "there is far too little information available to make well-founded decisions on this topic" on p. 43 of the Study on the Human Rights Dimensions of Automated Data Processing Techniques and possible regulatory implications, *available at* https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10

[53] *See* p.38 of Study on the Human Rights Dimensions of Automated Data Processing Techniques and possible regulatory implications, *available at* https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10

[54] At p. 39: "Another avenue to explore is whether existing product liability regulation should be extended to include software? Or are rather the public or private actors to be held accountable who purchase the algorithm and introduce it into their services, even without

held accountable when they use algorithms to inform their decisions. Pointing to recent stories about dark political ads on Facebook, they urge attention to electoral integrity and the negative effects algorithmic targeting has had on democracy. At the same time, they criticise the new requirement for internet platforms to monitor content in an automated way, saying this conflicts with freedom of expression.

They suggest new ethical frameworks to AI, such as "professional ethics" codes for algorithm designers modeled on those for doctors or lawyers. They point to impact assessments and emerging industry standards, including those set by the Institute of Electrical and Electronics Engineers as useful protections for human rights.

The Council of Europe further provides a set of specific recommendations which includes:

➔ a proposal for licensing schemes for algorithms, already used in the gambling context, may be appropriate for wider sectors of the economy;
➔ a call for data protection authorities to be funded and well-supported;
➔ a proposal for Europe to consider whether to regulate news on the internet platforms as it did traditional broadcasters;
➔ a note that regulators (insurance, banking, and others) will need guidelines to control the use of algorithms on their sectors; and
➔ a call for techniques for auditing algorithms, including "zero knowledge proofs," to be developed to test for bias without seeing the underlying source code.
➔ Looking ahead to what *may* need to be regulated, the Council of Europe has identified a comprehensive set of rights potentially risked by unregulated AI. It contains extensive discussions of the implications for automated data of human rights, including on the right to a fair trial and due process, the right to data protection, freedom of expression and freedom of assembly and social rights.

understanding its operation?" *See* Study on the Human Rights Dimensions of Automated Data Processing Techniques and possible regulatory implications, *available at* https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10

C. MEMBER STATE INITIATIVES

As with the EU, most member states' AI strategies centre on how to support AI research and development. That said, many advisory bodies, including national data protection authorities, have done forward-thinking work on the management and regulation of AI.

A.

FRANCE

🇫🇷

Overall assessment: Much of France's strategy discusses diversion of funds and attracting engineering talent. There is solid and creative thinking in the advisory paper that informed the strategy around the ethical and regulatory challenges posed by AI, but at the moment the proposed solutions largely involve the creation of groups to study them rather than the proposal of new or modified norms.

Main regulatory proposals

France's AI strategy generally cleaves to the "ethics" framework and makes scant reference to hard legal constraints on AI development. The Villani paper does point to some sources of hard law (noting that France has had a national data protection right against fully automated decision making since 1978, for example), and proposes some creative approaches to AI regulation. CNIL, the French data protection authority, has also laid out proposals to improve the auditability and transparency of AI systems.

National AI Strategy, "AI for Humanity"

France published its strategy, AI for Humanity, in March 2018.[55] The major planks of their platform are:

➔ **Developing an open data policy**

The strategy states that "France has a key asset: massive centralised databases. The problem is that they are underexploited." The document simultaneously asserts that these datasets will be opened up for use in, *e.g.*, the agricultural, transport, or health sectors, but that they will be done so "accompanied by a European framework for the protection of personal data." This balance will be a major regulatory challenge for all EU member states: unlocking the use of valuable state personal and non-personal data but in a way that involves meaningful consent and which ensures that the benefits of this data processing go to the public and not simply to processing corporations.

➔ **Creating a "favourable" regulatory environment for the creation of "AI champions"**

This appears to refer to relaxing or amending some regulations to support AI development for example, driverless cars. We note here that the french strategy uses

[55] *See* AI for Humanity, *available at* https://www.aiforhumanity.fr/en/

A.

FRANCE

🇫🇷

positive language to describe a deregulation process instead of proposing discussion on regulation as a way to enable favourable innovation.

- **“Giving thought” to regulation and ethics**

In the main, the regulatory recommendations carried through to the announced strategy from the Villani paper are soft law, rather than hard. They propose supporting research on AI ethics—France will set up an international panel of experts on AI, with priorities on “transparency and fair use.” One of the major areas of research will be info “explainable models” for AI to support greater transparency. To the same end they also pledge that all algorithms used by the state should be public, and plan to encourage diverse talent in the development of AI to mitigate the risk of bias.

The Villani report

The Villani report is considerably more detailed about the ethical and legal challenges posed by AI.[56] The report complements the national strategy and proposes a number of regulatory approaches, albeit it often recommends “soft” norms and consent-based models of regulation.

Some of the key regulatory mechanisms the Villani paper recommends include:

- Encouraging data controllers to pool data;
- Opening up “public interest” access to some data sets by the government;
- Enabling the right to data portability (as set out in GDPR);
- that France focus AI development on health, transport, the environment, and “defense and security” – this would appear to include a focus on AI-run armaments and surveillance apparatus;
- review the EU-US Privacy Shield arrangement allowing for data transfer to assess whether it is sufficiently protective of privacy or gives away too much data;
- a public lab for transformation of work – to prepare society for the displacement caused by automation, as well as law-making project to deal with working conditions in the digital age;
- promoting AI to mitigate the ecological effects of expanded computer use—estimated to use between 20-50% of global electricity consumption by 2030;
- AI processors should carry out “discrimination impact assessments” along the lines of existing privacy impact assessments in French law;
- set boundaries for predictive algorithms in law enforcement and “discuss” development for autonomous weapons, and create an ‘observatory’ for their non-proliferation; and
- a call for diversity: in particular, pulling more women and underrepresented groups into the development of AI.

[56] *See* Report of the Mission Villani on AI, *available at* https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

The paper also makes several recommendations to make sure that AI and machine learning systems obey the principles of transparency and accountability:

- ➔ making sure organisations who deploy AI and machine learning systems remain legally liable for damage caused;
- ➔ creating an expert group who will analyse “more explainable models” and “more intelligible user interfaces” for AI. This group would help create audit tools and processes for the use of algorithms, particularly in the litigation context.;
- ➔ suggests state funding of public interest groups who would be qualified to investigate and report on AI uses;
- ➔ including “ethics” in training for AI engineers and researchers;
- ➔ formulating improved “collective rights” concerning data (recognising that AI systems often have a mass, rather than an individual, effect), such as support for data class actions and right to compensation - albeit the paper proposes that no monetary penalties should be awarded, only injunctive relief.
- ➔ establishing a consultative ethics committee which would organise public debate.

The Villani paper calls for greater transparency, auditability, and accountability for algorithms, and “research” into how to make them more accountable, and contains a number of suggestions for soft norm development. France promotes the principle of transparency by mandating that all French state-developed and run algorithms, should be public. On the other hand, the attempt to open its national databases to businesses and researchers may, depending on its method, present privacy and data protection concerns. The suggestion of data class actions with compensation for injury sustained is a unique and interesting hard law recommendation that would improve accountability provided that necessary safeguards are put in place.

Of concern, neither the Villani paper nor the AI strategy rule out France’s development of autonomous weapon systems - they simply say it needs debate before development. The Villani paper also contains forward-thinking proposals on environmental uses of AI that are missing from most other national AI analyses.

Broadly missing is a sense of what limits there should be on use of AI systems by public authorities to ensure compliance with rights such as fair trial, freedom of expression and association.

Commission Nationale de l’Informatique et des Libertés (CNIL)

CNIL, France’s data protection authority, has published a paper on the ethics of AI, called “How can humans keep the upper hand?”.[57] This paper was taken into account in the development of the French national AI strategy and it was presented to representatives of the French government. The CNIL paper is the result of around 65

[57] *See* Comment permettre à l’homme de garder la main sur les enjeux éthiques des algorithmes, *available at* https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de

A.

FRANCE

🇫🇷

public debates involving around 3,000 individuals in France. Its six main recommendations are:

- ➔ Improving citizens’ digital literacy and their capacity to think critically about AI systems (as well as the capacities of designers);
- ➔ Making algorithmic systems comprehensible by strengthening rights and “rethinking mediation with users”;
- ➔ Improving the designs of algorithmic systems to prevent the “black box” effect;
- ➔ Increased incentives for research into ethical AI, including a major national research project; and
- ➔ Strengthening ethics in the AI companies themselves.

In parallel, the CNIL identifies what it sees as the six major ethical challenges raised by AI, which will require further work on:

- ➔ The threat to free will and responsibility posed by autonomous machines;
- ➔ Bias, discrimination, and exclusion;
- ➔ Algorithmic profiling: personalisation versus collective benefits;
- ➔ Preventing massive files while enhancing AI: seeking a new balance;
- ➔ Quality, quantity, relevance: the challenge of data selection; and
- ➔ Human identity in the age of AI.

B.

GERMANY

🇩🇪

Overall assessment: While Germany has set out the cornerstones of its national platform for AI, the country seems to be comparatively early in its thinking about AI regulation. The publication of its national strategy has reportedly been delayed until November 2018. The country has however set principles governing self-driving cars are already serving as a regulatory example, with China saying it intends to model its regulation on them.[58]

Main regulatory proposals

Germany currently has a Platform for AI—essentially an expert advisory committee, combining officials, businesspeople, and academics—and has published the “Cornerstones” for the forthcoming national strategy.[59] [60]

The Cornerstones briefly address many of the signal rights debates in AI. For example, they acknowledge that citizens expect “justified trust” in AI “on the basis of transparent procedures and traceability”. The Cornerstone pledges to “promote the development of procedures for control and traceability of algorithmic forecasting and decision systems.”

[58] *See* Reuters, China may adopt some of Germany’s law on self driving cars, *available at* https://www.reuters.com/article/us-autos-autonomous-germany-china/china-may-adopt-some-of-germanys-law-on-self-driving-cars-expert-idUSKCN1GR2TJ

[59] *See* Platform for AI, *available at* https://www.plattform-lernende-systeme.de/ai-strategies.html

[60] *See* German national strategy, layout, *available at* https://www.bmbf.de/files/180718 Eckpunkte_KI-Strategie final Layout.pdf

B.

GERMANY

🇩🇪

Germany has passed “NetzDG,” a law on hate speech online with severe financial penalties that is likely to affect algorithmic processing of material by large internet platforms.[61] This hate speech law has come in for severe but, in our view justified, criticism for its restrictions on freedom of expression. The Cornerstones propose a series of further regulatory changes. It notes that Germany will work changes its competition and copyright laws to open up data while still protecting privacy. It supports both the applicability of the GDPR and the proposed ePrivacy Regulation in this context.

Germany’s most developed regulatory thinking on AI is its Ethics Commission on Automated Driving, which has published principles.[62] These deal extensively with accountability and allocation of liability, and ban any discrimination between people in the event of accident.

Finally, the document also acknowledges the need to avoid bias and discrimination when AI is used in public decision-making. It affirms the importance of effective legal protection (or due process) for citizens controlling the public use of AI. It also discusses the need to ensure that AI development is “people-centered” and proposes observatories assessing the future of work, monitoring the impact on employment, and a national retraining strategy among other ameliorating policies on AI’s effect on the workforce. Germany has also established a national commission to study the social effects of algorithmic decision-making.[63]

C.

THE UK

🇬🇧

Overall assessment: While the UK National Strategy on AI only addresses the need for “regulatory innovation” in general terms, solid thinking has been done on the point by other UK bodies. Both the UK House of Lords (HOL) and the UK’s data protection regulator (the ICO) have published extensively on AI regulation. Whereas the HOL focuses on high-level areas like “explainability”, the ICO has made detailed recommendations about how to manage and regulate AI systems, both in general and in the recently-relevant context of social media and the integrity of elections.

Main regulatory proposals:

UK Government policy

Like most of the national strategies on AI, the UK AI Sector Deal sets out a raft of

[61] *See* NetzDG, *available at* https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf;jsessionid=829D39DBDAC5DE294A686E374126D04E.1_cid289?__blob=publicationFile&v=2

[62] *See* Principles of Ethics Commission on Automated Driving, *available at* https://www.bmvi.de/SharedDocs/EN/Documents/G/ethic-commission-report.pdf?__blob=publicationFile

[63] *See* National Commission on algorythm, *available at* https://www.bundestag.de/dokumente/textarchiv/2018/kw26-de-enquete-kommission-kuenstliche-intelligenz/560330

C.

THE UK

funding commitments and educational initiatives meant to grow the AI sector in the UK.[64] Lip service is paid to the notion that the UK could be a regulatory "innovator", with funding earmarked for the same, but there is little discussion of what that would mean.

The UK's approach to AI regulation can also be found in its response to the House of Lords report.[65] The UK government is establishing a ministerial Working Group on Future Regulation, a £10m Regulators' Pioneer Fund, and a Centre for Data Ethics and Innovation—an £8m body to serve as a sounding board for the development of ethical AI.

The UK also acknowledges the relevance of the Data Protection Act and GDPR to the regulation of AI and has pledged to follow the GDPR after Brexit.[66] The UK notes GDPR's provisions on automated processing but it does not currently plan to legislate to require businesses to tell the public on how and when AI is used to make decisions about them.

House of Lords AI report

The HOL published an extensive expert report, drawing on conversations and papers from a huge range of witnesses, that discussed how the UK could best support the development of AI technology and law.[67]

While the HOL paper does not explicitly use a "human rights" framework in its analysis of the risks and potential of AI technology, it does cover several key rights and data protection principles. In particular, the paper's ethical and legal focus is on transparency and explainability of AI systems, as well as the risk of prejudice. The HOL discussed, for example, how machine learning could become "money laundering for bias." The HOL proposed the creation of secure data trusts to support the implementation of the right to data portability; building new approaches to the auditing of datasets; and the holding of a summit for global AI norms.

The report suggests that the Competition and Markets Authority review the concentration of power in the hands of a small number of technology companies. Finally, the report concludes that "blanket AI-specific regulation" is inappropriate at this stage, and that regulation is best left to "existing sector-specific regulators."[68]

[64] *See* AI Sector Deal, *available at* https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal#key-commitments

[65] *See* Government response to AI House of Lord report, *available at* https://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/AI-Government-Response2.pdf

[66] At p. 7: "The Data Protection Act 2018 reflects the need to ensure there are stringent provisions in place to appropriately regulate automated processing. The Act includes the necessary safeguards such as the right to be informed of automated processing as soon as possible and also the right to challenge an automated decision made by a data controller or processor." Government response to AI House of Lord report, *available at* https://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/AI-Government-Response2.pdf

[67] *See* House of Lord paper, AI in the UK: ready, willing and able?, *available at* https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

[68] *See* p.116 of House of Lord paper, AI in the UK: ready, willing and able?, *available at* https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

C.

THE UK

🇬🇧

The Information Commissioner's Office (ICO)

The UK's Information Commissioner Elizabeth Denham has published several thoughtful papers on AI's intersection with data protection rights in the UK and the GDPR.

i. Report "big data, artificial intelligence, machine learning, and data protection"

The report on big data, artificial intelligence, machine learning, and data protection concerns the regulatory frameworks that should govern machine learning techniques.[69] It contains extensive discussions on anonymisation and purpose limitation, and advances the idea of baking in "auditability" to AI.

The main conclusions of the report are that organisations using AI should:

- ➔ anonymise data before mass analysis where possible;
- ➔ be transparent about their processing of personal data and offer provide "meaningful privacy notices";
- ➔ "embed a privacy impact assessment framework" into AI applications which "should involve input from all relevant parties including data analysts, compliance officers, board members and the public"
- ➔ "adopt a privacy by design approach" in developing AI, with a particular focus on "data security, data minimisation and data segregation;
- ➔ supplement data protection principles with ethical principles. Larger AI organisations "should create ethics boards to help scrutinise projects and assess complex issues" AI presents;
- ➔ innovate to create more "auditable machine learning algorithms," running internal and external audits "with a view to explaining the rationale behind algorithmic decisions and checking for bias, discrimination and errors."

"Democracy disrupted' - report on electoral manipulation and Facebook

> *"We are at risk of developing a system of voter surveillance by default. This could have a damaging long-term effect on the fabric of our democracy and political life."*[70]

The ICO's most recent report "Democracy Disrupted?",published in June 2018, analyses how AI-powered micro-targeting in campaigns have gotten ahead of citizens' understanding or ability to debate.[71] It called for an "ethical pause" on targeted political ads until society and regulators can catch up with technology.

The ICO also criticises the use of proxies for protected characteristics in EU law—e.g., ethnicity and age—to target or exclude people from political advertising.

[69] *See* Report on big data, artificial intelligence, machine learning, and data protection, *available at* https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

[70] P. 9. of Democracy Disrupted?", *available at* https://ico.org.uk/media/action-weve-taken/2259369/democracy-disrupted-110718.pdf

[71] *See* Democracy Disrupted?", *available at* https://ico.org.uk/media/action-weve-taken/2259369/democracy-disrupted-110718.pdf

D. THE NORDIC-BALTIC STATES

Overall assessment: The joint Nordic-Baltic statement on AI collaboration pledges to enhance access to data for AI, while developing ethical and transparent guidelines, standards, principles and values to guide how AI applications should be used.[72] The only explicit reference to regulation states is rather negative, as the signatories of the statement seek to "avoid unnecessary regulation" in order to keep pace with a fast-developing field. Countries do pledge to cooperate on "the objective that infrastructure, hardware, software and data, all of which are central to the use of AI, are based on standards, enabling interoperability, privacy, security, trust, good usability, and portability."

Main regulatory proposals: At this stage the joint proposal is only for the development of standards, and has yet to state what the standards should be.

It is preliminary to assess the compliance of this joint plan at such a high level of generality, but there is at least acknowledgement of the legal requirement for transparency as well as privacy and portability. Issues of bias and inclusion are not dealt with, nor due process or other human rights. In general, the statement tends to frame regulation as a burden on business, rather than something more nuanced and positive: a legal framework that protects users' rights and can open up new avenues for competition, innovation, and growth.

E. FINLAND 🇫🇮

Overall assessment: The regulatory conversation in Finland appears to be at an early stage. The Finnish government has established a high-level expert group on spurring AI in Finland, which is due to publish its final report in 2019.

Main regulatory proposals: The interim report on "Finland's Age of Artificial Intelligence" is mostly focused on growing Finland's domestic AI industry and says little about regulation other than that it "should be developed".[73] It nods to the need to protect privacy while developing AI, as well as the principles of transparency and accountability.

A second report on Work in the Age of Artificial Intelligence does contain a section on ethics.[74] It notes some principles that should guide the application of AI in the Finnish workplace, including transparency of decisions, responsibility (accountability) for the use of AI and the need to regulate liability.

[72] *See* Joint Nordic-Baltic statement, *available at* https://www.regeringen.se/49a602/globalassets/regeringen/dokument/naringsdepartementet/20180514_nmr_deklaration-slutlig-webb.pdf This was issued jointly by the governments of Denmark, Estonia, the Faeroes, Finland, Iceland, Latvia, Lithuania, Norway, Sweden, and the Aland islands.

[73] *See* Finland's Age of Artificial Intelligence, *available at* https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y

[74] *See* Work in the Age of Artificial Intelligence, *available at* http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160931/19_18_TEM_Tekoalyajan_tyo_WEB.pdf?sequence=1&isAllowed=y

F. DENMARK 🇩🇰

Overall assessment: Denmark has placed its AI policy in a wider plan for “digital growth”.[75] The country does not yet situate the conversation a regulatory context, other than to say that regulation of tech needs to be more agile.

Main regulatory proposals: The main regulatory proposals involve opening data and loosening regulation: in particular, government data. There is also reference to improvements in cybersecurity to protect mass data sets, as well as regulatory sandboxes. Denmark’s general digital strategy says that regulation in Denmark “needs to be more agile than it is in other countries in order to provide optimal support for new business models,” and refers to supporting the sharing economy, for example.[76]

G. ITALY 🇮🇹

Overall assessment: Italy’s main AI policy paper “Artificial Intelligence: At the Service of Citizens,” focuses on encouraging the public administration to take up AI technologies.[77]

Main regulatory proposals

In one sense this paper helps fill a gap in many other states’ conversations: it considers what principles should guide government bodies as they use AI. This paper contains an extensive discussion of the ethical and regulatory challenges inherent in rolling AI out across Italy’s public sector.

One of the main challenges it notes is to involve citizens in the transparent procurement of AI technologies. It discusses the rights of citizens to explanation when the public sector uses AI to make a consequential decision about them. It considers the need to address the liability and accountability frameworks for AI if robots cause harm, for example. It also notes government use of AI to predict citizens’ behavior, “from traffic management to crime prevention”, may compromise citizens’ right to privacy. It nods to the importance of the GDPR in this context. And it contains extended discussions on the need for public sector authorities to avoid discrimination in the use of AI, and recommendations how public authorities should test for bias in AI-driven decisions. The paper notes the possible benefits of AI to the administration of justice, but also echoes the general criticisms around the bias of the COMPAS system in the United States.

The paper notes several legal challenges with Italian public sector AI: transparency, determining standards for liability, privacy, information security and intellectual property. The main strategy for ensuring regulatory and rights compliance suggested in the paper is to test any proposed public sector use of AI at a small scale before rolling it out, to be sure that the issues of “data protection and privacy, ethical

[75] *See* Digital growth strategy, *available at* https://em.dk/english/~/media/files/2018/digital-growth-strategy-report_uk_web.ashx?la=en
[76] Digital growth strategy, *available at* https://em.dk/english/~/media/files/2018/digital-growth-strategy-report_uk_web.ashx?la=en p. 7.

G. ITALY 🇮🇹

dilemmas, the risk of bias" are minimised, and "exposing data and algorithms in a transparent and replicable manner."

Some examples given of current use of AI/Machine Learning driven systems raise questions. For example, the paper refers to the Italian government's use of an EU-funded machine learning - data mining counterterrorism tool called DANTE (Detecting and Analysing Terrorists) "to trace terror networks."[78] Little is said about how this tool works but it raises potential risks of profiling and, depending on how it is used by law enforcement, due process.[79]

H. SPAIN 🇪🇸

Overall assessment : The AI national strategy in Spain appears to be at an early stage. The Spanish government established at the end of 2018 a high-level expert group on AI and Big Data in Spain, which is due to publish a report in late 2018 or early 2019.[80] The publication has been delayed given the change of Government in Spain on June 1st. The Spanish Government organised a preliminary conference on May 31st 2018 where many of the members of the high-level expert group, in addition to other international experts, discussed the main pillars that will be included in the Spanish strategic document.[81] [82]

Main regulatory proposals

The Spanish strategic document will discuss AI technologies, ethical implications of AI systems, existing legal frameworks and AI, key economic areas that AI could enable in Spain, AI and Big Data to empower society, the impact of AI on the labor market and the need to invest in nurturing, attracting and retaining talent.

[77] *See* Artificial Intelligence: At the Service of Citizens, *available at* https://ia.italia.it/en/assets/whitepaper.pdf

[78] P. 50 of Artificial Intelligence: At the Service of Citizens, *available at* https://ia.italia.it/en/assets/whitepaper.pdf

[79] Digital growth strategy, *available at* https://em.dk/english/~/media/files/2018/digital-growth-strategy-report_uk_web.ashx?la=en p. 7.

[80] *See* Red, Constituido el grupo de sabios sobre la inteligencia artificial y el big data, *available at* http://www.red.es/redes/es/actualidad/magazin-en-red/constituido-el-grupo-de-sabios-sobre-inteligencia-artificial-y-big-data

[81] *See* Red, Primer conversatorio sobre inteligencia artificial, *available at* http://www.red.es/redes/es/actualidad/magazin-en-red/el-primer-conversatorio-sobre-inteligencia-artificial-pone-sobre-la-mesa

[82] *See* Red, Alcala de Heneras acoger el primer conversatorio sobre inteligenci artificial, *available at* http://www.red.es/redes/es/actualidad/magazin-en-red/alcal%C3%A1-de-henares-acoge-el-primer-%E2%80%98conversatorio-sobre-inteligencia

D. COMPARATIVE ANALYSIS OF THE AI PROPOSALS AND THEIR IMPACT ON HUMAN RIGHTS

A. Criteria for Assessment _

To carry out the assessment of the strategies mapped in this report, it was necessary to identify a set of principles and human rights that are most relevant in the context of development, deployment and use of AI to be able to benchmark and compare these documents to. We developed the list of criteria below based on the principles and rights explicitly mentioned in the strategies on the one hand, and the most widely acknowledged relevant issues that are impacted by AI on the other.[83]

In the annexed chart (see Section F), we give a broad overview of where the AI strategies stand against this scale that serves as a basis for the below detailed analysis.

1. Transparency: People and societies should be meaningfully informed when a decision has been made about them that involves AI. This goes beyond the outdated model of impenetrable terms of services or other privacy notices. In addition, people should receive clear, meaningful explanation of the technology used and how it arrived at its decision or prediction from the input to the processing and output. This may require additional work to develop AI/Machine learning models that are more explainable and auditable. To the extent there is a technical trade-off in accuracy and explainability, the balance in consequential areas of public life should generally be in favour of the explainable one.

2. Accountability: Any given algorithm—and the entities, public or private, who design, develop and deploy it—must be accountable to people and society given the impact that the processing of a large volume of personal data have on users' rights. That includes providing access to effective and meaningful remedy and redress when harm occurs. This is in part a products liability question and is being actively developed in the context of self-driving cars. But well before that point there is a need for wider user and social control of AI. Citizens should also be included in public debates about procurement of AI systems: where the use is socially significant, such as the use of facial recognition software by the police, states should assist meaningful public debate about whether the algorithmic tool is a desirable or acceptable use of AI at all.

[83] *See* Artificial Intelligence and Human Rights, Berkman Klein Center, 2018, *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3259344

3. The Right to Privacy: AI systems must respect users' rights to privacy and autonomy. The need for large training data sets for machine learning tools must not be permitted to override users' genuine desire and actionable right for privacy. Pervasive surveillance of users' online behaviour by private and public actors has been shown not only unlawful but also to have corrosive effects on public life and on trust in the digital ecosystem. Studies show that individuals wish their right to privacy to be respected better than the internet platforms allow - they just think they have no choice.[84]

4. Freedom of Conscience and Expression: Neither states nor private entities should deploy AI systems to limit free expression and opinion. AI-powered content monitoring on internet platforms, for example, will need to be carefully scrutinised for these effects in addition to its privacy implications. But the harms can be more subtle: where AI systems interfere with privacy, that tends to have a ripple effect on users' rights to follow their conscience or express themselves.

5. The Right to Equality and Non-Discrimination: AI systems cannot be permitted to produce social outcomes that are biased, amplify existing human bias, or which give pretext for biased decision-making. All AI systems will need rigorous audit to show compliance with the right to non-discrimination. Where AI is used to profile people in ways that are, on analysis, a proxy for race or other protected characteristics, that should be assessed with the same strict scrutiny as for race itself. Public and private sector actors must uphold their obligations and responsibilities under human rights laws and standards to avoid and prevent discrimination in the use of machine learning systems where possible. Where discrimination arises, measures to deliver the right to effective remedy must be in place.[85]

6. Due Process: AI must never be permitted to undermine individuals' right to a fair trial—particularly in the criminal justice, immigration, or national security context.

[84] *See, e.g.*, the 2018 Digital Attitudes survey conducted by UK think tank Doteveryone: http://attitudes.doteveryone.org.uk./

[85] *See* The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems, https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf.

7. The Right to Data Protection and User Control: This is an additional but separate right to privacy, protected under the EU Charter - citizens must have control over their personal data and should not have to relinquish it to use internet services. The data protection principles protected under the GDPR (purpose limitation; minimisation; accuracy; storage limitation and more) all add additional layers of protection to users. This is an area where AI applications will challenge the enjoyment of the right, and states will need to give the issue greater attention.

8. Collective Rights (Free **Press**, Free and Fair **Elections**): some essential rights in society are not best understood through the individual human rights framework but are collective. Protecting these collective rights is likely to require concerted thought, over and beyond frameworks like data protection.

9. Economic Rights and the Future of Work: The economic gains of AI should not come at the cost of society's poorest or most vulnerable. AI also has potential costs to economic and social rights—including potential for mass job displacement, or creating disparities in access to social services or financial products. Without foresight, inequality could deepen. This, in turn, would have ripple effects on our social fabric and political life.

10. The Laws of War: An in-depth analysis of the Convention on Certain Conventional Weapons and the debate over autonomous weapons systems is beyond the scope of this paper. We note, however, that despite the CCW's position that weapons systems should be subject to "meaningful human control," major member states have not ruled out autonomous weapons systems. This is a clear lacuna—either in states' regulatory thinking, or in their public engagement.

B. The European Way to AI: harmonised approaches or diverging routes?

1. Similarities and Differences between all the AI strategies

Diverse as these strategies are, they reflect several common patterns of thinking about AI regulation.

Ethics as a substitute for law, or ethics as a foundation for law?

"Ethics in artificial intelligence" – this thread runs through most of the AI strategies. The EU

Commission's Communication on AI has trailed "ethics guidelines" that will be published in early 2019. The French strategy says it will "establish an ethical framework" for AI; the UK's major initiative is to set up a "Centre for Data Ethics and Innovation;" ethics is the main limiting principle in Finland's paper on the future of work, and so on.

Most of the papers exhibit a marked preference for the "ethical" dimensions of the development of AI, but have considerably less to say about how the existing, or potential future, hard laws of the EU or of any member state may influence or shape AI development.

There may be several reasons for this. Some entities, of course, may stress ethics simply to avoid thorny issues of regulation. Some of the papers caution against the use of ethics in this way: the Council of Europe, for example, notes that reference to ethics in this field may reflect "a tactical move by some actors who want to avoid strict regulation" and prefer soft norms to hard ones.[86] Equally, however, the Council of Europe acknowledge that it may simply reflect a need for "deeper reflection" to balance the different, sometimes competing, norms that AI will bring into play. The Villani paper suggests that while legislation must control AI, ethics fill the space between what is legally permissible and what the technology allows, particularly given how slow the law may be to catch up with the pace of technological change.[87]

The data protection authorities note that ethics, done properly, can support the existing legal framework. Indeed, the EDPS commissioned the EAG paper on the ethical challenges of AI precisely because "data protection authorities now face ethical questions that legal analysis alone cannot address."[88] The EAG insist that the purpose of ethical analysis is not, and should not be, to weaken legal doctrine or "to fill regulatory gaps in data protection law with more flexible, and thus less easily enforceable ethical rules."

The experts we consulted for this report expressed a variety of opinions about the meaning and value of "ethics" in this context. Some saw the extent of corporate engagement with these issues as a positive sign, and an unusual move by business to get things right early in the development of these technologies. Others voiced more skepticism, saying that "ethics" was an ill-defined term and created a meaningless meeting ground for people with highly diverse perspectives. Most agreed that whether "ethics" was a starting point of discussion between regulators, citizens, and companies, or a barrier to discussion, would vary across contexts and actors. Some pointed to the history of bioethics in medicine - including the use of ethics boards for significant medical research projects - as an example of how ethics could usefully be built into AI research and development.

The strategies' tendency to lean on ethics in lieu of law is not absolute. The EU Commission tempers its comments on ethics by calling on public bodies to "ensure that the regulatory frameworks for developing and using of AI technologies are in line with…fundamental rights." At the moment, however, the EU Commission simply proposes to "monitor developments" and review laws "if necessary." Similarly, the UK pledges to become a centre for "regulatory innovation"—but has yet to determine which areas or AI technology, or which applications, need regulatory focus, and says little

[86] *See* Council of Europe study at p. 42, *available at* https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10

[87] *See* Villani paper at p. 113: ethics "occupy the available space between what has been made possible by AI and what is permitted by law." and p. 120: Ethical training for computer scientists, the paper adds, might be *more* protective than a minimal standard of following the law: "The aim of teaching ethics is rather to pass on to the future architects of a digital society the conceptual tools they will need to be able to identify and confront the moral issues they will encounter—within the context of their professional activities—in a responsible fashion."https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

[88] *See* EDPS EAG paper, Foreword at p. 1, *available at* https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf

about human rights in any public statement on AI. More concerning are authorities who, while paying a nod to "ethics", mainly express willingness to loosen the regulatory environment such as the Denmark digital strategy.

So, in short, authorities should be vigilant that ethics does not become a smokescreen for an unregulated technical environment. As the data protection authorities have observed, many companies developing this technology have avoided meaningful regulation for many years:

> *[F]or two decades tech has been largely outside the reach of regulation. They have been allowed to move fast and break things. Now there are a lot of broken things which need to be mended, and we need to guard against future breakages.*[89]

The essential question will be: are regulators and legislators able to keep up with the pace of technological change, and will state authorities give them the resources, and logistical and political support they need, to manage the safe, equitable, legal development of AI? It cannot be a question of trading human rights law for ethics. Instead, the added value of ethics comes on top of legal obligations and minimum binding requirements, including international human rights norms. It can serve as guidelines for both the private and public sector to aspire to higher standards to truly achieve the concept of AI for humanity.

AI-specific standards: is it better to "wait and see"?

Underpinning this preference for ethics appears to be a sense among states and the community that it is too soon to codify AI-specific regulation.

There are several apparent reasons for this: one is the repeated concern not to detract investors and stifle innovation. This motivates, for example, the Danish emphasis on "agile regulation," and the Finnish reference to a "regulatory sandbox," which explicitly states that the emphasis must be on supporting business and not on data protection: "A clear legislative framework that will ensure the availability of data must be created. This must be based on the importance of the data to business operations (not on data protection first)."[90] This is echoed by the EU Communication on AI, which also supports the idea of sandboxes to test AI applications, including in safety-critical areas like transport and healthcare.[91]

Another reason is uncertainty: the technology is so new that there is a sense that hasty legislative change will be swiftly left behind. This theme emerged, for example, from many witnesses to the UK House of Lords' study on AI, including those from civil society and academia.[92] The HOL questioned many witnesses about regulatory proposals, but "Few …gave any clear sense to us as to what specific regulation should be considered." Similarly, the Finnish strategy sets out the core essentials of what regulation would require—"some type of vision of what is a good artificial intelligence society"—but without reference to the existing human rights norms that have guided European notions of a good society for decades.[93]

[89] EDPS speech to the Telecommunications Forum at 3, *available at* https://edps.europa.eu/sites/edp/files/publication/18-04-24_giovanni_buttarelli_keynote_speech_telecoms_forum_en.pdf)

[90] *See* Finnish strategy at p. 4, *available at* https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y

[91] *See,* EU Commission Communication on AI at p. 10, *available at* http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625

[92] *See* House of Lords discussion at p. 113-114, *available at* https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

[93] *See* Finish interim report at p. 40, *available at* https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y

This is also, perhaps, a hope that some of the key challenges may prove to have *technical* solutions that make regulatory change unnecessary. This may be why many of the proposed approaches to the signal challenges with AI, such as the explainability of neural networks, involve the funneling of money into research—policymakers and computer scientists hope that a technical solution could be found. References to technical solutions are found in the UK and French strategies, as well as the EU Commission's Communication.

Finally, it may be that existing human rights and data protection principles—if backed up by enforcement resources and a will to accountability by public bodies—are already effective tools to manage AI for the public good. The GDPR is itself so young that its effect on the shape of AI is as yet untested, and data protection authorities themselves have said they prefer at this stage to use the "carrot" – of advice and assistance to comply – before turning to the "stick" of fines.[94]

Most published strategies contain at least a nod of recognition of many crucial areas where AI will implicate human rights – in particular transparency, accountability, privacy, and the future of work. They acknowledge the central importance of the GDPR in regulating AI, without saying precisely how it will. Some also point to the proposed ePrivacy Regulation as an additional layer of protection: reference is made in the German cornerstone strategy, for example, and by all the data protection authorities.[95] [96] The EU Commission, for its part, urges the adoption of this reform "as soon as possible."[97]

On the one hand, a cautious attitude about regulation has some merit, in particular if technical solutions are found. On the other, if AI applications are being tested and used in crucial areas of public life already, with potentially serious consequences for citizens' lives—employment, school systems, parole, credit decisions—can it truly be too soon to consider hard legal constraints? The European Data Protection Supervisor notes that the major internet companies, for most of their existence, have operated with minimal regulatory intervention and that because of that "Now there are a lot of broken things which need to be mended, and we need to guard against future breakages.".[98] Similarly, the Villani paper rightly points out that ethical restraints cannot simply be tacked on at the end. As with network and infrastructure security, Big Data, the Internet of Things and many more technological development, to wait until the technology is built and in use and to add law or ethics in "as an afterthought" will be too late.[99]

Regulatory challenges in context

The experts we consulted voiced different views about the need to regulate the use of AI. They shared a sense that it was impossible to assess the need to "regulate AI" in the abstract—that it was so pervasive and diverse a technology that its necessary contours and scope could only usefully be

[94] *See, e.g.*, interview with UK ICO Elizabeth Denham, *available at* https://www.computing.co.uk/ctg/news/3027593/ico-theres-so-much-misinformation-out-there-on-gdpr

[95] *See* German cornerstone document at 4, *available at* https://www.bmbf.de/files/180718%20Eckpunkte_KI-Strategie%20final%20Layout.pdf

[96] *See, e.g.,* EDPS press release, "Going Beyond the GDPR," *available at* https://edps.europa.eu/sites/edp/files/edpsweb_press_releases/edps-2018-02-annual_report_2017_en.pdf

[97] *See,* EU Commission Communication on AI at p. 15, *available at* http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625

[98] *See* EDPS speech to EU Telecommunications forum: "Older sectors like telecoms, broadcast and print media appreciate the historical imbalances in regulation of sectors. But for two decades tech has been largely outside the reach of regulation. They have been allowed to move fast and break things. Now there are a lot of broken things which need to be mended, and we need to guard against future breakages." *available at* https://edps.europa.eu/sites/edp/files/publication/18-04-24_giovanni_buttarelli_keynote_speech_telecoms_forum_en.pdf

[99] *See* Villani paper at p. 43, *available at* .https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

discussed by focusing on specific sectors, specific social opportunities, trade-offs, or risks posed by AI.[100]

We shared this sense as we undertook the mapping: that much of the ethics and regulatory conversation is taking place at too high a level of generality, while at the same time public authorities are acquiring specific technologies and private companies are using AI for a particular purpose. There needs to be more engagement with the uses of AI *already underway* to have a colorful and meaningful regulatory debate. To that end, we have included four hypothetical examples that elucidate some of the dilemmas facing European regulators.

EXAMPLE 1: POLICING AND CRIMINAL JUSTICE

Hypothetical: The head of state of a Middle Eastern nation is coming to London. The country, a major ally of the United Kingdom, has been widely criticised for its human rights record and, in particular, for its role in a war in the Arabian Peninsula that has killed thousands of civilians. Campaigners organise a protest on social media. Some, but not all, of the organisers have previously been convicted of minor offenses in connection with protests. A high percentage of the protestors are Muslim. As campaigners arrive in the neighborhood where the protest is to be held, facial recognition and other AI-powered tools flag the protesters' identities and location with the police. Automated network analysis identifies other likely protestors. Several dozen people, including those with no prior convictions, are arrested. They cannot attend the protest. Police release them without charge after six to twelve hours.

Increasing numbers of police forces are seeking to use mass data tools to detect and prevent crime. Some of these tools seek to anticipate where a crime will occur (the best-known purveyor of this software is PredPol). Others claim to assess the risk a given individual may pose (for, e.g., absconding from bail, or reoffending).

The use automated tools for law enforcement activities is already a reality in a few countries. After the terror attacks of July 2016, the French city of Nice hired defense contractor Thalès to develop a suite of predictive policing tools.[101] Gerard Collomb, the former French Interior Minister, drew criticism from French privacy groups by his support for high-tech policing.[102] In 2017 UK police trialled facial recognition software at Notting Hill Carnival—but discontinued the trial in 2018 in the face of public resistance.[103] Durham Constabulary used a bail assessment software that included in its risk variables an individual's post code (closely linked with economic status).[104] In most instances the police have released limited information about their use of these tools.

Can algorithmic tools, correctly calibrated, make policing more neutral across class and race – or do they inevitably risk amplifying existing biases? If a community is historically over-policed, will AI-powered tools ease or exacerbate a climate of mistrust? How is the impact of the use of such technology on freedom of assembly, the rights to privacy and

[100] *See also* Artificial Intelligence and Human Rights, Berkman Klein Center, 2018, *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3259344

[101] *See* In France, Smart City Policing Is Spreading Like Wildfire, *available at* https://www.laquadrature.net/en/smartcity_policing_like_wildfire

[102] *See* Intervention Gerard Collomb, *available at* https://www.interieur.gouv.fr/Archives/Archives-ministre-de-l-interieur/Archives-Gerard-Collomb-mai-2017-octobre-2018/Interventions-du-ministre/Adaptation-de-la-doctrine-d-emploi-au-maintien-de-l-ordre

[103] *See* The Guardian, Police use facial recognition software at Notting Hill Carnival, *available at* https://www.theguardian.com/uk-news/2017/aug/05/met-police-facial-recognition-software-notting-hill-carnival

[104] *See* Wired, Police use of algorithm, *available at* https://www.wired.co.uk/article/police-ai-uk-durham-hart-checkpoint-algorithm-edit

data protection assessed? Who is responsible for possible violations of laws linked to the use of these technologies: the public authorities using it or the private entities building it? Questions like these caused community activists and officials in the city of Oakland to pass an ordinance: police are now required to take any proposed acquisition of surveillance technology (including AI-powered technology) to a municipal privacy board for approval.[105]

EXAMPLE 2: HEALTH AND THE PUBLIC SECTOR

Hypothetical: A major artificial intelligence company contracts with a state health provider to improve the diagnosis of age-related macular degeneration through AI-powered analysis of historical eye scans. The deal between the state and the company entails, in essence, vast quantities of data being sold to the company in exchange for its expertise. The company then use the data to train and improve their diagnostic software. The state health service seeks to anonymise the data – names, addresses, and other obviously identifying details are removed – but it later turns out de-anonymisation in a number of cases is technically possible. The contract is entered into without public debate nor review by a state ethics board.

This is a loosely modified version of what happened with DeepMind's partnership with the Royal Free Hospital trust. In this case, an app in development to identify acute kidney injury was developed and tested with the data of 1.6 million NHS patients. The UK's ICO found that the Royal Free Trust had "failed to comply with data protection law" in the process by which it assigned the data to DeepMind.[106] Other reporting criticised DeepMind for planning to train its AI on this public sector health data without making this aim explicit when announcing the project.

Broadly, experts we consulted said that the use of artificial intelligence in the health sector is one of the most promising areas for AI—and one which raises deep ethical and technical issues. AI in health presents multiple challenges for policymakers: what level of consent is required and desirable, how to preserve patient confidentiality (studies tend to show that data can be de-anonymised with a surprisingly small number of data points), and whether governments will extract the true value of a significant asset—years of patient data—from the private sector.

EXAMPLE 3: NAVIGATING AND URBAN PLANNING

Hypothetical: Early one Saturday evening, you are due to attend a party at a friend's house across town. You hail a popular ride-sharing service to get there, and your driver uses an automated map application to navigate to the party. Traffic is heavy this afternoon, so the map software routes your car off arterial roads and onto a series of residential side streets. The route turns out to be surprisingly busy, and you notice half a dozen other cars that seem to have been directed through the same residential roads. The speed limit is 20 mph; all these drivers average between 25 and 30 mph as they transit the neighborhood. Three kids playing soccer in the street pause as you pass.

Many of us will have used an automated map software to plan our journey. Some of the experts we consulted raised the ripple effects of this routing software (such as Waze, Google Maps) on local communities as a cost of AI. Some extreme examples have been reported in the press. One

[105] *See* Oakland city council ordinance, *available at* https://www.documentcloud.org/documents/4450176-View-Supplemental-Report-4-26-18.html

[106] *See* ICO letter on provision of data to DeepMind, *available at* https://ico.org.uk/media/2014353/undertaking-cover-letter-revised-04072017-to-first-person.pdf

city in New Jersey had to ban local roads to non-residents after thousands of commuters were choking residents' side streets.[107] The local police chief said the town "had days when people can't get out of their driveways." In another city in Maryland, frustrated residents resorted to reporting (fake) car crashes in their area to deter navigation software.[108]

The use of AI for urban planning may come at cost for certains communities. Previously quiet residential area might become alternative busy route during peak traffic hours, increasing pollutions and risks of accidents. How will government determine the "cost" of new urban planning and the impact it may have on communities? How will these communities be engaged in these discussions? A time-efficient solution might end-up not being the most desirable for the quality of life of resident, what factors should be considered to mitigate these impacts?

EXAMPLE 4: EMPLOYMENT

Hypothetical: Your 19-year-old cousin is hunting for a job in retail to help pay for his university fees. He has applied for multiple jobs, and as part of the screening process he is regularly asked to answer questions online using a chatbot and webcam portal. He is not selected for interview, receiving rejections (often instant; at other times delayed by a few hours) saying that an automated assessment software analysing his interview has determined he was not fit for the job. The rejections contain no further detail. Your cousin has a solid academic record and is otherwise qualified for these posts but struggles with social anxiety and finds the process of speaking to the camera uncomfortable.

Employers – particularly large retailers such as supermarkets – are now using automated scoring of webcam interviews to sift applicants, particularly for entry-level retail jobs.[109] A new industry has sprung up purporting to assess applicants for high-volume, lower-skilled jobs. Critics are concerned that these systems will be biased against, for example, qualified applicants with a disability (or simply those who do not look like the current employee pool). In turn, another cottage industry has sprung up to help applicants game these assessment systems, through tricks such as putting "Oxford" or "Cambridge" in white text on a CV. Because of the intersection with employment and labour rights, the use of automated hiring software is an area ripe for policy attention. How to guarantee the right to non-discrimination in automated processes? Should applicant be able to ask for the intervention of a human in the process and at what stage? These are a few of the questions that governments should look into when assessing the use of automated processes in the public and private job market.

[107] *See* The New York Times, Use of traffic apps, *available at* https://www.nytimes.com/2017/12/24/nyregion/traffic-apps-gps-neighborhoods.html

[108] *See* The Washington Post, Traffic weary homeowners, *available at* https://www.washingtonpost.com/local/traffic-weary-homeowners-and-waze-are-at-war-again-guess-whos-winning/2016/06/05/c466df46-299d-11e6-b989-4e5479715b54_story.html

[109] *See* The Guardian, Dehumanising, impenetrable, frustrating: the grim reality of job hunting in the age of AI, *available at* https://www.theguardian.com/inequality/2018/mar/04/dehumanising-impenetrable-frustrating-the-grim-reality-of-job-hunting-in-the-age-of-ai and CNBC on AI and recruitment, *available at* https://www.cnbc.com/2018/03/13/ai-job-recruiting-tools-offered-by-hirevue-mya-other-start-ups.html

DPAs and national decision-makers: speaking at cross-purposes?

Another theme running through the papers is a persistent gap in understanding between the governments developing AI strategies—who are keenly focused on attracting talent for sound economic reasons—and the data protection authorities, who are sensitive to the ways in which internet platforms have built an economic model based on tracking and monetisation of data.

A number of experts we spoke to echoed this view. They stressed that - as expert regulators with both the ability and the statutory duty to police and enforce the laws relating to data practices - the data protection authorities' thoughts on AI needed to be better incorporated into national strategies. The DPAs themselves are actively approaching the private sector for their assistance in developing regulation; there is ample room for national authorities to listen to and learn from DPAs in this regard. [110]

One question missing from most of the analyses in the national strategies, are the extent to which most AI applications are likely to constitute "high risk" processing. To take just two examples:

- ➔ **"Evaluation or scoring" - e.g., predictions and profiling:** This, of course, is a major AI application, from health to credit to policing. The data protection authorities express deep concerns about how to square these predictive uses of AI with the principle of human dignity. EDPS has expressed a concern that under AI we risk "a 'dictatorship of data' where 'we are no longer judged on the basis of our actions, but on the basis of what all the data about us indicate our probable actions may be'."[111]
- ➔ **"Sensitive data" or data of a highly personal nature**: the vast quantities of data currently harvested about users as they go about their online lives, including cookies which track them across websites to build extremely detailed pictures of their browsing and purchasing habits – are highly personal. It recently surfaced that Google paired its search data with actual credit card purchase information given to them by Mastercard—in order to sell more tailored ad targeting to ad buyers.[112] Google claims that the material was anonymised, and only given to the advertisers in the aggregate, but the fact remains that google itself, as the data controller, suddenly owned not simply information about what a given user did on its search engine, but potentially what was spent at a physical store.

The GDPR will require data protection impact assessments for these sorts of AI applications: is that sufficient? The data protection authorities engage these problems in detail. None of the national strategies do. Important as the GDPR is, it will be vital for governments to engage more with DPAs - and to resource them properly - if they are to enforce the law, have a realistic prospect of assessing, and regulating, the large internet companies who are likely to have a central role in developing AI. And this becomes even more essential as new regulatory challenges arise in the area of big tech.

Large companies are not inherently more likely to violate laws or ethics. Practically, however, the consequences when they do are more significant. Moreover, given AI's need for massive sets of

[110] *See, e.g.*, the ICO's public consultation on developing a regulatory sandbox, *available at*. https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2018/09/ico-call-for-views-on-creating-a-regulatory-sandbox/

[111] *See* EDPS 2015 big data paper p. 8, *available at* https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf

[112] Bloomberg, Google and Mastercard Cut a Secret Ad Deal to Track Retail Sales, *available at* https://www.bloomberg.com/news/articles/2018-08-30/google-and-mastercard-cut-a-secret-ad-deal-to-track-retail-sales.

training data, the small number of dominant market players who have long enjoyed unfettered (and largely unregulated) access to vast personal data streams may be the ones who develop and implement many consequential AI applications. And those companies are moving into new areas, including politically and democratically sensitive ones, such as Google's recent extension into military contracting, or Amazon's sale of its facial recognition software to law enforcement authorities. Empowering DPAs to exercise their powers and ensuring that they are involved in debate around the regulation of technology is of paramount importance.

Differences in focus—different strengths and weaknesses, as well as identified gaps in human rights impacts

A separate chart in annex maps the extent to which a given national strategy engages with each of these rights; this section focuses on areas where a given country is particularly strong or where states exhibit a sharp gap in coverage.

1. Transparency: All the national strategies (save Denmark's) assert that AI should be developed in a transparent and explainable way. However, many national strategies pin their hopes, in the main, on identifying a *technical* solution to the problems of explainable AI over a *regulatory* solution.[113] This may be because the technical challenges to meaningful transparency of AI systems are real: both data protection authorities and the Council of Europe acknowledge them.[114] This is not to say a rights analysis is totally absent: the Council of Europe, Italy, and the UK ICO and House of Lords have all cited the "right to an explanation" as a crucial part of upholding transparency. Indeed, all analysed EU DPA papers note the more stringent transparency requirements in the GDPR, and the EDPS has gone so far as to call for an end to covert profiling altogether.

But most of national strategies are silent about specific regulatory requirements for transparency. The EU Commission's Communication hails the importance of the GDPR as "a major step for building trust, essential in the long term for both people and companies," but very little about how this is to be upheld in the context of AI as discussions on how to interpret certain provisions of the law continue.[115]

Finally, there is a different sort of transparency that states should engage in: transparency around AI policymaking. This is more positive. Here the UK has taken extensive public evidence about the dilemmas that AI presents, and published extensive expert reports on these issues. The quality of information the UK has placed online is arguably a benchmark for engaging the public about AI policy. The CNIL for France's extensive public consultation on its AI ethics paper is a similar exemple.

[113] The EU Commission, France, and the UK all focus on the need for research into more explainable AI models. Technical transparency is a particular focus of the Villani paper and of the French strategy. "Opening the black box," as it is put, is heavily stressed in the French paper as an important lynchpin for ensuring AI systems are more accountable.

[114] At the current state of the technology, the Council of Europe posit that it may be easier to "ensure critical engagement" about AI than to change algorithms. However, the Council of Europe also propose various audit standards short of full explainability: disclosure of variables, the goals for which the algorithm is optimised, information about the training data.

[115] We refer here to the extensive academic debate on the applicability - both in theory and in practice - of the right to an explanation laid down under the GDPR. See "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation", S. Watcher; Brent Mittelstadt; and Luciano Floridi, International Data Privacy Law, 2017, *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469 ; "Meaningful Information and the Right to Explanation", A. D. Selbst; and J. Powles, 2018, *available at* http://proceedings.mlr.press/v81/selbst18a/selbst18a.pdf ; and "Slave to the algorithm? Why a 'Right to an explanation' is probably not the remedy you are looking for", L. Edwards; and M. Veale, 2017, *available at* https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1315&context=dltr

Given the importance of these issues and the scale of the transformation in prospect, public bodies should engage the public in these decisions to the fullest possible extent.

Principal gaps: How to square the GDPR's robust transparency requirements with the technical issues with AI is a major gap (and best articulated by the EDPS EAG's ethics paper.) But the existence of technical challenges does not mean there is nothing governments and companies can and should do to inform citizens better about AI. As the Council of Europe and data protection authorities urge, the purposes of the processing, the nature and source of training data, and input-output auditing are all ways of checking algorithms that do not depend on further technical innovation. Member states should do more to support these initiatives.

There are also major exceptions to the data access and transparency rights in the Police Directive. This is understandable, given the need to maintain the confidentiality of law enforcement investigations. But given the controversies around law enforcement use of AI, there ought to be a mechanism for greater public scrutiny and debate around the use of AI technologies in this context. As discussed below, criminal justice and due process is a major gap in the national AI strategies. Finally, more could be said about making all AI-powered systems *legible* to people: individuals ought to know when an automated process is being used to make or influence a decision about them. The entity applying automated processes must have clear notice schemes to inform anyone who is subject to such processes.

2. Accountability: Accountability appears in most national strategies, but many of the references focus on a question of products liability: what would happen if an AI-powered vehicle injured or killed someone, for example.

Germany has done the most extensive work on developing norms and standards for this crucial area of AI accountability – self-driving cars. Their AI cornerstone document does refer to wider principles, using the language of greater user control and traceability—both important prerequisites for accountability.

Products liability and private law claims are an important component of any regulatory regime, but only form part of the picture. If an employer uses an AI-sifting software to hire employees that, in practice, has a disparate impact on women or people of colour, holding the employer accountable may require the use of discrimination law or state regulatory enforcement. And state bodies themselves, of course, will also need to be held accountable for its uses of AI: democracies will need to consider whether there are areas where, for public policy reasons, the use of AI should be curtailed or off limits. These are questions of public law, which go far beyond products liability.

Wider theoretical work on questions of accountability is to be found in the Villani paper for France, which has an extended discussion of the importance of accountability for AI systems and says this is "one of the conditions of its social acceptability."[116]

Data protection authorities have also published extensively on the question of accountability, as it is one of the new, formalised and strengthened principle in the GDPR.[117] Some other bodies, like the UK

[116] *See* p. 115 -116 of the Report of the Villani mission, *available at* https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf However, the national strategy (which the Villani paper informs) does not discuss accountability.

[117] *See, e.g.,* the ICO's paper on big data and algorithms https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

House of Lords, note that legal changes may be necessary to the current allocation of liability.[118] The Council of Europe sees the technical challenges to explainability as potentially a major hurdle to achieving meaningful accountability of AI systems, and propose sector-specific regulation.

Principal gaps: Whether for technical or other reasons, this is one of the areas where the conversation of the data protection authorities—who have sought to grapple with the new accountability principle in the GDPR—seems ahead of the rudimentary conversations in the national strategies. Effective enforcement and upholding this principle in the GDPR cannot be done by DPAs alone: it will require resourcing and support by national executives if it is to function properly.

The national strategies say very little about whether private companies or public authorities will be held to similar or different standards where AI is concerned. One problem with early uses of algorithmic systems, such as PredPol, has been that trade secrets or intellectual property have been used to stymie the public accountability that is citizens' right. While public bodies generally have higher duties to their citizens, given the power and dominance of a small number of internet platforms (and the likelihood that this dominance will persist in the AI area), states and regulators should pay greater attention to the compliance of the private sector with these rights as well. It will be crucial to further analyse the impact of AI developed by public-private partnerships. As often with the development of new technologies, private entities are likely be innovation leader, and thus develop technical standards and dictate deployment of AI. This means that potentially most of the AI technologies that will be used could be have been developed, fully or partly, by the private sector, even if these are will be used by public authorities. Which is why *both* public and private entities shall abide the highest standards of transparency and accountability in the design, deployment and use of AI.

3. Right to Privacy: A few strategies refer to privacy tools or point to ongoing legislative debate that could have an impact on the protection of this right in the context of AI. For instance, both the EU Commission and the data protection authorities refers to the proposed ePrivacy Regulation as a potential bulwark against invasive uses of AI, and Germany recognises its importance. Others approach the privacy question from a different angle: the French propose a "privacy impact assessment." Meanwhile, the House of Lords propose "data trusts" where citizens have a safe, central repository of data (although how this could technically be achieved is not discussed in detail.) The UK authorities are concerned to unlock the potential commercial value of public sector data, but at least note in passing that this must be balanced against users' privacy rights.

The UK ICO have published an extensive discussion of how to protect privacy in the "big data" context, including layered privacy notices; wifi location analytics; and privacy impact assessments. It shall be noted that only the Council of Europe refers to Article 8 of the European Convention on Human Rights (ECHR) on the right to private and family life in any detail. Other papers mention the right to privacy in passing.

Principal gaps: More analysis should be conducted as to how AI systems can be designed and developed in compliance with Article 7 of the EU Charter and Article 8 of the ECHR, possibly analogising to existing European jurisprudence on surveillance and mass data collection practices by states. Again, the gap between the detailed analysis done by the DPAs on the GDPR in this context and the national strategies is striking.

[118] *See* p. 97-98 of UK House of Lord paper on AI, *available at* https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf The paper recommend the law commission of the UK to investigate whether changes in the law are necessary.

4. Freedom of Conscience and Expression: Only the Council of Europe and some data protection authorities have published much thinking on this. The EDPS notes the potential chilling effect of pervasive internet tracking on freedom of expression and also on the ability of society to innovate.[119] The Council of Europe discuss this in some detail, nothing the "gatekeeper" function of dominant search engines. They also note that Google and Facebook, serving as a "quasi-public sphere", engage in problematic conduct when they filter speech by algorithm, raising concerns with Article 10 of ECHR compliance.

Principal gaps: Few national authorities have considered the effects of AI on freedom of expression in any detail. The one entity that has, Germany, has passed a hate speech law online that has been criticised as unduly restricting free expression. States may wish to consult the work of the UN Special Rapporteur on Freedom of Expression, David Kaye, who has recently published a paper on the consequences for freedom of expression on automated content regulation online, and whose upcoming report will focus on Artificial Intelligence and the freedom of opinion and expression.[120]

5. Right to Equality and Non-Discrimination: Many of the national strategies discuss the risk of bias and the need not to discriminate in AI. Most of the regulatory solutions proposed, however, are for soft models, and fail to refer to existing principles of anti-discrimination law.

The European Commission has supported a pilot on "algorithmic awareness building" to design policy responses to the risk of bias in AI systems, and its AI high-level expert group is reportedly considering the same. One potential mitigating factor that is regularly proposed is to involve a *diverse* range of voices in the development of AI, in the hopes this will help companies and public authorities spot and mitigate biased applications of AI early.

The Villani paper, for France, proposes discrimination impact assessments; the German self-driving car regulations explicitly ban any discrimination between individuals (or groups) in the event of accident; and the Italian paper makes proposals regarding testing for bias. The Council of Europe state flatly that "differential treatment will be unjustified and unlawful where it relies on biased data to generate a risk assessment." The UK House of Lords discuss bias extensively, but without reference to existing legal protections against discrimination—the main reference is to the creation of a "challenge fund" to innovate around bias testing. The Article 29 working party have published a full paper analysing the requirements on profiling, which holds that AI-powered profiling inherently creates a risk of bias and must be carefully monitored. The UK ICO analyse this issue through the lens of the "accuracy" requirement in the GDPR.

Principal gaps: On a granular level, what is missing from most member states' analyses, aside from the data protection authorities, is a clearer sense of how profiling and inferred data can capture other sensitive and protected characteristics (postcodes or names can track socioeconomic class or

[119] *See* p.9 of EDPS 2015 big data paper, *available at* https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf : " the rights to privacy and to the protection of personal data are a precondition for individuals to develop their personalities and to lead their own lives as independent human beings, as well as a precondition to exercise cherished rights and freedoms, and indeed, also a precondition for individuals and also for society to innovate."

[120] David Kaye, A human rights approach to platform content regulation, *available at* https://freedex.org/a-human-rights-approach-to-platform-content-regulation/

The paper has five key recommendations for states: 1) repeal of any law that unduly restricts expression [including online]; 2) "smart regulation…should be the norm," and states should "refrain from imposing disproportionate sanctions…on internet intermediaries"; 3) states should not require "'proactive' monitoring or content filtering as it violates privacy rights and may amount to pre-publication censorship; 4) states should not regulate in ways that delegate the responsibility to adjudge content to companies, rather than judges or state bodies; and 5) states should publish transparency reports on all content requests to internet companies.

ethnicity, for example). DPAs have handled this by stating that this inferred data needs to be treated as personal data, but some mass data applications—which purport to assess places, rather than people, for example—may evade the data protection framework. Proper handling of bias may require provisions from other human rights law, such as "disparate impact" and other concepts from anti-discrimination jurisprudence.

Generally, most strategies do not properly consider the impact the development, deployment and use of AI on vulnerable and at risk communities.

6. Due Process: This is a consistent gap in the AI strategies, and in general in the advisory papers as well. Only the Council of Europe deals with the right to due process in terms. The EU Commission's Communication on AI notes the EU will support the use of AI in the justice system, but without detail. Many papers (the UK House of Lords, the Villani paper, the Italian paper) note the well-documented problems with the COMPAS system in the United States. But in general, and given the extensive push by some European law enforcement authorities to acquire AI-powered technologies, this is a major gap in the public discussion of AI uses.

Principal gaps: Most potential applications of AI to public order and criminal justice processes are under-discussed. This should urgently be remedied. This is a timely part of the public debates on socially acceptable uses of AI: an increasing number of European police forces are using, testing, or turning to the private sector to develop AI-powered policing tools. The French city of Nice, for example, is reportedly cooperating with defence contractor Thalès to carry out a "Safe City Experimentation Project", which includes analysis of video surveillance, and "developing new analysis and correlation algorithms to better understand a situation and develop predictive capacities"—in short, predictive policing.[121] In the UK, Kent Police have been using PredPol's predictive policing software since 2013; Durham Constabulary recently drew criticism because they used a custody assessment algorithm to determine whether to grant police bail that included in its variables an individual's postcode.

7. Right to Data Protection and User Control: The EU Commission's strategy acknowledges the GDPR but does not treat its application to AI in detail. The EDPS, Article 29 Working Group, the CNIL, and ICO have published extensive discussions on how AI applications are likely to affect these rights. Other strategies which do not discuss the GDPR in detail contain provisions seemingly aimed at meeting some of its requirements. For instance, the French strategy emphasises the importance of data portability and citizen control while the House of Lords support data trusts for similar reasons.

Principal gaps: This is where the gap between national strategies and data protection authorities looms largest. What is needed is more engagement with these legal issues by government agencies *other* than the data protection authorities. And there needs to be much broader and deeper regulatory thinking about how to overcome some of the basic challenges set out in the EDPS EAG paper: how purpose limitation can work when AI applications are inherently speculative and seek previously-unknown patterns, for example. Another essential area of debate involves the right to object to fully automated processing: in reality, most "narrow AI" applications have *some* human involvement, but an algorithm may still have played a problematic role in influencing a significant decision. The Article 29 Working Party's paper on profiling is an essential starting point for this conversation.

[121] *See* Nice Safe City Project, *available at* https://www.laquadrature.net/files/Convention d expérimentation Safe City ville de Nice.pdf

It is also important to note the limits of data protection law which can make it less suited to the protection of some other human rights that may be affected by AI; for example, the WP29 opinions are largely silent on questions relating to the future of work, and they don't deal with the effects of AI-powered content regulation on freedom of expression. Perhaps more surprisingly, the opinions have not yet assessed the human rights implications of AI use by states in, for example, policing or security in significant detail.

More generally, data protection is not a panacea. For example, if data is anonymised it is taken out of the data protection framework. Yet many problematic uses of AI don't necessarily attach to an individual. The predictive policing "heatmaps" generated by US corporation PredPol, for example, refer in the aggregate to arrest records to make racially biased predictions about criminal activity in a given neighbourhood. The bias in these protections makes them problematic for EU authorities to use, but elude data protection regulations.

For issues where the individual's right to control their information is less obviously at stake, and the issues are collective—questions of bias or discrimination may affect a whole community, as do the threats to a free press and diverse debate that algorithmic news feeds cause—a different analysis will be required under human rights law (ECHR anti-discrimination principles, for example.) Some initial regulatory thinking on these issues has been carried out by the Council of Europe.

8. Collective Rights to Free Press and Free Elections: The UK's ICO has published an extensive paper on these issues as they have affected the UK in the wake of the Cambridge Analytica/Facebook/Brexit debate in Britain. The Council of Europe has also discussed the issue, calling for regulation of spending in this area, as have other national data protection authorities.

Otherwise virtually none of the states have incorporated this major issue into their regulatory thinking. The UK House of Lords has recommended that the Competition and Markets Authority consider the position of some of the most dominant actors in the tech market.

Principal gaps: These are gaps in virtually every national strategy and one important area where they should look to the work of data protection authorities, as well as investigative journalists and civil society, to catch up.

9. Economic Rights and the Future of Work: By contrast, both member states and the EU Commission have given some thought to AI and the future of work: most national strategies set aside funds for retraining and amelioration of AI-driven job displacement. (Some version of this is contained in the German cornerstone, the French strategy, the UK strategy, as well as the Nordic strategy, and it is the single social area considered in the most detail by the Finnish authorities, for example.) Dedicated retraining schemes and money into education, as well as the European stabilisation fund, are all important ways of protecting economic and social rights in the wake of the AI revolution.

Principal gaps: The pragmatic steps pledged by states are a positive start, but we note none of the analyses carried out involve a labour rights or human rights framework. The Council of Europe, in passing, refers to the problem that AI-driven "social sorting" may unfairly constrict some people's access to public services. But other issues arise. Already we have seen how "disruptive" start-ups have, by efforts to skirt around labor regulation, distorted sectors of the labor market.[122] States should take steps to protect AI from undermining people's rights in the employment context.

10. Laws of War: The only nations to address this issue are the UK and France, both of which seem to rule *in* the idea that some autonomous weapons could lawfully and safely be developed.

Principal gaps: This is a highly contentious area of the law of war which needs regulatory attention. Prior moving into this area, if at all, which posit significant risks for human rights, including the right to life, we would recommend that governments first address questions related to the current use of automated technologies, including AI, for law enforcement and national security purposes.

General gaps: In addition to the above detailed gaps in relation to specific rights and principles, we have identified three main general areas where structural questions are to be answered. It is imperative that the identified criteria for rights and principles are respected in the policies that will be relevant for collective issues, AI and public authorities and the consideration of development of no-go areas for AI.

Collective issues: AI is, to a large extent, a mass technology. Many of AI's greatest benefits—and risks—reach all of us. This means an individual, exercising her individual rights, will not always be best placed to assess or challenge any harm wrought by a mass system. Few states have dealt with this—the difficulty encompassing the collective nature of harms within individual rights and data protection frameworks. This is among the most interesting contributions of the Villani paper: to question whether an individual framework is, standing alone, fit for purpose, and to propose more collective thinking about society's regulatory needs:

> *"[C]urrent legislation, which focuses on the protection of the individual, is not consistent with the logic introduced by these systems—i.e. the analysis of a considerable quantity of information for the purpose of identifying hidden trends and behavior—and their effect on groups of individuals. To bridge this gap, we need to create collective rights concerning data."*[123]

AI and public authorities: Perhaps the most important gap in states' thinking is the question of how public law principles that constrain state behaviour also should constrain their use of AI in certain contexts. Italy has done the most to consider how public authorities, in particular, should apply AI.

No-go areas for AI? Related to the question of public authority use: are there areas where it is undesirable as a matter of public policy to have AI-powered tools in the hands of public or private bodies in general? Are there uses of AI the law should, or already may, ban? None of the papers address this and it seems an essential problem to discuss.

This goes to another important and missing debate around state use of AI--whether there are areas in public life where it is deemed undesirable to have an AI-powered system at all. Some experts we consulted have suggested the use of AI-supported facial recognition software by law enforcement or autonomous weapons as examples. Some police are already using AI (DANTE in the EU, Durham police), but there is little mention of this in the published discussion papers. There is also no developed discussion on the use of AI by the military even if the UK and France briefly mention their interests in using technologies for that purpose.

[122] *See* The Guardian, Inside the gig economy: the 'vulnerable human underbelly' of UK's labour market, *available at* https://www.theguardian.com/inequality/2017/aug/24/inside-gig-economy-vulnerable-human-underbelly-of-uk-labour-market.

[123] *See* p. 114 of Report of Mission Villani, *available at* https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

Much more advance thought is needed around the relationship between AI and consequential public sector decision-making, particularly immigration, national security, and criminal justice.

2. The EU's place in the AI race: the human factor

The Spectre of Silicon Valley: "Move Fast and Break Things" or "Unsafe at Any Speed?"

> *"Innovation is not sustainable without public trust"*—Elizabeth Denham, UK ICO[124]

Underpinning some of these gaps may be an anxiety: that the US' free-wheeling regulatory environment is a major reason the internet behemoths formed and became so wildly successful. However, the grim realities of internet shutdows, walled-gardens, censorship, zero-day exploits, hate speech, data protection and privacy violations and disinformation are increasingly threatening to overshadow the internet's transformative powers to realise human rights.

Analogies to 20th century automobile history may be instructive. In 1965, Ralph Nader shook the US car industry with a trenchant exposé of carmakers' refusal to introduce basic safety measures, *Unsafe at Any Speed.*[125] It became a bestseller and helped birth a new axis of competition for automakers: safety. And European cars became early and effective competitors in this market, with superior safety standards and quality engineering becoming synonymous in global markets with European quality. Even automakers at the time recognised that regulation had the effect of levelling the playing field to an extent: "It sets ground rules where everybody has to do something and nobody has to worry about being at a competitive disadvantage," one former automobile executive recalled.

European data protection authorities have rightly called out the existing culture of unregulated internet services as unsustainable—and unsafe. Citing the Cambridge Analytica fiasco in a keynote speech to the telecommunications forum, EDPS Giovanni Butarelli stated: "This year practices have been coming to light which contradict the most basic principles of not only data protection but basic respect for people." [126] He added that the entire purpose of the EU's regulatory framework – the GDPR and others – is "to change market incentives, encourage innovation so that access to information on the internet does not depend on being watched all the time."

This debate goes to the heart of business on the internet today, and the data protection authorities' argument is gaining steam. The dominant economic model of web services, in which companies collect the maximum possible data on people and then seek ways "to monetise that data," need to be rethought afresh if the internet as a productive and *trusted* public space is to survive.

A cautionary tale: China - Meanwhile, it has become common to hold up China – with its sinister system of "social credit", and use of facial recognition software to pull suspects out of pop concerts – as *the* exemplar of an AI-powered dystopia to avoid.[127]

It is simpler to say, however, that the pluralist democracies of the EU would never go down China's

[124] *See* Computing, ICO: There's so much imisinformation out there on GDPR, *available at* https://www.computing.co.uk/ctg/news/3027593/ico-theres-so-much-misinformation-out-there-on-gdpr

[125] *See* The New York Times, 50 Years Ago, Unsafe at Any Speed Shook the Auto World, *available at* https://www.nytimes.com/2015/11/27/automobiles/50-years-ago-unsafe-at-any-speed-shook-the-auto-world.html

[126] *See* EDPS speech to EU Telecommunications forum, *available at* https://edps.europa.eu/sites/edp/files/publication/18-04-24_giovanni_buttarelli_keynote_speech_telecoms_forum_en.pdf

[127] *See* QZ, China facial recognition, *available at* https://qz.com/1285912/chinas-facial-recognition-cameras-keep-catching-fugitives-at-pop-star-jacky-cheungs-concerts/

path than it is to spot and avoid problematic local equivalents. If an insurance company datamines your social media posts to assess your lifestyle, your risk, and therefore set your fees, is that not a soft, privatised form of social credit scoring? Note that much of China's social credit system was originally developed and is maintained by private companies.[128] If the police's AI-powered intelligence algorithm searches Twitter or Facebook for hidden patterns of erratic or unusual social media posts, and those analyses then drive arrest decisions, in what ways does this differ from the Chinese example? The UK police trialled facial recognition software at the 2017 Notting Hill Carnival, a popular festival that is also historically associated with London's Afro-Caribbean communities. This underscores a question that is insufficiently discussed in national AI strategies: what are the circumstances in public life where—for historical reasons of mistrust between a community and public authorities, or because of the sensitivity of the issue—it may undesirable to deploy AI-powered sifting technologies at all?

The Chinese case also serves as a reminder that the world's largest data companies—including those at the forefront of AI development—evolve over time, and not always for the better. Recently it emerged that, years after having pulled out of China because of pervasive censorship, Google has been developing a tailored product for the Chinese market: Dragonfly, a censored version of its search engine.[129] This would, in effect, capitulate to the Chinese government's demands for total information control.

This is a hotly contested decision by Google, but it underscores a crucial point about AI regulation: leaving powerful corporate entities to self-regulate may be a recipe for dilution of Europe's most cherished and hard-won liberties. The time has passed when the simple narrative of tech giants as garage-founded, "don't be evil" values-driven corporations accurately describes their role in modern life. They are too powerful and pervasive—more like public utilities or broadcasters—simply to be left to manage themselves.

Critical community engagement in the United States: a positive example - What are the borders of socially acceptable use of AI? Some communities in the United States have begun to engage precisely this question. Partly because of the controversies around predictive policing, and American police forces' troubling record of abusing black citizens and people of colour, researchers and advocates in the US have found innovative ways to manage public authorities' use of AI.

In the city of Oakland, California, after researchers investigated the use of predictive and algorithmic tools by police, the city decided to bring these tools under better democratic control. A recent initiative by the City Council requires a Privacy Advisory Commission to review any new procurement before the acquisition of any new algorithmic tool by police. [130] [131] European regulators would do well to monitor these forward-thinking examples for best practices and adopt useful regulatory frameworks where appropriate.

Other international actors - Other international actors, too, have joined US cities and European authorities in highlighting the opportunities and risks of an automated society.

[128] *See* Wired, China social credit, *available at* https://www.wired.co.uk/article/china-social-credit "Ant Financial, the finance arm of e-commerce giant Alibaba, launched a product called Sesame Credit in 2015. It was China's first effective credit scoring system but was also much broader, functioning as a social credit scheme and loyalty programme as well."

[129] *See* The New York Times, Google Tried to Change China. China May End Up Changing Google. *available at* https://www.nytimes.com/2018/08/22/technology/google-china-conventionality.html

[130] *See* Oakland City Council Ordinance, *available at* https://www.documentcloud.org/documents/4450176-View-Supplemental-Report-4-26-18.html

[131] *See* Slate, How Cities Are Reining In Out of Control Police Tech, *available at* https://slate.com/technology/2018/05/oakland-california-and-other-cities-are-reining-in-out-of-control-police-technologies.html

The Organisation on Economic Cooperation and Development (OECD) has also published various papers on AI regulation.[132] The central theme emerging from the OECD appears to be that AI policy is urgent, and should be coordinated internationally as far as possible, but that where AI poses potential "threats to privacy and democratic principles," law or regulation will need to address these.[133] The OECD is proposing to publish a set of guidelines that aim to help its member states "balance innovation with regulations" that foster the OECD's core values--"democracy, prosperity, and inclusivity"--resulting in a draft recommendation for the Council in 2019.[134]

David Kaye, the UN Special Rapporteur on Freedom of Expression, recently published an extensive report on recent laws that force internet platforms to run algorithmically-driven censorship policies, noting their potential harm on freedom of expression. This is one area, perhaps, where Europe's historical and justified concern with the far right may have led to regulatory differences which are viewed as over-reach in some context or jurisdictions. The German law on hate speech and its penalties, in particular, have come in for trenchant criticism and may need to be reassessed.

Europe's offer: the human factor - Europe's challenge will be to develop artificial intelligence policy that promotes innovation but steers between the "Wild West" approach that characterised the early Silicon Valley era, and the statist approach of China. This is not a simple gold rush—nor is it a doomsday scenario that requires iron-clad regulation across the board. Rather, every socially significant use of artificial intelligence should be assessed in context, critically judged for its effect on European rights and freedoms, and regulated accordingly.

To get this right, it will also be essential to expand the debate well beyond the corners of the specialist technological press or the privacy community. Citizens are waking up to the potential misuse of their data and at the same time realising the opportunities of AI. After the Cambridge Analytica fiasco, a Rubicon has been crossed: fewer people see the internet as a private or inconsequential space.[135] More and more, people will expect states and corporations to respect their rights and personal boundaries as they live online.

While states and regulators may always be playing catch-up with technological change, that is no reason to cede the regulatory field. Human rights anchored, ethical and legal principles that guide a better, more tailored offer are possible. It is possible for European regulation to protect citizens from trading their right to a private life to use essential internet services, because they think they have no other choice.[136] This—developing smart AI regulation that keeps the human factor at the centre of the frame—could and should be Europe's unique offer.

[132] While the OECD is not a regulator, it "supports governments through policy analysis, dialogue and engagement and identification of best practices." *See* http://www.oecd.org/going-digital/ai/

[133] *See, e.g.,* AI: Intelligence machines, smart policies, *available at* https://www.oecd-forum.org/users/68225-wonki-min/posts/38898-harnessing-ai-for-smart-policies

[134] *See also* OECD creates expert group to foster trust in artificial intelligence, (announcing creation of expert AI policy group), *available at* http://www.oecd.org/going-digital/ai/oecd-creates-expert-group-to-foster-trust-in-artificial-intelligence.htm

[135] *See, e.g.,* HarrisX Tech Media Telecom Survey, *available at* http://harrisx.com/wp-content/uploads/2018/04/Inaugural-TMT-Pulse-Survey_-20-Apr-Final.pdf Showing over 80% of Americans believe technology companies should be held legally responsible for their content and that privacy protections for users should be strengthened.

[136] While this is only a subset of consequential AI applications, it is an important one: The data model that has permitted Google, Facebook, and Amazon in particular to rise to hyper-dominant market positions is one of harvesting and analysing vast stores of data to sell to advertisers. The current model of the data economy turns on this: services are "free", because the companies sell users' data to advertisers. This "behavioural advertising" model is coming under increased scrutiny in the wake of the Cambridge Analytica case. The vast stores of data this handful of market players have collected also positions them to dominate the coming artificial intelligence race.

E. CONCLUSION - AN AI SPECIFIC HUMAN RIGHTS ASSESSMENT

> *“The application of a human rights framework is crucial because it goes beyond just ensuring transparency and accountability, as it ensures that all rights are effectively considered in automated decision-making systems such as algorithms.”- Council of Europe, March 2018*[137]

Key Conclusions and Recommendations

- ➔ **Transparency:** Given the importance of these human rights issues at play in the context of AI and the scale of the transformation in prospect, public bodies should engage the public in these decisions to the fullest possible extent. This includes designing meaningful participative and inclusive process allowing for external inputs into AI strategies, from public consultation to expert peer reviews. On the use of AI, when public authorities seek to acquire an AI system for public use, the procurement of such a system should be done openly and transparently according to open procurement standards. This should include the purpose of the system, goals, parameters, and more. Procurement should also include a period for public comment, and States should reach out to potentially affected groups where relevant to ensure they have input. States should also ensure people’s right to know when an automated process is used to affect or make a decision about them.
- ➔ **Accountability:** States bear the primary duty to promote, protect, respect, enforce and fulfill human rights. States must not engage in or support practices that violate rights when designing or implementing AI systems. Accountability can partly be ensured by preserving a human in the loop for consequential decisions. For high-risk areas such as criminal justice, significant human oversight is necessary. The desire to eliminate human bias from decision-making is understandable--even laudable. But democracy, due process, and human dignity all urge that where an important decision is made about a person, they should be able to query that decision of another human. There are areas of public life where “computer says no” is, for the time being, democratically unacceptable. Additionally, States must thoroughly investigate AI systems to identify potential human rights risks prior to development or acquisition. Finally, governments must draw lines for themselves to act as guardrails in their use of AI. If a given use of AI is deemed to cause human rights harms, government use of AI in this context should be prohibited. These red lines should be reexamined regularly to account for technological advancements and shifts in government policy.
- ➔ **Right to Privacy:** In a large part, national and European strategies on AI do not adequately address the impact and risk of the use of AI on the right to privacy. More work on this area is urgently required which should be built on the swift adoption of a comprehensive and robust ePrivacy Regulation.
- ➔ **Freedom of Conscience and Expression:** Many of the national strategies discuss the risk of bias and the need not to discriminate in AI. Most of the regulatory solutions proposed, however, are for soft models, and fail to refer to existing principles of anti-discrimination law or rights. In addition, most strategies are missing analyses on how profiling and inferred data can capture other sensitive and protected characteristics.
- ➔ **Right to Equality and Non-Discrimination:** Surprisingly little governmental work to date

[137] *See* p. 43 of Council of Europe study on automated decision, *available at* https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10

applies existing principles of anti-discrimination law to AI systems; more thought in this area is timely.

- ➔ **Due Process:** There is a consistent gap in the AI strategies in addressing due process. ore thought needs to go into the use of AI by public authorities in consequential areas: criminal justice, immigration, national security. There is the germ of a useful debate in the Council of Europe paper, but considerably more work needs to be done at national and EU level.
- ➔ **Right to Data Protection and User Control:** The GDPR, in theory, provides robust protections for citizens. But data protection authorities may lack sufficient resources to deal with them properly. Their scope in the AI context, particularly in areas such as the right to object to fully automated processing, and the rules around profiling, are likely to be hotly contested. Moreover, the technical challenges are real: how to use AI effectively while complying with purpose limitation, or how to improve explainability of deep learning techniques, has yet to be resolved. There is a need to support extended public comment and debate on the application of these principles to AI.
- ➔ **Collective Rights to Free Press and Free Elections:** Many of the risks of AI are by their nature collective. This insight, set out best in the Villani paper for France, is worth careful consideration and development, particularly in the context of the threats AI systems have already posed to a free press and free elections.
- ➔ **Economic Rights and the Future of Work:** The national authorities have thought about AI's effects on the future of work and at least begun to set aside funding to ameliorate these effects. This is positive.
- ➔ **Laws of War:** This is likely to become a highly contentious area of the law of war in the medium term and will need regulatory attention, in particular as few states are considering developing autonomous weapons. There is a clear lacuna in states' regulatory thinking.

To conclude, in many ways the EU is well-placed to regulate AI by identifying gaps in safeguards. The GDPR was an essential beginning to this process, and the EU Commission is right to hail it. But experts are right to say that the current "patchwork" of regulatory initiatives risks ineffectiveness – either a race to the bottom or forum shopping. Europe should aim for a consolidated approach to the regulation of AI, that is sensitive to the various contexts in which AI is already being developed and used, and to be sure it is applied and enforced in a consistent, rights-respecting way across the Union.

Finally, it is perhaps for states to acknowledge that there is not a single race for AI but multiple ones, going in opposite directions. While some seem to only have military developments in minds, the EU has the potential to lead the development of a user-centric AI by reaffirming its values and safeguarding rights. By doing so, Europe has an opportunity to define the direction it wants AI innovation to go, one that hopefully can truly work towards AI for Humanity.

F. ANNEXES

The chart below assesses the strategy papers on a 1-3 scale as follows:

1. the right is not accounted for,
2. the right is noted/accounted for in the strategy, but without a resolved concrete proposal (or a soft law proposal),
3. the right is accounted for with a concrete or hard law proposal.

The chart was used for the development of the comparative analysis presented in this report.

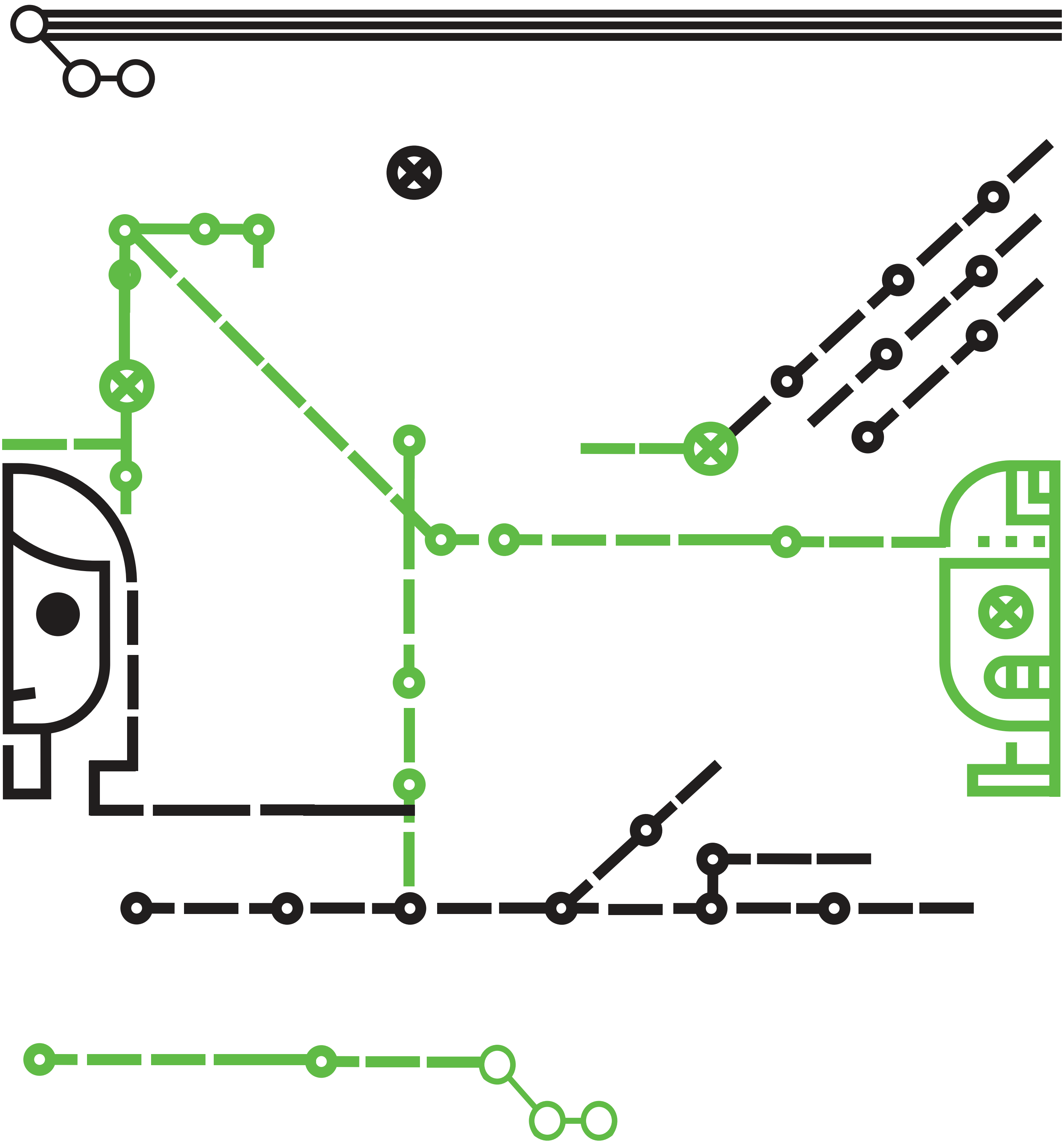
	EUROPE-WIDE	FRANCE 🇫🇷	GERMANY 🇩🇪	THE UK 🇬🇧	NORDIC-BALTIC	FINLAND 🇫🇮	DENMARK 🇩🇰	ITALY 🇮🇹	SPAIN 🇪🇸
Transparency & Explainability	EU Strategy: 2 DPA: 3 Council of Europe: 2	Strategy: 2 Villani: 2	2	Strategy : 2 House Of Lords (Hol): 3 ICO: 3	2	2	1	2	N/A
Accountability & Right to a Remedy	EU Strategy: 2 DPA: 3 Council of Europe: 2	Strategy: 1 Villani: 2	2	Strategy: 2 Hol: 2 ICO: 3	1	2	1	2	N/A
Right to a Human Decision-maker; No Fully Automated Decision-making	EU Strategy: 3 Dpa: 3 Council of Europe: 2	Strategy: 1 Villani: 3	2	Strategy: 2 Hol: 2 ICO: 2/3	1	1	1	2	N/A
Right to Privacy	EU Strategy: 3 DPA: 3 Council of Europe: 2	Strategy: 2 Villani: 3	3	Strategy: 2 Hol: 2 ICO: 3	2	2	1	2	N/A
References to binding Data Protection Rights (Data Portability; Purpose Limitation; Data Minimisation; Accuracy; Storage Limitation)	EU Strategy: 2 DPA: 3 Council of Europe: 2	Strategy: 3 Villani: 3	2	Strategy: 2 Hol: 3 ICO: 3	2	1	1	2	N/A
Freedom of Expression	EU Strategy: 1 DPA: 2 Council of Europe: 2	Strategy: 1 Villani: 1	3	Strategy: 1 Hol: 1 Ico: 1	1	1	1	1	N/A
Right to Equality and Non-discrimination	EU Strategy: 2 DPA: 2 Council of Europe: 2	Strategy: 2 Villani: 3	2	Strategy: 2 Hol: 2 ICO: 3	1	1	1	2	N/A
Due Process and Right to a Fair Trial	EU Strategy: 1 DPA: 2 Council of Europe: 2	Strategy: 2 Villani: 2/3	2	Strategy: 1 Hol: 2 ICO: 1	1	1	1	2	N/A
Free and Fair Elections	EU Strategy: 1 DPA: 2 Council of Europe: 2	Strategy: 1 Villani: 1	1	Strategy: 1 Hol: 1 ICO: 3	1	1	1	1	N/A
Economic, Social, Cultural Rights, Incl. Mitigation of AI's Effects on the future of work	EU Strategy: 3 DPA: 1 Council of Europe: 2	Strategy: 2 Villani: 2	2	Strategy: 2 Hol: 2 ICO: 1	2	2	1	1	N/A
Collective Data Rights	EU Strategy: 1 DPA: 1 Council of Europe: 1	Strategy: N/A Villani: 3	1	Strategy: 1 Hol: 1 ICO: 1	1	1	1	1	N/A
Conflict; Law Of War; Autonomous Weapons Systems	EU Strategy: 1 DPA: 1 Council of Europe: 1	Strategy: 1 Villani: 2	1	Strategy: 2 Hol: 3 ICO: 1	1	1	1	1	N/A

accessnow

Access Now is a global non-profit organization that defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots grantmaking, and convenings such as RightsCon, we fight for human rights in the digital age.

With the support of the Vodafone Institute:

Vodafone Institute
for Society and
Communications



accessnow

Vodafone Institute
for Society and
Communications