PRIVACY INTERNATIONAL

ARTICLE 19

Privacy and Freedom of Expression In the Age of Artificial Intelligence

April 2018

About Us

ARTICLE 19 is a global human rights organisation, which works around the world to protect and promote the right to freedom of expression and information ('freedom of expression'). Established in 1987, with its international office in London, ARTICLE 19 monitors threats to freedom of expression in different regions of the world, and develops long-term strategies to address them.

ARTICLE 19 is actively engaged in promoting fair, accountable, and transparent Artificial Intelligence (AI), and investigates the human rights impact of algorithmic decision making through policy engagement and research. Our work on AI and freedom of expression thus far includes a policy brief on algorithms and automated decision making, several submission to the AI and ethics initiative of the Institute of Electrical and Electronics Engineering (IEEE), a recent submission to the UK House of Lords Select Committee on AI, co-chairing several working groups of the IEEE initiative, membership in the Partnership on AI (PAI), and guidance on the development of AI for network management in the Internet Engineering Task Force (IETF).

Privacy International is a non-profit, non-governmental organisation based in London, dedicated to defending the right to privacy around the world. Established in 1990, Privacy International undertakes research and investigations into government surveillance and data exploitation in the private sector with a focus on the technologies that enable these practices. To ensure universal respect for the right to privacy, Privacy International advocates for strong national, regional and international laws that protect privacy around the world. It has litigated or intervened in cases implicating the right to privacy in the courts of the United States, the UK, and Europe, including the European Court of Human Rights and the European Court of Justice.

Our work on AI and privacy thus far includes a submission to the AI and ethics initiative of the Institute of Electrical and Electronics Engineering (IEEE), a recent submission and oral evidence to the UK House of Lords Select Committee on AI, as well as several submissions on profiling and automated decision-making to the Article 29 Working Party - an advisory body made up of a representative from the data protection authority of each EU Member State, the European Data Protection Supervisor and the European Commission that is drafting guidance to the EU General Data Protection Regulation.

Contents

Executive Summary

Artificial Intelligence (AI) is part of our daily lives. This technology shapes how people access information, interact with devices, share personal information, and even understand foreign languages. It also transforms how individuals and groups can be tracked and identified, and dramatically alters what kinds of information can be gleaned about people from their data.

AI has the potential to revolutionise societies in positive ways. However, as with any scientific or technological advancement, there is a real risk that the use of new tools by states or corporations will have a negative impact on human rights.

While AI impacts a plethora of rights, ARTICLE 19 and Privacy International are particularly concerned about the impact it will have on the right to privacy and the right to freedom of expression and information.

This scoping paper focuses on applications of 'artificial narrow intelligence': in particular, machine learning and its implications for human rights.

The aim of the paper is fourfold:

1. Present key technical definitions to clarify the debate;

2. Examine key ways in which AI impacts the right to freedom of expression and the right to privacy and outline key challenges;

3. Review the current landscape of AI governance, including various existing legal, technical, and corporate frameworks and industry-led AI initiatives that are relevant to freedom of expression and privacy; and

4. Provide initial suggestions for rights-based solutions which can be pursued by civil society organisations and other stakeholders in AI advocacy activities.

We believe that policy and technology responses in this area must:

- Ensure protection of human rights, in particular the right to freedom of expression and the right to privacy;

- Ensure accountability and transparency of AI;

- Encourage governments to review the adequacy of any legal and policy frameworks, and regulations on AI with regard to the protection of freedom of expression and privacy;

- Be informed by a holistic understanding of the impact of the technology: case studies and empirical research on the impact of AI on human rights must be collected; and

- Be developed in collaboration with a broad range of stakeholders, including civil society and expert networks.

Introduction

Artificial Intelligence (AI) and its applications are a part of everyday life: from curating social media feeds to mediating traffic flow in cities, and from autonomous cars to connected consumer devices like smart assistants, spam filters, voice recognition systems and search engines.

The sudden rise of these applications is recent, but the study and development of AI is over half a century old: the term was coined in 1956, though the concept goes back even further, to the late 1700s. Current momentum is fuelled by the availability of large amounts of data, affordable and accessible computational power, continued development of statistical methods, and the fact that technology is now embedded into the fabric of society. We rely on it in more ways than most are even aware of.[1]

If implemented responsibly, AI can benefit society. However, as is the case with most emerging technology, there is a real risk that commercial and state use has a detrimental impact on human rights. In particular, applications of these technologies frequently rely on the generation, collection, processing, and sharing of large amounts of data, both about individual and collective behaviour. This data can be used to profile individuals and predict future behaviour. While some of these uses, like spam filters or suggested items for online shopping, may seem benign, others can have more serious repercussions and may even pose unprecedented threats to the right to privacy and the right to freedom of expression and information ('freedom of expression').[2] The use of AI can also impact the exercise of a number of other rights, including the right to an effective remedy, the right to a fair trial, and the right to freedom from discrimination.

The threat posed by AI thus does not take the form of a super-intelligent machine dominating humanity: instead, core problems with AI can be found in its current everyday use.[3] This scoping paper focuses on applications of 'artificial narrow intelligence', in particular machine learning, and its implications for human rights.

1 Cath, C., Wachter, S., Mittelstadt, B. et al., 'Artificial Intelligence and the 'Good Society': The US, EU, and UK Approach', Science and Engineering Ethics, 2017, p. 1 – 24.

2 See for example, V. Dodd, Met police to use facial recognition software at Notting Hill carnival, The Guardian, 5 August 2017; available from: https://www.theguardian.com/uk-news/2017/aug/05/met-police-facial-recognition-software-notting-hill-carnival.

3 ARTICLE 19, Submission to the House of Lords Select Committee on Artificial Intelligence, 6 September 2017, available from: https://www.article19.org/wp-content/uploads/2017/10/ARTICLE-19-Evidence-to-the-House-of-Lords-Select-Committee-AI.pdf.

The right to freedom of expression and the right to privacy are mutually reinforcing – all the more so in the digital age.[4] Privacy is a prerequisite to the exercise of freedom of expression: without it, individuals lack the space to think, speak and develop their voice. Without freedom of expression, individuals would be unable to develop their sense of self.

Both of these rights are therefore essential foundations for open and democratic societies and among the basic conditions for progress, as well as for each individual's self-fulfilment. For democracy, accountability and good governance to thrive, freedom of expression must be respected and protected. The same is true of the right to privacy, which is a powerful bulwark against state and corporate power.

While many talk about the need to think carefully about the incorporation of AI applications in the 'safety-critical systems' of societies - like electricity grids and water supplies - it is important to re-centre the larger debate on the use of AI in 'human rights critical contexts'.

It is imperative that policy makers, regulators, companies, civil society, and other stakeholders working on the right to privacy and freedom of expression understand the implications, risks and potential of AI.

In this scoping paper, ARTICLE 19 and Privacy International aim to contribute to such understanding. First, we present key technical definitions to clarify the debate and examine key ways in which AI impacts the right to freedom of expression and the right to privacy, and we outline key challenges. We then review the current landscape of AI governance, including various existing legal, technical, and corporate frameworks and industry-led AI initiatives relevant to freedom of expression and privacy. Finally, we also provide initial suggestions for rights-based solutions which can be pursued by civil society organisations and other stakeholders in AI advocacy activities.

4 ARTICLE 19, The Global Principles on Protection of Freedom of Expression and Privacy, 2017, available at http://article19.shorthand.com/.

What is Artificial Intelligence?

The term 'AI' is used to refer to a diverse range of applications and techniques, at different levels of complexity, autonomy and abstraction. This broad usage encompasses machine learning (which makes inferences, predictions and decisions about individuals), domain-specific AI algorithms, fully autonomous and connected objects and even the futuristic idea of an AI 'singularity'.

This lack of definitional clarity is a challenge: different types of AI systems raise specific ethical and regulatory issues.

From a conceptual point of view, it is important to consider the following key concepts in AI related debates:

- **Artificial narrow intelligence** is the ability of machines to resemble human capabilities in narrow domains, with different degrees of technical sophistication and autonomy. Examples include: chatbots that assist by answering specific questions; Deep Blue, a chess-playing computer developed by IBM which famously beat world chess champion Garry Kasparov in May 1997;[5] or the computer system which defeated the reigning master of the Chinese board game Go in May 2017. [6]

- **Artificial general intelligence** is the overarching, and as yet unachieved, goal of a system that displays intelligence across multiple domains, with the ability to learn new skills, and which mimic or even surpass human intelligence. It is theorised that the creation of artificial general intelligence could lead to the 'singularity', or a period of runaway technological growth that profoundly changes human civilisation. This is still, at the very least, decades away, if not entirely implausible. [7]

- **Algorithm** can refer to any instruction, such as computer code, that carries out a set of commands: this is essential to the way computers process data. For the purposes of this paper, it refers to 'encoded procedures for transforming input data into the desired output, based on specific calculations.' [8]

5 G. Kasparov, Learning to Love Intelligent Machine, The Wall Street Journal, 14 April, 2017, available from: https://www.wsj.com/articles/learning-to-love-intelligent-machines-1492174086.

6 C. Metz, 'What the AI behind Alpha Go can teach us about being human', Wired, 19 May, 2017, available from: https://www.wired.com/2016/05/google-alpha-go-ai/.

7 L. Floridi, Should we be afraid of AI?, Aeon, 9 May 2016, available from: https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible.

8 T. Gillespie, The relevance of algorithms, Media technologies: Essays on communication, materiality, and society, MIT Press, 2014, p. 167.

- **Automated decision-making** 'generally involves large-scale collection of data by various sensors, data processing by algorithms and subsequently, automatic performance.'[9] It is an efficient means of managing, organising, and analysing large amounts of data, and structuring decision-making accordingly. It may or may not rely on AI, with varying degrees of human involvement.[10] It can make decisions, or generate knowledge or information, that significantly shapes or influences the exercise of human rights.

- **Machine learning** is a popular technique in the field of AI which has gained prominence in recent years. It often uses algorithms trained with vast amounts of data to improve a system's performance at a task over time.[11] Tasks tend to involve making decisions or recognising patterns, with many possible outputs across a range of domains and applications. Arthur Samuel, who coined the term, referred to machine learning programs as those which have 'the ability to learn without being explicitly programmed.'[12] Many of the technologies commonly referred to as AI today are, strictly speaking, machine learning systems.

- **Supervised, unsupervised, and reinforced learning** Machine learning is usually classified into these three types. Supervised machine learning forms the majority of AI application today. It seeks to teach the computer to predict an output, assuming that the input data is labelled correctly. Supervised machine learning can either be used to predict a continuous valued output through regression, or a discrete valued output through classification.[13] Unsupervised learning, on the other hand, depends on the computer program to find structure within data, based on particular features. Reinforced learning is the third type, wherein the program is placed in an environment and must learn how to behave successfully within that environment, based on feedback of successes and failures.[14]

9 M. Perel & N. Elkin-Koren, Accountability in algorithmic copyright enforcement, Stanford Technology Law Review, 2016.

10 What degree of human involvement renders a decision automated is subject to debate: Article 22 of the European General Data Protection Regulation, for instance, merely covers automated decisions that are 'based solely' on automated processing, which leaves room for interpretation. The current draft guidelines by the Article 29 Working Party, for instance, argues that human intervention has to be meaningful and cannot just be a token gesture. See Article 29 Data Protection Working Party, (2017), Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679.

11 H. Surden, 'Machine Learning and the Law', 89 Washington Law Review 87, 2014, p. 89-90.

12 A. Munoz, Machine Learning and Optimisation, New York University, available from: https://cims.nyu.edu/~munoz/files/ml_optimization.pdf.

13 N.G. Andrew, Machine Learning, Coursera, available from: https://www.youtube.com/watch?v=PPLop4L2eGk&list=PLLssT5z_DsK-h9vYZkQkYNWcItqhlRJLN.

14 S.J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, Third Edition, Pearson, 2015, p. 846.

AI and the Right to Freedom of Expression

Establishing the Nexus

AI will significantly impact the right to freedom of expression.[15] It is applied in a vast number of situations that influence how individuals access and find information online. Increasingly, online intermediaries, such as social media platforms and search engines, use AI systems to control information that users engage with in opaque and inscrutable ways.[16] While some uses, for instance spam filters or suggested items for online shopping, may seem harmless, others can have more serious repercussions.

AI-powered surveillance presents one such serious repercussion. The pervasive and invisible nature of AI systems, coupled with their ability to identify and track behaviour, can have a significant chilling effect on the freedom of expression. This can take place through self-censorship, altered behaviour in public spaces and private communications alike. The rise of techniques such as video surveillance, facial recognition, behaviour analysis etc., by public authorities and private companies hinder freedom of expression and also infringe the very essence of the right to privacy.[17] Mass surveillance in particular is a disproportionate interference with privacy and the freedom of expression, while targeted surveillance may only be justified when it is prescribed by law, necessary to achieve a legitimate aim, and proportionate to the aim pursued.[18]

Moreover, the data generated and used by online intermediaries in AI systems is another concern. To leverage the affordances of AI, large amounts of data are generated and collected. This is worrying because datasets are often built through problematic methods of collection, leading them to hold biases that reflect existing patterns of societal stereotyping.[19] Even when this is not the case, data samples can be unrepresentative of the population at large. In either situation, data can lead AI applications to negatively impact freedom of expression.

15 L. Rainie & J. Anderson, Code-Dependent: Pros and Cons of the Algorithm Age, Pew Research Center, February 2017, available from: http://pewrsr.ch/2BpdxYx.

16 J. Balkin, Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation, 2017, available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3038939.

17 ARTICLE 19, The Global Principles on Protection of Freedom of Expression and Privacy, op.cit.

18 International Principles on the Application of Human Rights to Communications Surveillance (“Necessary and Proportionate Principles”), available from: https://necessaryandproportionate.org/principles.

19 R. Calo, Artificial Intelligence Policy: A Primer and Roadmap, 2017, available from: https://ssrn.com/abstract=3015350 or http://dx.doi.org/10.2139/ssrn.3015350.

For example, if AI content moderation systems are not trained on slang or non-standard use of certain expression often used by minority groups it can potentially lead these systems to censor legitimate speech.

Beyond surveillance, it is difficult to provide a comprehensive overview of the ways in which AI impacts the right to freedom of expression: its effects are sector- and context-specific. AI raises different issues depending on its application, i.e. to online content moderation on social media platforms or to smart homes. At this stage, ARTICLE 19 considers the importance of AI through the lens of our five key global areas of work and highlights the following issues:[20]

- **Digital:** Technology shapes how people exercise their right to freedom of expression, how they access information and how they interact online. AI impacts how individuals can access information and express themselves on the Internet, including through search engines and social media. It also impacts the fundamental technical functioning of the Internet itself with the increased use of AI in Internet networking. Technical standard setting bodies like the Institute for Electrical and Electronics Engineers (IEEE) are also currently in the process of developing standards for ethical and safe AI systems.

- **Civic Space:** Civic space is the physical and legal place where individuals realise their rights. This space is increasingly impacted by various AI applications, from newsfeed algorithms to connected devices in smart cities. Building on these applications, AI systems will soon shape decision-making systems and spaces where people and communities organise, associate and assemble, from homes to cities.[21]

- **Media:** Media pluralism and media freedom are essential for protecting and promoting freedom of expression and the public interest in an increasingly globalised, digitised, and converging media landscape. There is a danger that a limited number of digital corporations will become the central conduit for media content online. This is particularly the case if companies use opaque AI systems that rank information, whether on news sites or by filtering email, according to indicators that often remain unknown to the users. [22]

- **Transparency:** Transparency and access to information - from both public and private bodies - are crucial in ensuring democratic governance. Increasingly, decisions on access to information, traditionally made by humans, are now driven by AI applications. These systems have the ability to selectively exclude

20 ARTICLE 19, The Expression Agenda Report 2016-2017, available from: https://www.article19.org/wp-content/uploads/2017/12/Expression-Agenda-Report-2017-for-web-_30.11.17.pdf.

21 R. Brauneis & E. Goodman, Algorithmic Transparency for the Smart City, Yale Journal of Law & Technology, GWU Law School Public Law Research Paper, 2 August 2017, available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3012499.

22 C. Cath et al., Media Development in the Digital Age: Five Ways to Engage in Internet Governance, 2017, available from: https://www.cima.ned.org/publication/media-development-digital-age-five-ways-engage-internet-governance/.

or emphasise critical information,[23] and can enable governments to deploy AI systems without transparency.[24]

- **Protection:** Those on the frontline of defending freedom of expression and information must be supported by a strong enabling legal framework for freedom of expression, effective mechanisms for protection of the right to freedom of expression, due process of law and active networks of institutions and activists. Various applications of AI are problematic, in particular for surveillance and censorship, or targeting those who exercise their freedom of expression in a manner that is controversial in the view of governments and other powerful institutions and corporations.[25]

23 For example, recent studies show that AI has the potential to privilege advertisement of high paying jobs to men over advertisement to women. See, for example, M. Day, How LinkedIn's Search Engine May Reflect a Gender Bias, The Seattle Times, 31 August 2016, available from: https://www.seattletimes.com/business/microsoft/how-linkedins-search-engine-may-reflect-a-bias/.

24 AI NOW, AI Now 2017 Report, 2017, available from: https://ainowinstitute.org/AI_Now_2017_Report.pdf.

25 See, J. Liu,' In your Face: China's all-seeing state', BBC, 10 December 2017, available at http://www.bbc.co.uk/news/av/world-asia-china-42248056/in-your-face-china-s-all-seeing-state. See also China's CCTV surveillance network took just 7 minutes to capture BBC reporter, Tech Crunch, 13 December 2017, available at https://techcrunch.com/2017/12/13/china-cctv-bbc-reporter/.

Mapping the Landscape

International Human Rights Law

The right to freedom of expression is guaranteed in Article 19 of the Universal Declaration of Human Rights (UDHR)[26] and the International Covenant on Civil and Political Rights (ICCPR),[27] as well as in regional human rights treaties. [28]

While the right to freedom of expression is a fundamental right, it is not guaranteed in absolute terms. Under international standards, in particular Article 19 para 3 of the ICCPR, restrictions on the right to freedom of expression must be strictly and narrowly tailored and may not put the right itself in jeopardy. The method of determining whether a restriction is narrowly tailored is often articulated as a three-part test. Restrictions must: (i) be provided by law; (ii) pursue a legitimate aim; and (iii) conform to the strict tests of necessity and proportionality. [29]

There are no international standards that would explicitly deal with AI and the right to freedom of expression. However, there is a body of international standards which are relevant to the use of AI, particularly in relation to online intermediaries.[30] For instance, states should not impose a general obligation on intermediaries to monitor the information that they transmit, store, automate or otherwise use,[31] and users should have the opportunity to challenge the blocking and filtering of content.[32]

At the regional level, some aspects of AI applications have been addressed in the European Union's new General Data Protection Regulation (GDPR), although the

26 G.A. Res. 217 (III) A, UDHR, art. 19 (Dec. 10, 1948). Article 19 of the UDHR states: "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers."

27 For interpretation of Article 19 of the ICCPR, see General Comment No 34, CCPR/C/GC/3.

28 Article 9 of the African Charter on Human and Peoples' Rights, Article 13 of the American Convention on Human Rights, Article 10 of the European Convention on Human Rights and Article 11 of the EU Charter on Fundamental Rights.

29 General Comment No 34, CCPR/C/GC/3, para. 21, 22.

30 See, in particular Report of the Special Rapporteur to the Human Rights Council on the promotion and protection of the right to freedom of opinion and expression, May 16 2011, para 47, available from: http://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf; Regulation (EU) 2016/679, 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

31 Directive 2000/31/EC June 8, 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (E-Commerce Directive).

32 Recommendation CM/Rec (2008)6 of the Committee of Ministers to Member states on measures to promote and respect for freedom of expression and information with regard to internet filters, s1.

details of its bearing on the right to access to information and expression remain up for discussion. We analyse this discussion[33] in detail at a later stage in this paper.

National Frameworks

At national levels, existing AI applications are regulated by traditional frameworks of legislation including freedom of expression, data protection, consumer protection, media and competition law, along with sectoral regulations and standards.

While these frameworks cover some aspects of AI application, the question remains whether these laws are adequate to address the myriad ways in which AI will impact freedom of expression.

In response to some of the specific questions raised by the use of AI systems, some countries like Japan [34] and Germany [35] have developed new frameworks applicable to specific AI issues. These exhibit an increasing tendency of governments to publish non-binding guidelines and ethical frameworks for AI, and even create advisory networks on ethics and AI.[36]

Going forward, governments must grapple with how existing laws can be applied to AI, and also identify the areas, sectors, and use cases where specific regulation is necessary. While proposals for national AI regulators are being made by experts in various jurisdictions,[37] these have not yet been adopted.[38]

Technical Standards

Recognition that human rights should form the basis of technical standards and protocols and the role such standards and protocols play in the exercise of human

34 Japan's Ministry of Economy, Trade and Industry, Japan's Robot Strategy was compiled', 23 January 2015, available from: http://www.meti.go.jp/english/press/2015/0123_01.html. Also see Japan's Robot Policy and the Special Zone for Regulating Next Generation Robots, Robolaw Asia, available from: https://pkurobotlaw.wordpress.com/2015/06/22/japans-robot-policy-and-the-special-zone-for-regulating-next-generation-robots.

35 B. Bergan, Germany Drafts World's First Ethical Guidelines for Self-Driving Cars, Futurism, 25 August 2017, available from: https://futurism.com/germany-drafts-worlds-first-ethical-guidelines-for-self-driving-cars/.

36 L. Kelion, UK PM Seeks 'safe and ethical' artificial intelligence, BBC, 24 January 2018, available from: http://www.bbc.co.uk/news/technology-42810678.

37 See, for example: I. Sample, AI watchdog needed to regulate automated decision making, say experts, The Guardian, 27 January 2017. Also see similar calls in the US: R. Calo, The case for a Federal Robotics Commission, Brookings Institution Center for Technology Innovation, 1 September 2014, available from: https://ssrn.com/abstract=2529151; O. Bracha and F. Pasquale, Federal Search Commission? Access, Fairness, and Accountability in the Law of Search, 93 Cornell L. Rev. 1149, 2008, available from: http://scholarship.law.cornell.edu/clr/vol93/iss6/11.

38 Most recently, in the United Kingdom. M. Burgess, The Government's Report on AI Doesn't Recommend Regulating It, Wired, 14 October 2017, available from: http://www.wired.co.uk/article/ai-report-uk-government-money.

rights is increasingly commonplace.[39] However, it is yet to be seen what this concretely means for AI.

Despite this increased recognition, human rights are not explicitly referred to in the policy processes of many technical or business organisations. These actors are fast becoming the gateways to and facilitators of the exercise of freedom of expression, since they develop the majority of AI systems.

In response to the ethical and legal questions posed by AI, various industry initiatives have been started. The two most prominent are:

- **Global Initiative on Ethics of Autonomous and Intelligent Systems** of the Institute of Electrical and Electronics Engineers (IEEE). This initiative focuses on developing technical standards that embed ethics in AI systems. The initiative also aims to raise awareness in the AI community about the importance of prioritising ethical considerations in the development of technology.

- **The Partnership on Artificial Intelligence to Benefit People and Society,** originally established by Microsoft, Google, Amazon, Facebook, and IBM to 'study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society.'

It is commendable that these initiatives are putting effort and resources into facilitating a cross-disciplinary and, at times, multi-stakeholder discussion. Ethical codes and industry standards can be an important compliance tool[40], or help organisations to go beyond compliance. It is also crucial to ensure that the technical standards and sector guidelines meet international human rights standards on freedom of expression, are accountable, and subject to public scrutiny. However, without strong institutional backing, these codes cannot function effectively, and can even be counterproductive.[41] In all of these, it is crucial to recognise that human rights should be the floor and not the ceiling, and that a number of established minimum principles can serve as guidance.[42]

39 UN Special rapporteur on freedom of expression, Report to the Human Rights Council on Freedom of expression, states and the private sector in the digital age, 2013, available from: http://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/Privatesectorinthedigitalage.aspx.

40 UK Information Commissioner, Discussion paper Big Data, artificial intelligence, machine learning and data protection, available from: https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf.

41 P. Boddington, Towards a Code of Ethics for Artificial Intelligence, Springer International Publishing, 2017.

42 See, the Guiding Principles on Business and Human Rights; the Global Network Initiative Principles on Freedom of Expression and Privacy; the European Commission's ICT Sector Guide on Implementing the UN Guiding Principles on Business and Human Rights; and the Telecommunications Industry Dialogue Guiding Principles. These encourage corporations to commit to protect human rights, undertake due diligence to ensure the positive human rights impact of their work and remediate adverse impacts of their work on human rights.

Self-Regulation By Companies

At the level of individual companies, the likely impact of AI on freedom of expression can be understood by scrutinising their Terms of Service. Research indicates that companies increasingly rely on AI systems in enforcement of their Terms of Service: for instance, platforms currently deploy automated filtering and blocking in response to 'violent extremism' online.[43]

Growing pressure placed on companies by governments to police 'harmful' or illegal content could lead to the development of more sophisticated systems, capable of identifying and removing vast amounts of content with limited, or no, human intervention.

This pressure also incentivises companies to err on the side of caution when it comes to content on their platforms, meaning that their limitations and restrictions on freedom of expression (in particular in the case of social media platforms and electronic payment systems[44]) are sometimes outside the scope of internationally-recognised legitimate limitations on freedom of expression.[45]

For fear of content removal, content-creators themselves then err on the side of caution in their expression, self-censoring and creating a 'chilling effect' on freedom of expression.[46]

The lack of transparency about AI systems used by companies is a problem in this context. This is compounded by the lack of clear complaint mechanisms to deal with inappropriate or overzealous removal of content. Users' ability to challenge these decisions before domestic courts is also extremely limited, which significantly undermines the right of affected parties to seek remedy.[47]

43 See, for example, S. Frenkel, Inside the Obama Administration's Attempt to Bring Tech Companies into the Fight Against ISIS, BuzzFeed, 26 February 2016, available from: https://www.buzzfeed.com/sheerafrenkel/inside-the-obama-administrations-attempt-to-bring-tech-compa?utm_term=.eoN44ER6aM#.xoaJJExkQ8. Also note, removals involve human interaction at various levels. For example, take-down requests originate from public scrutiny, whereas filtering systems are largely automated with human input only at the final stages of review.

44 See, for example, EFF, Free Speech Coalition Calls on PayPal to Back Off Misguided Book Censorship Policy, March 2012, available from: https://www.eff.org/deeplinks/2012/03/free-speech-coalition-callspaypal-back-misguided-book-censorship-policy; or PayPal Rains On Regretsy's Secret Santa Campaign Over Use Of Wrong Button, Consumerist, December 2011.

45 For example, see, K.M. Hovland & D. Seetharaman, Facebook Backs Down on Censoring 'Napalm Girl' Photo, The Wall Street Journal, 9 September 2016, available from: https://www.wsj.com/articles/norway-accuses-facebook-of-censorship-over-deleted-photo-of-napalm-girl-1473428032.

46 ARTICLE 19, Internet Intermediaries: Dilemma of Liability, available at https://www.article19.org/data/files/Intermediaries_ENGLISH.pdf.

47 In particular, the legal basis for any court challenge is contract law, where the standard is generally the lack of fairness of contractual terms, i.e. a very high threshold for consumers. Moreover, social media platforms Terms of Use usually contain jurisdiction clauses forcing users to use the courts in California rather than the courts of their place of residence.

Challenges

While we resist the tendency to construe AI as fundamentally different from preceding technologies, which have been at the forefront of policy debates in recent years, there are a number of unique challenges that AI poses for freedom of expression. They can be classified as follows:

- **Lack of respect for the rule of law:** Current industry initiatives around AI are narrowly focused on the development of technical standards, ethical frameworks, and concepts such as fairness, transparency, and accountability. However, these frameworks must be enforceable and comply with the rule of law. A great deal of the work currently undertaken in this area lacks such enforcement mechanisms, whether self-imposed or through voluntary regulation, limiting impact.

- **Lack of transparency:** Many companies developing critical AI systems do so in ways that are non-transparent and inscrutable. As most of AI is developed or owned by industry, critically engaging with the freedom of expression impact of these technologies is increasingly difficult for civil society actors because of trade secrets rules and high barriers to transparency around application and development, as well as the inherent complexity of these systems.

- **Lack of accountability:** The hidden nature of AI systems makes it difficult to study or analyse the impact of AI on the right to freedom of expression unless a tangible harm occurs. For example, profiling people who take part in protests will become increasingly easy, even if they cover their faces. It is not always clear when machine learning algorithms are being used, so harms arising out of the use of AI are hard to detect. Even when a potential harm is found, it can be difficult to ensure accountability for violations of those responsible.

- **Public perception and the role of the media:** Much of the popular discourse around AI focuses on the dangers of AI general intelligence instead of on current, practical, and realistic implications of AI systems. This discourse has a real impact. It misdirects attention and funding away from current problems surrounding freedom of expression and privacy to favour hypothetical dystopian scenarios. The media has a role to play in ensuring that coverage of AI is focused on the issues at hand.

- **Data collection and use:** Various freedom of expression concerns stem from the way data is collected and used in AI systems. Understanding how data use

47 In particular, the legal basis for any court challenge is contract law, where the standard is generally the lack of fairness of contractual terms, i.e. a very high threshold for consumers. Moreover, social media platforms Terms of Use usually contain jurisdiction clauses forcing users to use the courts in California rather than the courts of their place of residence.

48 S. Walker, Face recognition app taking Russia by storm may bring end to public anonymity, The Guardian, 17 May 2016, available from: https://www.theguardian.com/technology/2016/may/17/findface-face-recognition-app-end-public-anonymity-vkontakte.

and quality influences AI systems is particularly pertinent to front-line defenders of human rights, and vulnerable or minority communities that will be under- or misrepresented in datasets.[49]

49 J. Angwin, et al., Machine Bias, ProPublica, 23 May 2016, available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

AI and Privacy

Establishing the Nexus

The privacy implications of AI stem from its ability to recognise patterns and increasingly 'derive the intimate from the available'.[50] This ability relies on the processing of vast amounts of data.

Different applications and uses of AI can affect the right to privacy in different ways:

- AI-driven consumer products and autonomous systems are frequently equipped with sensors that generate and collect vast amounts of data without the knowledge or consent of those in its proximity;
- AI methods are being used to identify people who wish to remain anonymous;
- AI methods are being used to infer and generate sensitive information about people from their non-sensitive data;
- AI methods are being used to profile people based upon population-scale data; and
- AI methods are being used to make consequential decisions using this data, some of which profoundly affect people's lives.

Each of these novel interferences with privacy are significant: privacy is indispensable for the exercise of a range of human rights, such as freedom of expression, freedom of association, as well as being fundamental for the exercise of personal autonomy and freedom of choice,[51] as well as broader societal norms.[52]

49 J. Angwin, et al., Machine Bias, ProPublica, 23 May 2016, available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

50 Calo, Artificial Intelligence Policy, op.cit.

51 T. Payton and T. Claypoole, Privacy in the age of Big data: Recognizing threats, defending your rights, and protecting your family, Rowman & Littlefield, 2014.

52 R.C.Post,., 'The social foundations of privacy: Community and self in the common law tort', California Law Review, 1989, pp. 957-1010. Summarizing Post see T. Doyle, 2012; D. J. Solove, Nothing to Hide: The False Tradeoff between Privacy and Security, Yale University Press, 2011: 'As the legal theorist Robert Post has argued, privacy is not merely a set of restraints on society's rules and norms. Instead, privacy constitutes a society's attempt to promote civility. Society protects privacy as a means of enforcing order in the community. Privacy isn't the trumpeting of the individual against society's interests but the protection of the individual based on society's own norms and values.'

- **Data exploitation:** Consumer products, from smart home appliances to connected cars and phone applications, are often built for data exploitation. Consumers are commonly faced with an informational asymmetry as to what kinds and how much data their devices, networks, and platforms generate, process, or share. As we bring ever-more smart and connected devices into our homes, workplaces, public spaces, and even our bodies, educating the public about such data exploitation becomes increasingly pressing.

- **Identification and tracking:** AI applications can be used to identify and thereby track individuals across different devices, in their homes, at work, and in public spaces. For example, while personal data is routinely (pseudo-) anonymised within datasets, AI can be employed to de-anonymise this data, challenging the distinction between personal and non-personal data, on which current data protection regulation is based.[53] Facial recognition is another means by which individuals can be tracked and identified, which has the potential to transform expectations of anonymity in public space. Machine learning systems have even been able to identify around 69% of protesters wearing caps and scarves to cover their faces.[54] In the context of law enforcement, facial recognition can allow the police to identify individuals without probable cause, reasonable suspicion, or any other legal standard that might otherwise be required for law enforcement to obtain identification by traditional means.[55]

- **Inference and prediction of information:** Using machine learning methods, highly sensitive information can be inferred or predicted from non-sensitive forms of data. People's emotional states e.g. confidence, nervousness, sadness, and tiredness, can be predicted from typing patterns on a computer keyboard.[56] When sensitive personal data, such as information about health, sexuality, ethnicity, or political beliefs can be predicted from unrelated data (i.e. activity logs, phone metrics, location data or social media likes) such profiling poses significant challenges to privacy and may result in discrimination.

53 A 2015 study by researchers at the French Institute for Research in Computer Science showed that 75% of mobile phone users can be re-identified within a dataset using machine learning methods and just two smartphone apps, with the probability rising to 95% if four apps are used. See J.P. Achara,,G. Acs, and C. Castelluccia,'On the unicity of smartphone applications' at the 14th ACM Workshop on Privacy in the Electronic Society, October 2015, pp. 27-36, available at https://arxiv.org/pdf/1507.07851v2.pdf.

54 Walker, 'Face recognition app', op.cit.

55 Electronic Privacy Information Center (EPIC) and 45 organisations, Letter to Senators Grassley and Leahy and Representatives Goodlatte, Chaffetz, Conyers, and Cummings regarding the FBI's Use of Facial Recognition and Proposal to Exempt the Bureau's Next Generation Identification Database from Privacy Act Obligations, 2016, available from: https://epic.org/privacy/fbi/NGI-Congressional-Oversight-Letter.pdf.

56 C. Epp, M. Lippold & R.L. Mandryk, Identifying emotional states using keystroke dynamics' in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems., May 2011, pp. 715-724, available from http://hci.usask.ca/uploads/203-p715-epp.pdf.

- **Profiling to sort, score, categorise, assess and rank individuals and groups:** AI-driven applications are used to automatically sort, score, categorise, assess and rank people, often without their knowledge or consent, and frequently without the ability to challenge the outcomes or effectiveness of those processes. In 2016, for instance, IBM promoted the use of AI to separate 'genuine' refugees from other types of migrants.[57] Machine learning also plays a growing role in scoring systems which shape access to credit, employment, housing or social services.

- **Decision-making:** AI systems can be used to make or inform decisions about people or their environments, potentially based on profiling. An environment that knows your preferences and adapts itself according to presumed interests raises important issues around privacy, autonomy and the ethics of such adaptations. Personalisation, not only of information but also of our perception of the world, will become increasingly important as we move towards connected spaces like smart cities[58] or augmented and virtual reality (AR and VR).

57 P. Tucker, Refugee or Terrorist? IBM thinks its software has the answer, Defense One, 27 January 2016, available at http://www.defenseone.com/technology/2016/01/refugee-or-terrorist-ibm-thinks-its-software-has-answer/125484/.

58 Privacy International, 'Smart Cities: Utopian Vision, Dystopian Reality', October 2017, available at https://privacyinternational.org/sites/default/files/2017-12/Smart%20Cities-Utopian%20Vision%2C%20Dystopian%20Reality.pdf.

Mapping the Landscape

Policy debates around AI and privacy are complicated by the fact that regulatory and policy discourses use the term to refer to a broad range of applications, usages and methods. Where AI is discussed in such a broad way, there is a tendency to assume that the technology poses challenges that are so radically new that all existing laws, regulations and standards are no longer applicable or appropriate. The 'flipside' of that discourse is to demand regulation of the technology itself, regardless of how and where it is applied.

To avoid succumbing to any of these fallacies, there is a need to examine how existing discourses, such as human rights law, data protection, sectoral privacy regulation, and research ethics, relate to different applications and methods of AI.

Below, we explain how several of these existing frameworks apply and where they fall short. We also discuss various AI-specific initiatives which have an explicit privacy focus, some of which are sectoral, others are more general.

International Human Rights Law

International human rights law recognises the fundamental right to privacy. Article 12 of the Universal Declaration of Human Rights (UDHR), for instance, proclaims that "[n]o one shall be subjected to arbitrary interference with his privacy, family, home or correspondence …. Everyone has the right to the protection of the law against such interference or attacks." [59]

International human rights law requires that any interference with the right to privacy must not only be in accordance with law[60] but must also be necessary and proportionate.[61] To the extent that states develop or use AI in a manner that interferes

59 G.A. Res. 217 (III) A, UDHR, art. 12 (Dec. 10, 1948)

60 See Article 17(1), ICCPR ; Article 11, ACHR ("2. No one may be the object of arbitrary or abusive interference with his private life, his family, his home, or his correspondence .. 3. Everyone has the right to the protection of the law against such interference . . . ."); Article 8(2) of the European Convention of Human Rights ("ECHR") ("There shall be no interference by a public authority with the exercise of [the right to respect for private and family life] except such as is in accordance with the law …"); see also U.N. Human Rights Committee, General Comment No. 16 (Article 17 ICCPR), 8 Apr. 1988, para 3, available at http://tbinternet.ohchr.org/Treaties/CCPR/Shared%20Documents/1_Global/INT_CCPR_GEC_6624_E.doc (noting that "[t]he term 'unlawful' means that no interference can take place except in cases envisaged by the law" and that "[i]nterference authorized by States can only take place on the basis of law, which itself must comply with the provisions, aims and objectives of the Covenant"); Principle 1, International Principles on the Application of Human Rights to Communications Surveillance ("Necessary and Proportionate Principles"), available at https://necessaryandproportionate.org/principles. The Necessary and Proportionate Principles apply international human rights law to modern digital surveillance. They were drafted in 2013 by an international coalition of civil society, privacy and technology experts and have been endorsed by over 600 organizations around the world.

with the right to privacy, that use must be subject to the three-part test of legality, necessity and proportionality.

International human rights law requires that any interference with the right to privacy must not only be in accordance with law[60] but must also be necessary and proportionate.[61] To the extent that states develop or use AI in a manner that interferes with the right to privacy, that use must be subject to the three-part test of legality, necessity and proportionality.

Advocates and authorities using the international human rights framework are increasingly recognising and acknowledging the impact that new forms of data-processing have on fundamental rights, including the right to privacy. With respect to profiling, for example, which may involve the use of AI methods to derive, infer or predict information about individuals for the purpose of evaluating or assessing some aspect about them, the United Nations Human Rights Council noted with concern in March 2017 that 'automatic processing of personal data for individual profiling may lead to discrimination or decisions that otherwise have the potential to affect the enjoyment of human rights, including economic, social and cultural rights.' [62]

International human rights authorities have also moved towards recognising a right to anonymity under the rights to privacy and freedom of opinion and expression. This has implications for AI used to identify individuals online, in their homes and in public spaces. The UN Special Rapporteur on Freedom of Expression, for instance, has repeatedly identified this relationship and emphasised that state interference with anonymity should be subject to the three-part test of legality, necessity, and proportionality, as is any other interference with these rights.[63]

Data Protection

Data protection frameworks apply to research, development and application of AI to

61 See U.N. Human Rights Committee, Toonen v. Australia, Comm. No. 488/1992, U.N. Doc. CCPR/C/50/D/488/1992 (31 Mar. 1994), para. 8.3 ("[A]ny interference with privacy must be proportional to the end sought and be necessary in the circumstances of any given case."); Office of the U.N. High Commissioner for Human Rights, The Right to Privacy in the Digital Age, U.N. Doc. A/HRC/27/37 (30 June 2014), para. 23, available at https://documents-dds-ny.un.org/doc/UNDOC/GEN/G14/088/54/PDF/G1408854.pdf?OpenElement ("These authoritative sources [HRC General Comments 16, 27, 29, 31, and 34 and the Siracusa Principles] point to the overarching principles of legality, necessity and proportionality... ."); U.N. Human Rights Council Resolution on the Right to Privacy in the Digital Age U.N. Doc. A/HRC/34/7, 23 Mar. 2017, para. 2 available at https://documents-ddsny.un.org/doc/UNDOC/GEN/G17/086/31/PDF/G1708631.pdf?OpenElement ("Recall[ing] that States should ensure that any interference with the right to privacy is consistent with the principles of legality, necessity and proportionality").

62 U.N. Human Rights Council Resolution on the Right to Privacy in the Digital Age, U.N. Doc. A/HRC/34/L,7, 23 Mar. 2017, para 2.

63 Office of the U.N. High Commissioner for Human Rights, Report on encryption, anonymity, and the human rights framework, U.N. Doc. A/HRC/29/32 (22 May 2015), available at http://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/CallForSubmission.aspx.

the extent that personal data (as defined in the frameworks) is involved.[64] Thus, even without explicit reference to AI, data protection frameworks already regulate how AI systems can process personal data. Regulatory frameworks around the world are diverse, but are all designed to protect individuals' data and reflect a sense that such protections are an important aspect of the right to privacy.

The EU General Data Protection Regulation (GDPR), which takes effect on 25 May 2018, requires a legal basis for processing data - and in addition to the principles of fairness, accountability and transparency, includes the core principles of purpose limitation and data minimisation,[65] which have implications for the development, use and application of AI systems.

The GDPR also limits the use of automated decision-making in certain circumstances, and requires individuals to be provided with information as to the existence of automated decision-making, the logic involved and the significance and envisaged consequences of the processing for the individual.[66] The law introduces an overall prohibition (with narrow exceptions) to solely automated decisions when such decisions have legal or other significant effects.[67]

Notably, the GDPR also defines profiling as the automated processing of data to analyse or to make predictions about individuals.[68] This definition recognises that personal data can be produced by machine learning applications and other forms of profiling.[69]

Finally, the GDPR introduces a range of provisions which encourage the design of less privacy-invasive systems, some of which have far reaching consequences for AI. The obligation to incorporate data protection by design and by default seeks to integrate data protection principles into the design of data processing operations.[70]

Data Privacy Impact Assessments (DPIA) - tools for organisations to manage privacy risks - will be mandatory for many privacy-invasive AI and machine learning

64 See for instance UK Information Commissioner, Discussion paper 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 or the European Data Protection Supervisor (EDPS)'s Room Document for the 38th International Conference of Data Protection and Privacy Commissioners, Artificial Intelligence, Robotics, Privacy, and Data Protection, available at https://edps.europa.eu/sites/edp/files/publication/16-10-19_marrakesh_ai_paper_en.pdf.

65 Article 5 of the General Data Protection Regulation (GDPR)- (Regulation (EU) 2016/679.

66 Articles 13, 14 and 22 of GDPR.

67 Article 22 of GDPR only applies to decisions "based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her".

68 Article 25 GDPR.

69 Article 4(4) GDPR.

70 See, for example, R. Binns, 'Data protection impact assessments: a meta-regulatory approach' International Data Privacy Law, 7(1), 2017, pp 22-35; L. Edwards, & M. Veale, 'Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'?', IEEE Security & Privacy, 2017.

applications that fall within the scope of data protection law and come with substantial anticipated risks, such as the processing of sensitive data.

Data protection plays a crucial role in safeguarding the right to privacy[71] but cannot address all privacy risks that arise from different applications and uses of AI. Data protection is limited to the protection of data that relates to an identified or identifiable person (even indirectly). That does not cover the privacy of groups, or other infringements on privacy that do not necessarily involve personal data.[73]

While provisions like those in the GDPR that deal with profiling and automated decision-making are crucial, they will only affect some uses of AI in automated decision-making[74] or profiling.[75] Additionally, data protection frameworks also frequently have exemptions for national security, limiting rights and safeguards in crucial privacy-invasive applications of AI, e.g. government surveillance.

Sectoral Privacy Regulation

In countries with data protection frameworks, sectoral privacy regulation complements data protection. The proposed ePrivacy Regulation in the EU, for instance, covers the privacy and confidentiality of communications and, as such, has implications for AI-driven consumer products, such as digital assistants. French administrative law gives a right to an explanation for administrative algorithmic decisions made about individuals.[76] This provision is broader and more comprehensive than GDPR provisions on automated decision-making but only applies to administrative decisions.[77]

Sectoral privacy regulation also plays an important role in jurisdictions which do not have a general data protection framework, such as the United States, where all applications of AI have to comply with existing laws like the US Health Insurance Portability and Accountability Act of 1996 (HIPPA).[78] The city of New York introduced legislation which will establish a taskforce to examine the city's 'automated decision systems' in order to make them fairer and more open to scrutiny. This will apply to

71 In 2011, the U.N. Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression issued a report similarly noting that 'the protection of personal data represents a special form of respect for the right to privacy.' U.N. Doc. A/HRC/17/27, ¶ 58 (May 16, 2011).

72 On the difficulty of data protection laws to protection groups see Taylor, L., Floridi, L., & van der Sloot, B. (Eds.), Group privacy: New challenges of data technologies, vol. 126, 2016, Springer.

73 An example would be automated lip reading systems if applied to images of people in public or a crowd. See Veale, M., Edwards, L., Bear, H. (draft, Jan 2018 for PLSC Europe). Better seen and not (over)heard? Automated lipreading systems and privacy in public spaces.

74 F.Kaltheuner, & E. Bietti, 'Data is power: Towards additional guidance on profiling and automated decision-making in the GDPR', Journal of Information Rights, Policy and Practice, vol 2(2), 2018.

75 See for instance Hildebrandt, M. and Koops, B.J., 'The challenges of ambient law and legal

76 protection in the profiling era', The Modern Law Review, 73(3), 2010, pp.428-460.

77 Loi pour une République numérique (Digital Republic Act, Loi n 2016-132).

78 Edwards & Veale, 'Enslaving the Algorithm', op.cit.

computerised algorithms which guide the allocation of everything from police officers and firehouses to public housing and food stamps.

Sectoral regulation can also play an important role in addressing more context- and domain-specific challenges of AI, for instance autonomous cars. However, not all existing sectoral privacy regulations are effective in protecting people from the new threats to privacy posed by AI applications. Many alternative credit-assessment tools that rely on machine learning methods for scoring, for instance, have been able to avoid coverage under the US Fair Credit Reporting Act (FCRA).[79]

AI can undermine the effectiveness of purely sectoral regulation data. For instance, even strong regulation of medical records typically does not address the fact that health data can be derived, inferred or predicted from browsing histories or credit card data.

Ethical Codes and Industry Standards

Industry initiatives, standards bodies, and governments are currently developing ethical codes on AI, some of which are general, others sectoral.

The Global Initiative on Ethics of Autonomous and Intelligent Systems of the IEEE, for instance, has dedicated a section relevant to privacy on 'Personal Data and Individual Access Control in Ethically Aligned Design'.

The German Ethics Code for Automated and Connected Driving is an example of a sectoral ethic code that also contains a specific principle on data privacy which addresses the tension between business models that are based on the data generated by automated and connected driving, and limitations to the autonomy and data sovereignty of users.[80]

While many ethical challenges are distinctive to AI or its use in a particular domain or context, some are not necessarily unique to it. For instance, there is a rich literature on business and human rights,[81] as well as the ethics of big data research,[82] some of

79 M. Hurley, & J. Adebayo, 'Credit Scoring in the Era of Big Data', Yale JL & Tech., 18, 2016, p. 148.

80 Permitted business models that avail themselves of the data that are generated by automated and connected driving and that are significant or insignificant to vehicle control come up against their limitations in the autonomy and data sovereignty of road users. It is the vehicle keepers and vehicle users who decide whether their vehicle data that are generated are to be forwarded and used. The voluntary nature of such data disclosure presupposes the existence of serious alternatives and practicability. Action should be taken at an early stage to counter a normative force of the factual, such as that prevailing in the case of data access by the operators of search engines or social networks.' See Federal Ministry of Transport and Digital Infrastructure, Ethics Commission, Automated and Connected Driving, June 2017, available at https://www.bmvi.de/SharedDocs/EN/Documents/G/ethic-commission-report.pdf?__blob=publicationFile.

which can be informative to the privacy risks of AI as well. It is important to note that the international human rights framework is also relevant to non-state use of AI.[83]

81 UN Guiding principles on business and human rights: implementing the United Nations "Protect, Respect and Remedy" framework, 2011.

82 J. Metcalf, & K. Crawford, 'Where are human subjects in big data research? The emerging ethics divide', Big Data & Society, 3(1), 2016p.2053951716650211; Zook, M. et al, 'Ten simple rules for responsible big data research', PLoS computational biology, 13(3), 2017, p. e1005399.

83 As the UN Special Rapporteur on Freedom of Expression recognised in 2015: "Corporations in a variety of sectors play roles in advancing or interfering with privacy, opinion and expression, including ... anonymity. … [I]t remains important to emphasize that "the responsibility to respect human rights applies throughout a company's global operations regardless of where its users are located, and exists independently of whether the State meets its own human rights obligations", see Kaye, D, 2015. A/HRC/29/32, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression. UN General Assembly, May 22.

Challenges

1. **The diversity of AI applications, systems, and uses:** Different types of AI and different domains of application raise distinct ethical and regulatory privacy concerns. For instance, processing data generated by autonomous cars raise different privacy challenges than the use of machine learning to identify 'terrorist' suspects. This lack of definitional clarity is a challenge, since different types of AI and different domains of application raise distinct ethical and regulatory issues.

2. **Informational asymmetry:** Individuals are commonly unable to fully understand what kinds and how much data their devices, networks, and platforms generate, process, or share. As we bring ever-more smart and connected devices into people's homes, workplaces, public spaces and even bodies, the need to educate the public about such data exploitation becomes increasingly pressing. In this landscape, uses of AI for purposes like profiling, or to track and identify people across devices and even in public places, amplify this asymmetry.

3. **Opacity and secrecy of profiling:** Some applications of AI can be opaque to individuals, regulators, or even the designers of the system themselves, making it difficult to challenge or interrogate outcomes. In this context it is important to distinguish between three sources of opacity: (1) opacity as intentional corporate or state secrecy; (2) opacity as technical illiteracy; and (3) opacity that arises from the characteristics of machine learning algorithms and the scale required to apply them usefully.[84] While there are technical solutions to improving the interpretability or auditability of some systems for solutions to improving the interpretability or auditability of some systems for different stakeholders,[85] a key challenge remains where this is not possible and where negative outcomes are either safety-critical or human-rights-critical.

81 UN Guiding principles on business and human rights: implementing the United Nations "Protect, Respect and Remedy" framework, 2011.

82 J. Metcalf, & K. Crawford, 'Where are human subjects in big data research? The emerging ethics divide', Big Data & Society, 3(1), 2016p.2053951716650211; Zook, M. et al, 'Ten simple rules for responsible big data research', PLoS computational biology, 13(3), 2017, p. e1005399.

83 As the UN Special Rapporteur on Freedom of Expression recognised in 2015: "Corporations in a variety of sectors play roles in advancing or interfering with privacy, opinion and expression, including ... anonymity. … [I]t remains important to emphasize that "the responsibility to respect human rights applies throughout a company's global operations regardless of where its users are located, and exists independently of whether the State meets its own human rights obligations", see Kaye, D, 2015. A/HRC/29/32, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression. UN General Assembly, May 22.

84 J. Burrell, 'How the Machine 'thinks': Understanding Opacity in Machine Learning Algorithms', Big Data and Society, 3(1), 2016.

85 A. Datta, S. Sen, & Y. Zick, 'Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems', In Security and Privacy (SP), 2016 IEEE Symposium, pp. 598-617.

4. **Discrimination, unfairness, inaccuracies, bias**: AI-driven identification, profiling and automated decision-making may lead to unfair, discriminatory, or biased outcomes.[86] Individuals can be misclassified, misidentified, or judged negatively, and such errors or biases may disproportionately affect certain groups of people. Accurate predictions may reveal sensitive attributes that could be used to discriminate. On the other hand, inaccurate or systematically biased data can feed into profiles, which may lead to biased or discriminatory outcomes.

5. **Re-identification and de-anonymisation:** Some applications of AI, in particular uses of machine learning, blur the line between personal and non-personal data [or personally identifiable information (PII) and non-personally identifiable information (non-PII) in the US], around which data protection and privacy laws around the world are organised. Data that is initially non-personal (non-PII) can become personal data (PII) in a different context or in different points in time, which is a particular risk for sectoral regulations. A similar challenge applies to sensitive personal data. Profiling using machine learning can derive, infer, or predict sensitive information from non-sensitive data, which might undermine additional safeguards for sensitive personal data.

86 S. Barocas, & A. Selbst, 'Big data's disparate impact', Cal. L. Rev., 104, 2016, p. 671.

Conclusions and Recommendations

ARTICLE 19 and Privacy International support the development and use of AI in compliance with human rights standards and regulatory standards in their respective fields. As AI systems become increasingly integrated into a larger number of critical societal processes, policy and technology responses in this area must meet the recommendations set out in this paper.

We have provided an initial overview of the impact of these technologies on the freedom of expression and privacy, mapping the regulatory landscape, and delineating the roles, responsibilities, and duties that accrue to various actors in the field. Beyond this overview, we hope this paper provides a concrete step towards building strong civil society networks for action and advocacy to ensure that the organisations using, building and governing AI are held accountable and meet international human rights standards.

As indicated in this paper, we believe that it is necessary to further study and monitor the impact of AI on human rights. However, at this stage, we **call on states** to:

- **Review the adequacy of existing frameworks and regulation:** Different types of AI and different domains of application raise specific ethical and regulatory human rights issues. In order to ensure that they protect individuals from the risks posed by AI, existing laws must be reviewed, and if necessary amended, to address the effects of new and emerging threats to privacy and freedom of expression.

We also call on states and companies to:

- **Ensure protection of international human rights standards:** The development, use, research and development of AI must be subject to the minimum requirement of respecting, promoting, and protecting international human rights standards. This should include developing an understanding of what constitutes 'AI human rights critical systems' and ensuring that laws and regulations, codes of conduct, ethical codes, and self-regulatory and technical standards meet the threshold set by international human rights.

- **Ensure accountability and transparency:** Corporate, technical, and state actors must allow for meaningful multi-stakeholder participation, including civil society actors, in setting technical standards, regulation, and industry guidelines for AI systems, technology policy and industry standards to ensure transparent processes and legitimacy of outcomes. In particular, non-binding frameworks must be accompanied by strong accountability and oversight measures.

We also call on civil society to:

- **Engage further** to ensure the mitigation of any potential negative impact on fundamental rights like freedom of expression and privacy. This will involve a detailed understanding of the technology, the actors developing it, and the context in which it is deployed.

- **Collect and highlight case studies of 'human rights critical' AI:** In understanding the myriad ways in which AI will impact human rights, it is important to collect and highlight case studies that demonstrate impact. These case studies need to include examples from across the globe.

- **Build civil society coalitions and expertise networks:** It is important to emphasise the need to develop knowledge-exchange programs and facilitate joint-strategy development between civil society organisations. So far, academia and industry have taken the lead in moving the debate on the societal impact of AI forward. While civil society actors play a crucial role in these debates, it is important to strengthen the voice of those working on technology in the public interest.

PRIVACY INTERNATIONAL
62 Britton Street
London EC1M 5UY
United Kingdom

Website: www.privacyinternational.org
Phone: +44 20 3422 4321
Twitter: @privacyint

Registered charity number 1147471

ARTICLE 19
Free Word Centre
60 Farringdon Rd
London EC1R 3GA
United Kingdom

Website: www.article19.org
Phone: +44 20 7324 2500
Twitter: @article19org

Registered charity number 327421