doi:10.1017/aju.2020.30

SYMPOSIUM: HOW WILL ARTIFICIAL INTELLIGENCE AFFECT INTERNATIONAL LAW?

USING HUMAN RIGHTS LAW TO INFORM STATES' DECISIONS TO DEPLOY AI

*Daragh Murray**

States are investing heavily in artificial intelligence (AI) technology,[1] and are actively incorporating AI tools across the full spectrum of their decision-making processes. However, AI tools are currently deployed without a full understanding of their impact on individuals or society, and in the absence of effective domestic or international regulatory frameworks.[2] Although this haste to deploy is understandable given AI's significant potential, it is unsatisfactory. The inappropriate deployment of AI technologies risks litigation, public backlash, and harm to human rights. In turn, this is likely to delay or frustrate beneficial AI deployments. This essay suggests that human rights law offers a solution. It provides an organizing framework that states should draw on to guide their decisions to deploy AI (or not),[3] and can facilitate the clear and transparent justification of those decisions.

However, using human rights law to inform states' decision-making processes is not straightforward. Although human rights law imposes (essential) ex ante obligations, our understanding of how that law applies, and the content of specific obligations in specific contexts, is primarily derived from ex post accountability mechanisms. These understandings do not apply straightforwardly to ex ante processes.[4] A change in thinking is required if human rights law is to more effectively inform decision-making processes from the outset: this essay unpacks core human rights law features to identify relevant guidance. For reasons of space, it focuses on how states can determine if AI deployments are "necessary."

The approach discussed here is relevant to AI generally, but this essay uses live facial recognition (LFR) as an example throughout to move the conversation from the abstract to the practical. LFR also effectively illustrates the broader issues. In the United Kingdom, citizens have challenged South Wales' Police use of LFR before the High

* *Senior Lecturer, University of Essex.* *Thanks to Ashley Deeks for invaluable editorial comments and suggestions. This work was supported by the Economic Social and Research Council, grant number ES/M010236/1.*

[1] This essay uses the term "AI" broadly, to include techniques such as machine learning and data analytics.

[2] This is demonstrated, for example, by the fact that states are only beginning to develop national AI strategies. These are typically loosely formulated, with a focus on broad principles or ethical considerations, as opposed to legal regulation.

[3] Lorna McGregor et al., *International Human Rights as a Framework for Algorithmic Accountability*, 68(2) Int'l & Comp. L.Q. 309 (2019). Although not addressed in detail here, human rights law provides a universally accepted definition of "harm," a means of identifying whether such harm is permissible or not, a means of determining the responsibilities of relevant state and business actors, established accountability mechanisms, and "red lines" indicating where AI should not be deployed.

[4] There are limited exceptions, and the due diligence requirements associated with the UN Guiding Principles on Business and Human Rights provide a case on point. However, this is also a new area, and the precise steps required to facilitate compliance are being worked out. *See, e.g.*, Tara L. Van Ho & Mohammed K. Alshaleel, *The Mutual Fund Industry and the Protection of Human Rights*, 18(1) Hum. Rts. L. Rev. 1 (2018).

Court, and an appeal is currently pending.[5] Public backlash has produced calls for a moratorium or ban on the use of this technology.[6] In the United States, a number of cities have put in place just such a ban.

Although derived from human rights law, this approach should be of practical use to all states, irrespective of their status of treaty ratification or level of human rights engagement. By examining *why* a deployment is "necessary," and *what* alternative approaches are available, states can more clearly explain their intentions, act more transparently, and better engage with any subsequent challenges and debates, legal or otherwise.

Obligation to Respect and Non-Arbitrariness

Two core human rights law components are relevant when states consider how to approach the decision to deploy AI. First, the law establishes an "obligation to respect," requiring states to refrain from taking action that will result in a human rights violation.[7] Second, a central objective of the law is to protect individuals against arbitrary rights interferences.[8] This requires clarity and certainty vis-à-vis the scope of state authority. To protect against arbitrariness and determine the legitimacy of any deployment, states should typically conduct a three-part test. The measure in question should: (a) be in accordance with the law, (b) pursue a legitimate aim, and (c) be necessary in a democratic society.[9] These features in turn require states to conduct a pre-deployment impact assessment. This is not an explicit human rights law requirement, but it is implicit: if states must ensure that their activities do not result in human rights violations, they must identify the potential impact of those activities.[10] This essay focuses on the "necessary in a democratic society" test. Case law is derived primarily from the European Court of Human Rights, as these issues have been addressed in detail there, but the conclusions remain broadly relevant both to the International Covenant on Civil and Political Rights and other regional human rights treaties.

Determining Whether an AI Deployment Is "Necessary in A Democratic Society"

The "necessary in a democratic society" test is intended to ensure the overall rights compliance of any measure. It addresses the "competing interests" arising in particular contexts. For example, a particular measure—such as AI-assisted surveillance—may be useful for the prevention and detection of crime, but pose risks to privacy, including of individual stigmatization.[11] These are the "competing interests" at play. In resolving these interests,

[5] *See Bridges v. Chief Constable of South Wales Police* [2019] EWHC 2341 (Admin), Sept. 4, 2019.

[6] *See* Carly Kind, *Biometrics and Facial Recognition Technology – Where Next?*, Ada Lovelace Instit. (July 2, 2019); Liberty, Resist Facial Recognition.

[7] *See* International Covenant on Civil and Political Rights art. 2(1), Dec. 16, 1966, 999 UNTS 171. This requirement also appears in all regional human rights treaties.

[8] With respect to the International Covenant on Civil and Political Rights, this is demonstrated by reference to the prohibition of arbitrary deprivation of life (Article 6) or liberty (Article 9) or interference with privacy (Article 17), and the limitations clauses established in relation to Articles 18, 19, 21, and 22. Most deployments of AI in decision-making roles will implicate a number of human rights protections.

[9] This formulation is most closely associated with the limitations clauses established in relation to rights such as the right to privacy, the right to freedom of expression, etc. *See* Szabo and Vissy v. Hungary, App. No. 37138/14, para. 54 (Eur. Ct. H.R., Jan. 12, 2016). However, it has also been applied to protect against arbitrary interferences vis-à-vis other rights, such as the right to life. *See* Human Rights Comm., General Comment No. 36, UN Doc. CCPR/C/GC/36, paras. 10-12 (Sept. 3, 2019); Human Rights Comm., General Comment No. 34, UN Doc. CCPR/C/GC/34, para. 22 (Sept. 12, 2011).

[10] In relation to the need for impact assessments in the context of live facial recognition, see Surveillance Camera Commissioner, Police Use of Automated Facial Recognition Technology with Surveillance Camera Systems: Section 33 Protection of Freedoms Act 2012, at 9 (Mar. 2019).

[11] *See* S. and Marper v. the United Kingdom, App. Nos. 30562/04 & 30566/04, paras. 112, 122, 125 (Eur. Ct. H.R., Dec. 4, 2008).

the state must identify both the potential utility and the potential harm of any deployment, in light of the constraints of a democratic society.[12]

Application of the "necessary in a democratic society" test involves a number of different elements. An interference may meet this test if it remains faithful to democratic principles,[13] "if it answers to a 'pressing social need', if it is proportionate to the legitimate aim pursued and if the reasons adduced by the national authorities to justify it are [']relevant and sufficient.'"[14] Applied to a state's decision to deploy an AI tool, these may be distilled into two central criteria: First, why is a deployment required, and second, what alternative mechanisms are available?

Why Is an AI Deployment Required?

A number of factors are relevant to clarifying why a particular AI deployment is required: (a) identifying the objective underpinning the deployment, (b) demonstrating why achieving that objective is necessary, and (c) specifying how the technology will be deployed. These components are a means of establishing purpose, thereby facilitating identification of utility and harm, and are central to ensuring foreseeability and protecting against arbitrary rights interferences. It is necessary that states undertake this process prior to any potential deployment—and that a record be maintained—so that a "pressing social need" can be demonstrated, and "relevant and sufficient" justifications for deployment presented and "convincingly established."[15] It is also a means of protecting against "mission creep," whereby a tool is deployed for a particular purpose, but is then used to achieve other objectives over time.[16] Adaptation of objectives will require fresh analysis, limiting creep.

Identifying the objective underpinning an intended deployment is a first step.[17] This should be done at a granular level, rather than in the abstract.[18] Using the LFR example, an objective of "preventing crime and protecting public order"—a legitimate aim—is overly broad: it is essentially reflective of all policing activity, and so does not provide any foreseeability as to the specific activities that state actors will undertake.[19] Examples of more focused objectives may include the identification of individuals suspected of belonging to proscribed terrorist organizations at border posts so that they may be stopped or questioned, or the identification of individuals subject to outstanding arrest warrants as they pass through a particular part of a city.

Once the state identifies the objective, it must then demonstrate why achieving that objective is necessary. This is relevant to determining specific utility, and demonstrating "a pressing social need." In the LFR context this may

[12] *See, e.g.*, Klass and Others v. Germany, App. No. 5029/71, para. 55 (Eur. Ct. H.R., Sept. 6, 1978) [hereinafter Klass]; Von Hannover v. Germany (No. 2), App. Nos. 40660/08, 60641/02, para. 101 (Eur. Ct. H.R., Feb. 7, 2012).

[13] *See* Klass, *supra* note 12, at para. 55.

[14] Catt v. the United Kingdom, App. No. 43514/15, para. 109 (Eur. Ct. H.R., Jan. 24, 2019); Segerstedt-Wiberg and Others v. Sweden, App. No. 62332/00, para. 88 (Eur. Ct. H.R., June 6, 2006); Vogt v. Germany, App. No. 17851/91, para. 52 (Eur. Ct. H.R., Sept. 26, 1995); Observer and Guardian v. the United Kingdom, App. No. 13585/88, para. 59 (Eur. Ct. H.R., Nov. 26, 1991).

[15] Autronic AG v. Switzerland, App. No. 12726/87, para. 61 (Eur. Ct. H.R., May 22, 1990); Weber v. Switzerland, App. No. 11034/84, para. 47 (Eur. Ct. H.R., May 22, 1990).

[16] *See generally* Gary Marx, Undercover: Police Surveillance in America (1989).

[17] Although this would seem to be a common-sense measure, it appears that AI deployments are often undertaken without detailed planning. *See* Pete Fussey & Daragh Murray, Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology 46-61 (The Human Rights, Big Data and Technology Project, July 2019).

[18] *See, e.g.*, Szabo and Vissy v. Hungary, App. No. 37138/14, para. 65 (Eur. Ct. H.R., Jan. 12, 2016). *See also* Catt v. the United Kingdom, App. No. 43514/15, para. 114 (Eur. Ct. H.R., Jan. 24, 2019).

[19] In *Catt*, the Court expressed "concerns about the ambiguity of the legal basis for the collection of the applicant's personal data," suggesting the need for further specificity. Catt v. the United Kingdom, App. No. 43514/15, para. 105 (Eur. Ct. H.R., Jan. 24, 2019).

relate to the nature of the crime, or the threshold for initiating surveillance: the social needs associated with preventing murder will be much higher than those associated with detecting petty theft.[20] "Relevant and sufficient reasons" are critical. Building on the arrest warrant example, it may be necessary to specify whether it is difficult to contact individuals subject to an arrest warrant, and whether this applies generally to all warrants or is restricted to specific offences.

A next step is to specify the circumstances of deployment. This is essential to evaluating impact—both in terms of utility and harm—and gives effect to the previous two steps. A number of factors are potentially relevant. For instance, will the AI deployment run for a set period/at particular intervals, or on a more continuous long-term basis; will the data produced be subject to further AI-driven analysis; and who has access to the resulting data, and under what circumstances? Clarity around the intended circumstances of use is important both to understand how a particular deployment will run (facilitating foreseeability) and what the potential human rights-related impact of that deployment may be. In the terrorism example, the definition of a proscribed terrorist organisation is likely to be established in law, thereby reducing the scope for arbitrariness. However, specificity may be necessary with respect to the criteria used to enroll individuals on the associated watchlist. For instance, if police intend to stop individuals on the basis of membership—or suspected membership—in such an organization, will this occur following a specified process with a required intelligence or evidentiary threshold, or on the basis of some other arrangement?

Identifying Alternative Mechanisms

The other element to demonstrating the utility of an AI deployment is a consideration of alternative, or preexisting, mechanisms. This element speaks to the "why AI" question, and helps to determine whether the state could use other, less invasive, approaches to achieve the same—or sufficiently similar—objectives. This assessment contributes to the proportionality assessment, which must evaluate "whether it is possible to achieve the aims by less restrictive means."[21]

The examples presented previously are helpful in unpacking some of the issues. In the first example, LFR technology was used at border ports in order to identify individuals suspected of belonging to proscribed terrorist organizations. Determining the availability of alternative mechanisms in this context is not straightforward. All individuals passing through a border post undergo an identity check, which may also involve initial questioning. At this point, border officials may check an individual's identity against a database and raise an alert in the event of a match. Equally, border officials may be briefed to monitor for particular behavioral or travel patterns, which may also be used to flag an individual for more detailed questioning. It is possible, however, that a member of a proscribed organization may travel on falsified papers and be trained not to raise suspicions on initial questioning. In these circumstances, LFR technology may be particularly useful, as it has the potential to counteract these two techniques. "Necessity" (and proportionality) will accordingly turn on the specific added value of LFR compared to traditional mechanisms. Relevant considerations may include whether sufficiently high-quality pictures of suspected individuals are available, or whether such persons are typically tracked on the basis of visual identification, known aliases, or patterns of movement.

In the second example, LFR is used to identify individuals subject to outstanding arrest warrants. In this case, frequently used alternative mechanisms also exist. These include, for example, identity checks when individuals come into contact with law enforcement, visits to places typically frequented by the individual, or interviews with

[20] *See* Cases C-203/15, C-698/15, Tele2 Sverige AB v Post-och telestyrelsen and Secretary of State for the Home Department v. Watson and Others, ECLI:EU:C:2016:970, para. 102 (Dec. 21, 2016).

[21] Zakharov v. Russia, App. No. 47143/06, para. 260 (Eur. Ct. H.R., Dec. 4, 2015).

associates and family members. In considering the effectiveness of these alternative mechanisms, a number of factors are likely to be relevant, such as the nature of the underlying offence, existing success rates and time frames regarding apprehension of individuals subject to an arrest warrant, and rates of re-offending during that time period and the gravity of that offense. In determining the added value of LFR in this context, states must also consider the likelihood that a wanted individual will pass through a facial recognition camera system.

An evaluation of alternative mechanisms demonstrates whether—in any given deployment—AI technology represents a continuation of preexisting police capability by other means, or whether it represents a step-change in capability. This is relevant to the determination of potential human rights-related harm. For instance, using LFR to confirm an individual's identity at a border crossing arguably represents a continuation of an existing capability, where a single border agent checks an individual against her documentation. On the other hand, deploying LFR across city-wide CCTV networks and integrating data analysis tools may facilitate the tracking of individuals' movements, the identification of patterns of life and personal/professional networks, and the flagging of unusual or suspicious behavior. This arguably constitutes a step-change in capability, as this would not have been possible absent LFR, even with significantly increased resources.

A step-change in capabilities is a useful indicator that more in-depth analysis and impact assessments are required. It is also useful when considering whether a state may cite resource efficiencies to justify an AI deployment. There is a strong argument that where AI represents a continuation of existing capabilities, resource efficiencies should be taken into consideration, given the positive impact this may have on states' ability to fulfil rights in other areas. If, however, AI represents a step-change in capabilities, then resource savings should arguably not play a role in justifying an AI deployment: the powers of the state (and the human rights impacts) are altered significantly and so a like-for-like cost comparison is not possible.

Conclusion

This essay has attempted to identify some of the steps that states should undertake when deciding to deploy an AI tool (or not), in order to facilitate human rights compliance. The focus has been on demonstrating utility. An assessment of potential harm is equally important but constitutes the next step in the analysis. Importantly, the measures outlined above will help to set the parameters of deployment, thereby establishing the framework within which potential harm can be evaluated.[22] Once potential utility and potential harm are identified, efforts may be made to resolve any "competing interests," and it is here that appropriate safeguards, or restrictions on circumstances of use, may be identified. Hopefully, this essay also demonstrates how taking a human rights-based approach to decision-making will advance states' interests.

[22] For instance, the human rights impact of a short-term fixed location LFR deployment is different from a permanent, 24/7, city-wide deployment.