Securing meaningful transparency of public sector use of AI

Comparative approaches across five jurisdictions

Public Law Project

Contents

Executive Summary

Public authorities are increasingly using artificial intelligence (AI) and automated decision making (ADM) in a wide range of areas, including health, education, immigration, welfare benefits, policing, and prisons. While this brings the promise of greater efficiency and accuracy, it also brings with it well documented risks, including the potential exacerbation of existing inequalities.

Within this context, effective regulation is essential to ensure that these systems work for the public good. This paper focuses on one particular element of a potential future regulatory framework: **the requirements for meaningful transparency regarding public sector use of AI and ADM systems.** The decision to focus on this aspect of regulation reflects the role of transparency not just as a goal in itself, but as a necessary requirement for other important aspects of AI governance including accountability and redress, protection against unlawfulness and unfairness and ensuring privacy and data rights and the safe, effective and trusted use of these technologies.

The intention behind this paper is to support the development of effective regulation in the UK by drawing on examples from other jurisdictions. It carries out a comparative analysis of the transparency requirements as they relate to AI and ADM use by the public sector across five jurisdictions **(Canada, the European Union (EU), France, Japan and the United States of America (USA))** and considers their reported effectiveness.

It draws on this analysis, and PLP's wider experience and research in this area, to inform recommendations for the development of equivalent regulation within the UK.

Key findings

- The work carried out by **Public Law Project (PLP)** emphasises the importance of securing transparency of public sector use of AI and ADM, particularly in high-impact areas such as healthcare, education, and immigration.

- PLP's TAG Register identifies over fifty instances of AI use in the public sector, exposing risks of discrimination, unlawfulness, and unfairness, occurring without proper public scrutiny due to opaque use.

- PLP's roundtables with experts and stakeholders highlight the need for specific legislative duties to ensure AI transparency, rather than relying on the existing non-specific patchwork of regulation and legislation, and general duties.

- The current UK framework lacks robust, legally enforceable transparency requirements for AI use in the public sector. There is an over-reliance on existing regulators to enforce transparency, which may dilute the effectiveness of oversight and accountability.

- The framework does not mandate proactive public disclosure of AI usage details, leaving individuals uninformed about decisions affecting them.

- This research into transparency requirements as they relate to AI and ADM use by the public sector across five jurisdictions demonstrates that some jurisdictions are mandating comprehensive reporting and proactive disclosure of AI and ADM usage, enhancing the availability for public oversight and enabling better public trust.

- Others exhibit significant gaps in enforcement, accessibility of information, and vagueness of requirements that do not provide the suitable level of specificity needed to ensure public authorities know how best to adhere to them.

- Positive practices from other jurisdictions include notifying those who are subject to AI assisted or ADM that such technology has been used, the provision of explanations as to how the technology assisted the decision-making process and its effect on the outcome, mandatory detailed public disclosures of AI and ADM systems and tools in centralised public repositories, and signposting affected or interested individuals to these public sites.

- These findings underscore the need for a robust, legally enforceable regulatory framework in the UK to ensure transparency in AI and ADM use, drawing on successful practices from other jurisdictions and addressing current gaps in regulation.

Recommendations

Individual level transparency

---

1) Public authorities should notify individuals of the presence of an AI, algorithmic or automated tool or system when communicating the decision to them

- Public authorities should be required to explicitly inform decision subjects or those affected by a decision or action taken by a public authority about the use of an AI, algorithmic or automated tool or system when communicating the decision to them.

- This should apply in instances where the tool or system has been used to partially make or support a decision-making process as well as to solely make a decision.

2) Explanations should be provided proactively to individuals

- Alongside the notification of the presence of an AI, algorithmic or automated tool or system, public authorities should be required to proactively provide explanations to affected individuals, providing information on how and why the decision was reached. The proactive provision of an explanation avoids placing the burden of requesting an explanation, or specific information, on the individual.

- The requirement for public authorities should include specific categories of information to be included in the explanation, such as tailored information on the contribution of the AI, algorithmic or automated tool or system in the decision-making process, including the tasks performed by the system.

- The explanation should make clear that individuals may request further information specific to the decision they received, such as how the data is obtained and used, provider details, purpose, accuracy, measures taken to ensure the appropriateness of results, and the presence of human oversight and intervention mechanisms.

Systemic level transparency

3) Statutory requirement for submission of reports to the Algorithmic Transparency Recording Standard (ATRS) Hub

- The recent commitment made in the Government's response to the AI Regulation White Paper consultation to make the ATRS a requirement for all government departments is a step in the right direction. To ensure full engagement with the requirements under the ATRS, compliance should be mandated through a legally enforceable requirement.

4) In contexts where people will interact with or be impacted by an AI, algorithmic or automated tool or system, the ATRS should be mentioned and a link to the ATRS Hub should be provided.

- This will promote individual awareness of the existence of the ATRS and direct them toward the information.
- This would allow individuals and the general public to understand how a decision will be made, and inform them of the presence of AI or ADM within the process, before a decision is made rather than only after they have been affected by it.

Glossary of key terms

Term	Explanation
Artificial Intelligence (AI)	An umbrella term for a range of algorithm-based technologies that solve complex tasks by carrying out functions that previously required human thinking [1].
Automated decision-making (ADM)	Public Law Project is specifically interested in the way public bodies use automated systems to make decisions and therefore many of the uses of AI or similar technologies referred to in this paper centre around its use in decision-making. **Automated decision-making, or an automated decision** for the purposes of this paper, is one in which an automated system performs at least part of the decision-making process. Automated decision-making can enter into public body decision-making in a range of different ways: 1. **Partial automated or decision-support tool:** Where an automated system provides additional information to aid a human decision-maker in their decision (e.g. a system assesses whether an offender poses a risk of reoffending, and presents that risk score to a parole officer to inform their decision) 2. **Fully (or 'solely') automated:** Where an automated system takes a decision and action in relation to a person or group without human input (e.g. a system automatically assesses and approves an application for a driver's licence)

[1] Information Commissioner's Office definition of artificial intelligence (AI) https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/definitions/.

Term	Explanation
Decision (taken by a public authority)	When referring to a **decision** we are referring to two types of decision regarding a person: • Substantial final decisions taken by public authorities (e.g. the decision that a person's marriage is a sham or their notification to marry is decided to be an attempt to enter a sham marriage) • Preliminary decisions taken by public authorities (e.g. the decision to extend someone's notification period and investigate whether they are attempting to enter into a sham marriage)
Public authority, body, or agency	Due to the cross-jurisdictional scope of this paper and the analysis of guidance and legislation as drafted in each jurisdiction, the terms used to refer to public sector bodies varies throughout. When used within this paper, public authority, body or agency refers to a body substantially publicly funded which performs statutory duties, objectives and other activities consistent with central or local governmental functions.
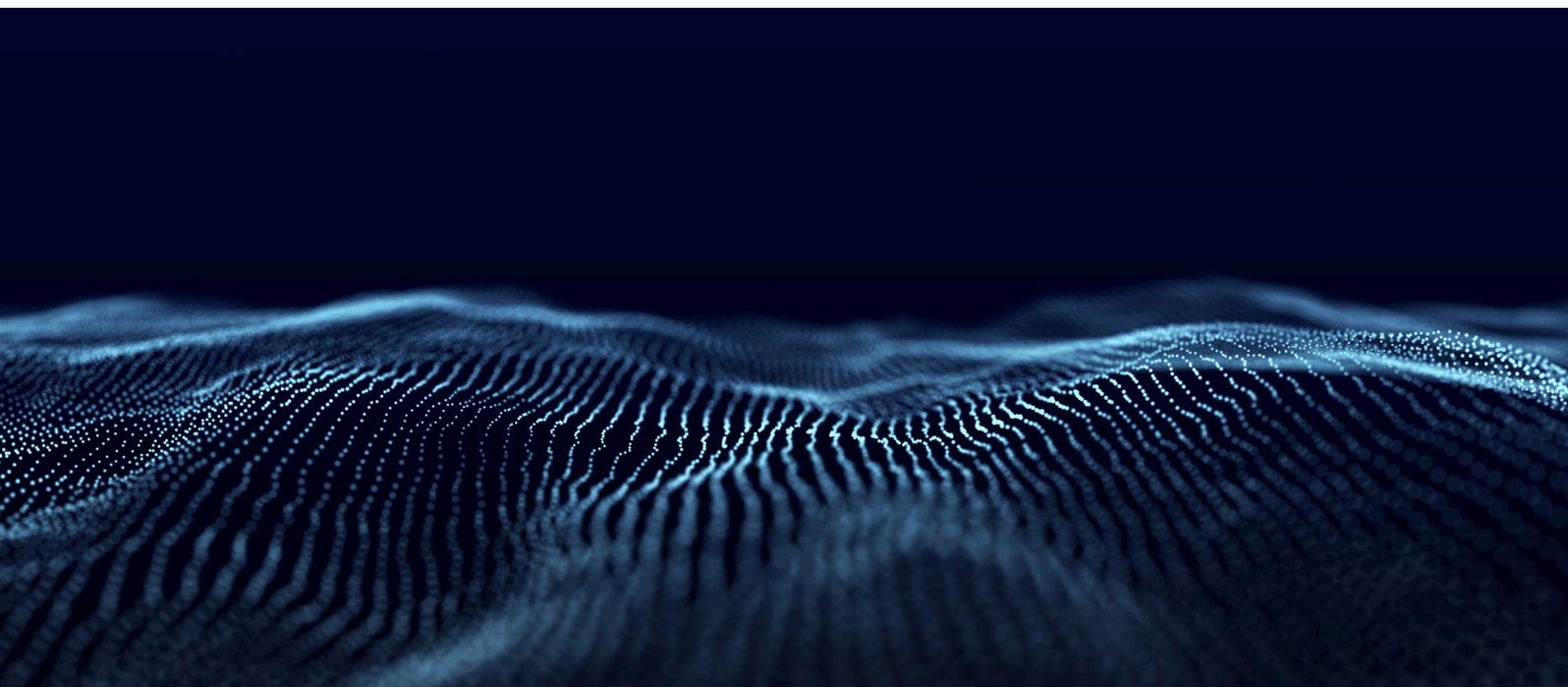
Public Law Project's work on AI regulation

One of PLP's five strategic objectives for 2022-2025 is 'ensuring that Government use of new technologies is transparent and fair'. Under this objective, the organisation has been carrying out a stream of work which looks at ADM within public authorities. The technology used in decision-making processes can vary in complexity from simple supervised rules-based algorithms to machine learning and AI.

PLP has an organisational interest in AI regulation because effective regulation of the technology, its development, and its deployment is likely to have a beneficial effect on its usage in decision-making processes in the public sector. Our research into automation and digitalisation has shown that public authorities are increasingly using AI and ADM in a wide range of high impact areas such as, healthcare, education, immigration, welfare benefits, policing and prisons.

To date, we have gathered more than fifty examples of public ADM systems through our investigative research, full details of which are publicly available in our 'Tracking Automated Government (TAG) Register' [2]. The TAG register is reflective of only a small amount of ADM systems used by public authorities, but of the tools we do know about, we have identified key risks and challenges in terms of the risk of discrimination, unlawfulness and unfairness [3].

The TAG Register is not a comprehensive picture of public sector use of AI because opacity is an inherent challenge in understanding and assessing the operation of AI. At present, most uses of AI by public authorities have been uncovered through resource-intensive research, including the submission of requests under the Freedom of Information Act 2000 (FOIA). Proactive disclosure from organisations of their use of AI would improve the extent to which individuals, organisations and legal practitioners are aware of the role of AI in administrative decisions.

---

[2] Public Law Project, Tracking Automated Government 'TAG' Register (9 February 2023) http://trackautomatedgovernment.org.uk/.

[3] See 'Unequal impact(s)' column of the 'Public Law Project, Tracking Automated Government 'TAG' Register (9 February 2023) http://trackautomatedgovernment.org.uk/' for more detail.

Roundtables on the governance of AI

Research commenced following the publication of the Government's National AI Strategy in September of 2021, which indicated that a "White Paper on a pro-innovation national position on governing and regulating AI" would be published in Spring 2022 [4]. Initially PLP hosted a two-part roundtable on the governance of AI, bringing together a broad range of participants from across civil society, academia, and legal practice.

From the roundtable discussions, PLP identified two sets of emerging themes: **'themes of consensus'** and **'themes for further development'**. Under themes of consensus, transparency, or lack thereof, was identified as a key concern, particularly in the development of a regulatory framework since transparency is seen as essential to effective evaluation and regulation of AI technologies and a building block to accountability.

Under themes for further development, roundtable attendees were less certain what types of 'entity' should be subject to transparency requirements or whether the nature of decision-making was a better way of defining the scope of requirements.

[4] Department for Science, Innovation and Technology, Office for Artificial Intelligence, Department for Digital, Culture, Media & Sport, and Department for Business, Energy & Industrial Strategy, National AI Strategy (22 September 2021) https://www.gov.uk/government/publications/national-ai-strategy.

Joint Statement on 'key principles for an alternative AI white paper'

Following on from the roundtables, the identification of themes of consensus between attendees, and the publication of the Government's AI Regulation White Paper, PLP convened a joint statement titled **'Key principles for an alternative AI White Paper'** [5]. The joint statement was designed to highlight the shortfalls of the Government's approach to AI regulation in terms of properly protecting individuals from the risk of unfairness and discrimination when automation is used to make decisions that affect them and represent the minimum required for people affected by AI to feel the benefits of this technology whilst being protected from the risks [6].

The five principles included in the joint statement were:

1. Transparency must be mandatory
2. There must be clear mechanisms for accountability at every stage
3. The public should be consulted about new ADM tools before they are deployed by government
4. There must be a specialist regulator to enforce the regulatory regime and ensure people can seek redress when things go wrong
5. Uses of AI that threaten fundamental rights should be prohibited

---

[5] Key principles for an alternative AI White Paper (June 2023) https://publiclawproject.org.uk/content/uploads/2023/06/AI-alternative-white-paper-in-template.pdf

[6] Public Law Project, Government behind the curve on AI risks (June 2023) https://publiclawproject.org.uk/latest/government-behind-the-curve-on-ai-risks/

The joint statement recommended that for these principles to be properly realised, they require obligations in statute that build on and work with existing data protection safeguards and our human rights framework.

For the purposes of this paper, the following section will expand on the principle that relates to transparency. The joint statement called for transparency requirements on public authorities to be in primary legislation, rather than in guidance.

It specified that wherever an ADM tool is being used to make, or support, decisions which have a legal, or similarly significant, effect on someone, requirements should include:

- A statutory duty on the public body to inform the person subject to the decision that ADM has been used, and how it is being used.
- Mandatory publication of the tool on a register of public use of ADM systems.

A statutory duty to publish a risk assessment (including the data protection, equality, human and child rights impacts) of the tool and measures of impact post-deployment.

PLP and the Government's AI Regulation White Paper consultation and response

The publication of the Government's AI Regulation White Paper launched a consultation period from 29 March 2023 to 21 June 2023, seeking views on the proposals set out within it [7]. Following the analysis of consultation submissions, the Government published its response and the consultation outcome in February 2024 [8].

Pro-innovation and light touch

- The Government established a pro-innovation approach to the development of an AI regulatory framework. Overall, the framework is intended to be light-touch to foster innovation and *'make the UK one of the top places in the world to build foundational AI companies'* [9].

Cross-sector focus

- The framework proposed within the White Paper is not sector specific, meaning that the same principles and approach to regulation will be applied to both private and public sector use of AI.
- Much of the White Paper focus is on how regulation will be implemented and experienced by actors in the private sector.

---

[7] Department for Science, Innovation and Technology (DSIT) and Office for Artificial Intelligence, A pro-innovation approach to AI regulation, Annex C: How to respond to this consultation (29 March 2023), https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#annexc.

[8] Department for Science, Innovation and Technology (DSIT), Consultation outcome, A pro-innovation approach to AI regulation: government response (6 February 2024) https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response.

[9] Department for Science, Innovation and Technology (DSIT) and Office for Artificial Intelligence, A pro-innovation approach to AI regulation (29 March 2023) https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper, page 4.

The 'principles-based approach'

- The White Paper sets out a 'principles-based approach' to guide the regulation of AI. These five principles are:

1. Safety, security and robustness
2. Appropriate transparency and explainability
3. Fairness
4. Accountability and governance
5. Contestability and redress

- The principle of **'appropriate transparency and explainability'** is most relevant to the scope of this paper.
- The White Paper states that transparency *"refers to the communication of appropriate information about an AI system to relevant people"*. Explainability is said to refer to the *"extent to which it is possible for relevant parties to access, interpret and understand the decision-making processes of an AI system"*. [10]

- The White Paper denotes two intended audiences for 'appropriate' transparency and explainability.

  - First, that an appropriate level of transparency and explainability for regulators will mean they have sufficient information about AI systems and their associated inputs and outputs to give meaningful effect to the other principle.

  - Second, that parties directly affected by the use of an AI system should also be able to access sufficient information about AI systems to be able to enforce their rights.

- The principle joins numerous other examples of non-statutory guidance already published by regulators, governmental and non-governmental bodies, both domestically and internationally [11]. This new principle adds little to this ecosystem. Instead, it acts as an instruction to regulators about the outcomes they ought to be working towards when AI is used in the areas for which they are responsible.

---

[10] Department for Science, Innovation and Technology (DSIT) and Office for Artificial Intelligence, A pro-innovation approach to AI regulation (29 March 2023) https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper, section 3.2.3 page 28.

[11] 'Guidance on AI and data protection', Information Commissioner's Office https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/; 'Guidelines for AI procurement. Office for AI (June 2020) https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement; 'Artificial intelligence in public services', Equality and Human Rights Commission (September 2022) https://equalityhumanrights.com/en/advice-and-guidance/artificial-intelligence-public-services; Organisation for Economic Co-operation and Development, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449 (22/05/2019) https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449; Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI) (December 2021) https://rm.coe.int/cahai-2021-09rev-elements/1680a6d90d.

Existing regulators

- The framework as set out in the White Paper requires existing regulators to interpret and implement the principles. The White Paper acknowledges that some regulators have warned they may *"lack the statutory basis to consider the application of the principles"* [12].
- There is no specific regulator for public sector use of AI, but the public sector is subject to cross-cutting regulation that relates to the use of data, the UK GDPR and decision-making, and the Equality Act 2010. These legal frameworks are enforced by the Information Commissioner's Office (ICO) and the Equality and Human Rights Commission (EHRC) respectively.
- The ICO and EHRC are therefore the responsible bodies for securing appropriate transparency and explainability of public sector use of AI under the AI regulation framework.
- Both regulators have published their strategic approach to regulating AI following the request in the Government's response to the AI regulation consultation [13].

―――――――――――――――――――――――――――――――

[12] Department for Science, Innovation and Technology (DSIT) and Office for Artificial Intelligence, A pro-innovation approach to AI regulation (29 March 2023) https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper, page 36.

[13] Department for Science, Innovation and Technology (DSIT), A pro-innovation approach to AI regulation, Government response to consultation (6 February 2024) https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response.

The Information Commissioner's Office

- The ICO's strategic approach sets out how it is driving forward the principles set out in the AI Regulation White Paper and government's guidance [14].
- It clarifies that the regulation of AI falls under the ICO's purview when the development and deployment of AI systems includes the processing of personal data. The strategic approach document also explains that because data protection law is 'technology-neutral', the ICO's approach will apply to any processing of personal data by new technologies, not only AI [15].

The Equality and Human Rights Commission (EHRC)

- The EHRC's update on its approach to regulating AI is comparably brief to that of the ICO [16]. It sets out that AI is a priority for the regulator and has been since 2022 and that it recognises the important role it has in supporting the responsible and fair use of AI.

- However, it makes clear that because the EHRC is a small regulator, and its budget has not been meaningfully increased to take on the additional responsibility of regulating AI its ability to scale up and respond to the risks to equality and human rights presented by the technology is limited.
- It goes as far as to say that it does not have the resources to develop dedicated guidance around the White Paper principles, as directed by Government.

Central functions

- To support regulators in delivering the proposed AI regulatory framework with their existing resources and remit, the Government proposed the delivery of 'central functions' to coordinate, monitor and adapt the framework [17]. These central functions will be carried out by the Government itself.
- As the regulatory framework evolves and the Government develops a clearer understanding of the functions needed, they plan to review the operational model outlined in the White Paper. In particular, they plan to consider if a government unit is the most appropriate mechanism for delivering the central functions in the longer term, or if an independent body would be more effective [18].

---

[14] Information Commissioner's Office, Regulating AI: The ICO's strategic approach (April 2024) https://ico.org.uk/media/about-the-ico/consultation-responses/4029424/regulating-ai-the-icos-strategic-approach.pdf.

[15] Information Commissioner's Office, Regulating AI: The ICO's strategic approach (April 2024) https://ico.org.uk/media/about-the-ico/consultation-responses/4029424/regulating-ai-the-icos-strategic-approach.pdf, para 6, page 4.

[16] Equality and Human Rights Commission, An update on our approach to regulating artificial intelligence (April 2024) https://www.equalityhumanrights.com/media-centre/news/update-our-approach-regulating-artificial-intelligence.

[17] Department for Science, Innovation and Technology (DSIT) and Office for Artificial Intelligence, A pro-innovation approach to AI regulation (29 March 2023) https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper, page 43.

[18] Department for Science, Innovation and Technology (DSIT) and Office for Artificial Intelligence, A pro-innovation approach to AI regulation (29 March 2023) https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper, para 79, pages 54-55.

Response to the Government's AI Regulation White Paper consultation

PLP responded to the White Paper consultation, confining comments to questions 1-10, 17-18, and L1-L2 due to areas of organisational focus and expertise on public sector use of AI and ADM [19]. These questions covered the revised cross-sectoral AI principles, a statutory duty to have due regard to those principles, the central functions, monitoring and evaluation of the framework and legal responsibility for AI.

The White Paper did not give specific consideration to the uses, risks and need for additional regulation in relation to the use of AI within ADM systems used by public authorities. PLP is concerned that this results in an unnecessarily 'light-touch' approach to the regulation of public authority use of AI and ADM. In the Government's response to the consultation, some specific attention was given to AI best practice in the public sector. Most notably the Government announced that the pilot of the ATRS was successful and that use of it will now become a requirement for all government departments. This requirement will be extended across the broader public sector over time [20].

In terms of transparency, PLP's consultation response drew attention to the challenges posed by opacity to understanding and assessing the operation of AI by public authorities. It highlighted that requiring public authorities to make it clear when they are using AI would be a helpful first step in improving transparency but that being 'clear' about the use of AI alone will not be enough to improve transparency. It instead recommended proactive disclosure from public authorities of their use of AI to improve the extent to which individuals, organisations and legal practitioners are aware of the role of AI in administrative decisions.

It was suggested that this could be achieved through the public being informed of the use of AI by public authorities alongside the notification of decisions and for this information to be provided in a way that is concise, transparent, intelligible, easily accessible, and in clear and plain language. In practical terms, this could include information such as the existence of the system, operating details, the (unequal) impacts on those with protected characteristics, specific groups or communities and the role of AI in the decision recommendation, or final decision.

---

[19] Public Law Project, Response to the AI White Paper consultation (June 2023) https://publiclawproject.org.uk/resources/hurt-first-fix-later-ai-regulation-white-paper-consultation-response/.

[20] Department for Science, Innovation and Technology (DSIT), A pro-innovation approach to AI regulation, Government response to consultation (6 February 2024) https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response, para 41, page22.

In response to the consultation question on what other measures the Government could require of organisations to improve AI transparency, PLP suggested **three** measures that were not considered by the White Paper:

1) Introduce specific compulsory transparency requirements for public sector use of AI and ADM systems

3) Introduce a requirement for public authorities to notify individuals when automation is used to reach a decision, similar to that under the Canadian Directive on Automated Decision Making (DADM),[21] and France's Loi pour une Republique Numérique (Law for a Digital Republic) 2016 [22].

2) Make the submission of reports to the ATRS mandatory for public authorities and increase the level of operational details required

The full rationale behind these suggestions is set out in paragraphs 21-30 of the consultation response [23].

---

[21] Government of Canada, Direction on Automated Decision-Making (1 April 2019), available at: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592.

[22] Law No. 2016-1321 of October 7, 2016, for a Digital Republic.

[23] Public Law Project, Response to the AI White Paper consultation (June 2023) https://publiclawproject.org.uk/resources/hurt-first-fix-later-ai-regulation-white-paper-consultation-response/.

Roundtable on practical policy and legislative proposals around the use of AI by public authorities

In autumn 2023, PLP convened a private roundtable between legal, policy and regulation researchers, academics and legal practitioners to discuss the concerns raised in consultation responses regarding the lack of proper consideration of regulation of the use of AI in the public sector.

The roundtable discussion focused on and explored the further development of practical policy and legislative proposals around five key themes of public sector use of AI:

- Transparency,
- Public consultation,
- Rights related to human involvement, accountability
- Redress
- Regulation [24].

Focusing only on the transparency theme for the purposes of this paper, the roundtable discussion demonstrated the range of objectives and the factors to consider when making proposals as to how best to achieve transparency from public authorities around the use of AI and ADM. Participants expressed hesitancy around the introduction of a general statutory duty for ensuring transparency, due to the complexity of implementation for Government, the need for exemptions and the risk a general duty will water down its impact.

Participants were however much more in favour of specific duties being put on a legislative basis, such as engagement with the ATRS. There was general agreement that not only did the ATRS need to be better engaged with by public authorities, but also that most of the reports submitted to the ATRS Hub at the time of the roundtable were of low potential impact, whilst reports for tools and systems that carry the potential for higher impact were not present.

---

[24] Public Law Project, Developing proposals for regulation of public sector use of AI: roundtable findings (November 2023) https://publiclawproject.org.uk/content/uploads/2023/11/Roundtable-findings-report-1.pdf.

It was raised within the course of the roundtable discussion that when considering transparency requirements, it might be useful to reflect on what transparency is trying to achieve. This might be considered through the lens of two separate or possibly interconnecting audience groups that may seek different levels of transparency because of the difference in what they are trying to achieve with it. The first audience group is the general public and/or decision-subjects and the second, a more expert audience [25].

With the two different audience groups in mind, roundtable participants suggested a range of transparency requirements spanning from obligations to provide notice that a decision will be undertaken in whole or in part by an automated decision system, the submission of detailed reports under the ATRS, and the need for explanations to be available or provided to enable individuals to understand the totality of how decisions about them are made [26]. Some participants expressed views that decision subjects as well as data subjects should get information through transparency mechanisms.

---

[25] Public Law Project, Developing proposals for regulation of public sector use of AI: roundtable findings (November 2023), page 6 https://publiclawproject.org.uk/content/uploads/2023/11/Roundtable-findings-report-1.pdf.

[26] For further details see Public Law Project, Developing proposals for regulation of public sector use of AI: roundtable findings (November 2023), page 6 https://publiclawproject.org.uk/content/uploads/2023/11/Roundtable-findings-report-1.pdf.
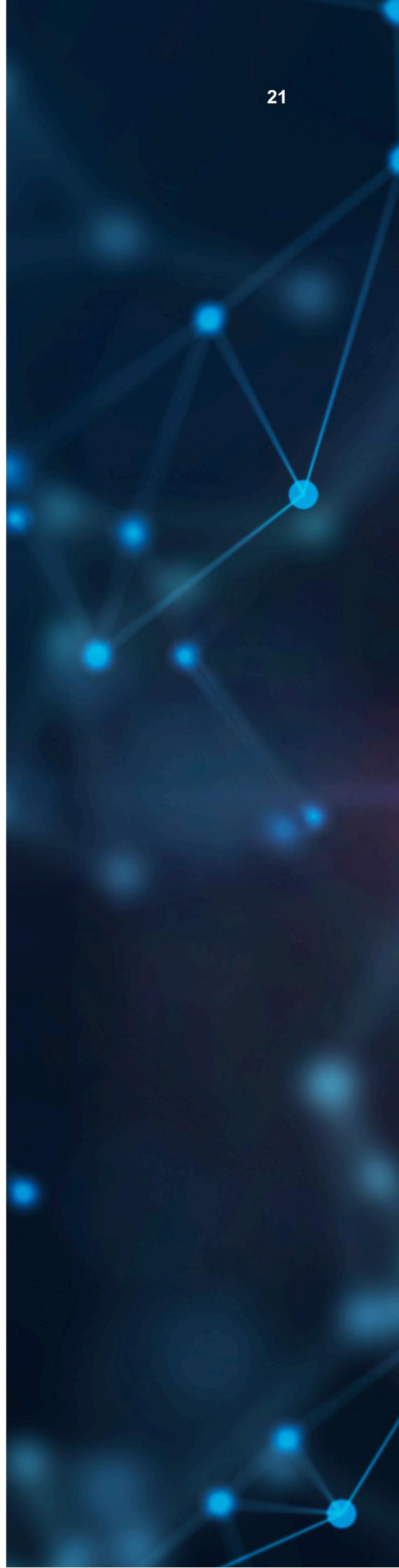
Key takeaways from PLP's work on AI regulation

Taken together, the findings from the roundtables and PLP work detailed above demonstrates:

- widespread demands for greater transparency around the use of AI and automated decision-making by public authorities,

- commonality within the dissatisfaction regarding current levels of transparency and requirements in the UK, and

- the variety of methods available for seeking to achieve better transparency and the differing views of the sector in terms of which are most preferable.

Existing regulators across the UK are beginning to implement the AI regulatory framework and enter the **'test and learn'** stage. This stage is envisioned by the Government to allow for adaptions to the regulatory approach to be made based on what works. It is therefore important as we enter this stage to have a clear vision of what type of transparency we require from public authorities regarding their use of AI and ADM, for which audiences and the most effective ways to secure this.

The following section explores this through cross-jurisdictional comparative research, considering transparency requirements as they relate to public sector use of AI across five jurisdictions. This research is intended to offer insight into which of the 'variety of methods' are in place in other jurisdictions, how effective they are at securing meaningful transparency, and what features enable this effectiveness. This is followed by a comparative analysis of the transparency requirements across the jurisdictions within scope of this paper, which informs a set of considered proposals as to what requirements the UK Government should put in place to secure transparency of public sector use of AI and ADM.

Transparency requirements across jurisdictions

PLP is grateful to the Fieldfisher LLP pro bono team for their work in conducting the research that has informed the following section.

This section explores the transparency requirements that apply to public authorities in relation to the use of AI for automated decision making in five jurisdictions:

- Canada
- The European Union (EU)
- France
- Japan
- The United States of America (USA)

World Map with Jurisdictions

In relation to each jurisdiction selected, the questions that are examined are:

Transparency requirement

An explanation of the transparency requirement, whether it is a statutory or non-statutory duty requirement, how it is enforced, and length of time in force.

Technology applied to

Levels of technological complexity required for the tool to fall into scope, such as whether the requirement applies only to AI, or other less complex forms of automated decision-support/making.

Scope

The type of automated system, type of decision made by a public authority, or the type of body making the decision that the requirement applies to.

Information to be shared

Categories of information required to be disclosed about automated decision making by public bodies.

Application

How the transparency requirement applies to public authorities and decision makers.

Exemptions

Any circumstances in which public authority or decisions are exempt from the transparency requirement (if any).

Canada

Overview

This section focuses on measures at the federal level in Canada, and does not consider proposed or existing provisions for the regulation of AI and the use of algorithms and ADM that have been introduced at a Province or Territory level.

The Federal Government of Canada introduced Bill C-27 (the Digital Charter Implementation Act, 2022) in Spring 2022. The first two parts update existing legislation, and the third introduces the Artificial Intelligence and Data Act (AIDA). The AIDA, if adopted, will be the first piece of legislation to focus specifically on regulating AI in Canada. However, it will not apply to government uses of AI, reflecting the approach taken to date in Canada of distinguishing between digital and privacy regulation of public and private sectors [27].

There are however 'soft law' measures in place that apply to most federal Government agencies and departments (Government of Canada Organisations). The Policy on Service and Digital (the 'Policy') was adopted in 2019 and took effect on 1 April 2020. It includes a requirement to ensure transparency and disclosure regarding the use of Automated Decision systems and their ongoing assessment and management of risks.

The Directive on Automated Decision Making (DADM) supports this overarching policy and includes specific actions that those organisations are required to take that relate to transparency. Initially adopted in 2019, it was updated in April 2023 following a stakeholder review to adapt use of the DADM based on the current Canadian and global AI landscape, and the evolving risks created by these changes.

---

[27] Bill C-27, Digital Charter Implementation Act, 1st session, 44th Parliament, 2021 https://www.parl.ca/legisinfo/en/bill/44-1/c-27.

Transparency requirement(s)
Canada

Non-statutory duty:

The Policy states that deputy heads of Government of Canada organisations are responsible for ensuring 'the responsible and ethical use of automated decision systems', including:

- ensuring decisions produced using them are 'efficient, accountable and unbiased' and
- 'ensuring transparency and disclosure regarding the use of the systems and ongoing assessment and management of risks' [28].

Non-statutory duty:

The DADM includes requirements on federal institutions to:

- Complete and release an Algorithmic Impact Assessments (AIA) prior to the production of any automated decision system, [29]
- Publish the source code of automated decision systems (subject to certain exemptions) [30]
- Document the decisions made by these systems, for the purposes of monitoring and reporting [31]

- Depending on the level of risk associated with the decision:
1. Provide notice that a decision will be undertaken in whole or in part by an automated decision system [32]
2. Provide meaningful explanations of how and why the decision was taken

Proposed statutory duty (private sector only):

The AIDA, if adopted, seeks to regulate the design, development and use of 'high-impact' AI systems in the private sector. In relation to transparency, it includes a requirement to publish a plain language description of the AI system that explains its key elements, such as how it is intended to be used, the type of content it is intended to generate and the mitigation measures that have been put in place to mitigate the risks of harm and biased output. Contraventions of the transparency requirements can result in fines up to the greater of CAD 10 million and 3% of global revenues in, or up to CAD 50,000 in the case of an individual. AIDA would not apply to government institutions [33].

---

[28] Policy on Service and Digital Section 4.4.2.4

[29] DADM section 6.1

[30] DADM section 6.2.6

[31] DADM section 6.2.8

[32] DADM section 6.2.1 and 6.2.2

[33] AIDA, section 3

What the requirements mean

While there are currently no statutory requirements that apply at a federal level to public bodies in Canada, the Policy and DADM are mandatory policy-setting instruments that require federal agencies and departments to take certain actions and achieve certain objectives [34]. However, while there is an accountability framework aimed at ensuring compliance, [35] those measures are internal to government; they do not create actionable rights for individuals or organisations.[36]

Both instruments apply to most federal agencies and departments.They do not apply to systems used by the provincial government, municipalities or provincial agencies such as the police service and child welfare agencies [37].

The Policy requirements that relate to ADM systems includes technology that 'either assists or replaced the judgment of human decision-makers' [38]. It includes a responsibility for deputy heads of the relevant organisation to ensure the transparency and disclosure regarding the use of those systems.

The supporting DADM, includes a requirement for Government agencies to complete and release the final results of an AIA prior to the production of any automated decision system (ADS) [39]. Designed to encourage authorities to assess the impact of their AI and ADM systems, the assessment template provides a set of questions that cover details of the project, reasons for automation, details of the system and algorithm and the decision it aids in reaching. It also helps identify the impact level of the ADM system. The final results of the AIA should be published in an accessible format via the Government of Canada websites [40].

---

[34] The Treasury Board of Canada oversees the spending and operation of the Government of Canada. It has the power to issue a range of policy instruments that are designed to establish mandatory requirements (rules) or voluntary best practice. The mandatory instruments available to them include Policies and Directives.Policies impose specific responsibilities on departments (what they are expected to achieve); Directives set out specific actions they must take or avoid (how they must meet a policy objective). See further, Foundation Framework for Treasury Board Policies: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=13616

[35] The Treasury Board is responsible for monitoring compliance and can intervene in cases of non-compliance. Interventions can take various forms, including informal follow-ups, requests for specific information or additional reports, external audits or other investigations, formal directions on specific preventative or corrective measures and withdrawal of authority

[36] Scassa, Teresa, Administrative Law and the Governance of Automated Decision-Making: A Critical Look at Canada's Directive on Automated Decision-Making (October 30, 2020). Forthcoming, (2021) 54:1 University of British Columbia Law Review, available at: http://dx.doi.org/10.2139/ssrn.3722192

[37]DADM- Canada.ca

[38] The Policy, Appendix A: Definitions

[39] DADM, section 6.1.

[40] DADM, section 6.1.4.

The extent of the transparency obligations that apply is linked to the impact level of the decisions being taken, with more extensive requirements for higher impact decisions. Depending on that impact level, there is a requirement to provide notice prominently and in plain language on websites when decisions will be made by or with the assistance of an ADS. In relation to very high impact decisions, there is also a requirement to publish documentation describing how the ADS components work, how it supports the administrative decision, the results of any reviews or audits, and a description of the training data.

For all but the lowest impact decisions, agencies must also provide meaningful explanations of their ADS informed decisions to affected individuals, with any decision that results in the denial of a benefit or service, or involves regulatory action.That explanation must inform them in plain language of the role of the ADS in the decision-making process, the training and client data, their source and method of collection, the criteria used to evaluate client data and the operations applied to process it, the output produced by the system and any relevant information needed to interpret it in the context of the decision, and a justification of the administrative decision, including the principal factors that led to it. A general description of these elements must also be made available through the AIA and discoverable via a departmental website.

In relation to the lowest impact decisions, there is still a requirement to ensure a meaningful explanation is published for common decision results and a requirement to publish a general description of the elements described in the paragraph above.

Reported effectiveness of requirements

The DADM has been praised as being the first of its kind and other countries have drawn inspiration from it when considering their approach to AI regulation [41].

However, it has been reported that in practice it can be challenging to know when an algorithm has been used as part of a decision-making process [42]. As of May 2024, there were only 21 AIAs publicly available through the designated open government portal.[43] In contrast, the Canadian Tracking Automated Government (TAG) Register displays details of 303 systems in use, piloted or previously used by public authorities [44]. These systems have been identified by researchers at the Starling Centre rather than information regarding their operation being proactively disclosed by the relevant public authorities [45]. 95% of the systems logged in the Canadian TAG Register were used at the federal government level and therefore fall within the scope of the DADM.

Furthermore, the AIAs that have been published on the open government portal are quite brief, often with little information provided about the impact or the potential impacts of a tool or system.

The DADM and the Policy are soft-law instruments and therefore their legal force only expands insofar as it informs common law administrative law obligations and internal government disciplinary measures. The consequences are determined by the impact level of the decision and there are different consequences for institutions and individuals which range from persuasion to restraint [46].

There has been criticism that the statutory measures contained in AIDA will not apply to government institutions (alongside wider criticisms that AIDA is not sufficiently robust or sufficiently consulted upon) [47].

[41] CANADA'S DIRECTIVE ON AUTOMATED DECISION-MAKING Policy - OECD.AI

[42] How artificial intelligence will change administrative law: The Government of Canada's Directive on automated decision-making - Lexology

[43] https://search.open.canada.ca/data/?collection=aia&page=1&sort=metadata_modified+desc

[44] Tracking Automated Government (TAG) Register Canada https://tagcanada.shinyapps.io/register/.

[45] Starling Centre, Faculty of Information and Media Studies at Western University, Canada.

[46] Framework for the Management of Compliance, see Appendix C and Appendix D

[47] CanadianLabourCongress-e.pdf (ourcommons.ca) Submission on the Proposed Artificial Intelligence and Data Act - The Dais

The European Union

Overview

The European Union (EU) proposed the EU AI Act in April 2021. On 13 March 2024, the European Parliament approved the AI Act. It will enter into force twenty days after publication in the Official Journal of the EU, with various provisions becoming applicable over 6-36 months.

The AI Act promotes transparency in the development, deployment, and use of AI systems. There is a specific requirement that when using generative AI systems, deployers should make human users aware that they are interacting with AI.

The AI Act will apply to decisions made by public bodies and local authorities which use AI systems to aid their decision-making.

The AI Act provides a governance framework for the coordination and implementation of the regulation, involving national authorities, the European Commission, and a new European AI Office and Artificial Intelligence Board. Public authorities in Member States will be responsible for enforcement, similar to the role of data protection authorities under the GDPR, with the exception of general-purpose AI (GPAI) enforcement of which will sit with the European AI Office.

The regulation would also provide for sanctions and remedies for non-compliance, such as fines, injunctions, or withdrawal of products from the market. Fines for violations may vary depending on the seriousness of the offence, with the highest fines reaching up to 7 percent of global turnover or 35 million euros, whichever amount is higher.

The Act classifies AI according to its risk and establishes varying levels of regulation based on risk categories:

- Unacceptable risk is prohibited,
- Most of the text addresses high-risk AI systems, which are regulated,
- Limited risk AI systems are regulated, but subject to lighter transparency obligations,
- AI systems of minimal risk are unregulated [48].

[48] EU Artificial Intelligence Act, High-level summary of the Act (27 February 2024) https://artificialintelligenceact.eu/high-level-summary/.

Transparency requirement(s)
European Union

Statutory duty

Article 50 of the AI Act puts in place transparency obligations for both providers and deployers of certain AI systems, and GPAI. Article 50(1) contains a transparency obligation for providers of AI systems that directly interact with natural persons. The provision states that systems that fall within this scope must be 'designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system, unless this is obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use'.

Article 50(3) requires deployers of emotion recognition or biometric categorisation systems to inform individuals who interact or are exposed to the system or the operation of the system. Systems used to detect, prevent or investigate criminal offences are mostly exempt from this requirement.

Article 50(4) requires deployers of AI systems that generate or manipulate image, audio or video content constituting a deep fake to disclose that the content has been artificially generated or manipulated.

Systems used to detect, prevent or investigate criminal offences are mostly exempt from this requirement.

Information disclosed under Article 50 is to be provided to those concerned in a 'clear and distinguishable' manner at the time of the first interaction or exposure, at the latest [49].

Article 71 of the AI Act establishes an EU database for high-risk AI systems, to be maintained by the European Commission in collaboration with Member States [50]. The information to be submitted by both providers and deployers to the EU database when registering high-risk AI systems is detailed in Annex VIII of the AI Act [51]. Most of the information contained in the EU database is to be accessible and publicly available, with the exception of high-risk systems used in law enforcement, migration, asylum and border control management [52]. The information is to be in a user friendly manner and be easily navigable and machine-readable [53].

---

[49] Article 50(5) EU AI Act https://artificialintelligenceact.eu/article/50/.

[50] Article 71(1) EU AI Act https://artificialintelligenceact.eu/article/71/.

[51] Annex VIII: Information to be Submitted upon the Registration of High-Risk AI Systems, EU AI Act https://artificialintelligenceact.eu/annex/8/.

[52] Article 71(4), the exceotion as referred to in Article 49(4), EU AI Act https://artificialintelligenceact.eu/article/71/.

[53] Article 71(4) EU AI Act https://artificialintelligenceact.eu/article/71/.

What the requirements mean

The Act applies to AI systems, which are defined as "machine-based systems that are designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations or decisions that influence physical or virtual environments" [54]. The EU AI Act would therefore apply to any form of automated decision-making that meets this definition, regardless of the complexity of the technology or methodology used.

As set out above, the transparency requirements within the EU AI Act will apply to public authorities and decision makers in different ways, depending on the risk level and the purpose of the AI systems they use or develop [55].

The AI Act considers two types of AI systems to fall under the category of high-risk:[56]

1. AI intended to be used as a product (or the security component of a product) covered by specific EU legislation, such as civil aviation, vehicle security, marine equipment, toys, lifts, pressure equipment and personal protective equipment.
2. AI systems listed in Annex III of the Act, such as remote biometric identification systems, AI used as a safety component in critical infrastructure, and AI used in education, employment, credit scoring, law enforcement, migration, and the democratic process [57].

---

[54] AI regulation: European Parliament approves legal framework for artificial intelligence (deloitte.com)

[55] EU AI Act: first regulation on artificial intelligence | News | European Parliament (europa.eu)

[56] Article 6, EU AI Act.

[57] Annex III, EU AI Act.

When using high-risk AI systems, public bodies must disclose provider details, purpose, accuracy, robustness, security, and the presence of human oversight and intervention mechanisms. They must include information about the AI system's logic, significance, consequences, and offer the opportunity for human intervention, expression of views, and the ability to contest the decision.

High-risk AI systems that operate in specific sectors must be registered in a dedicated EU database. These sectors cover many critical and sensitive areas under the control of public authorities, such as the management and functioning of essential infrastructure, the educational sector, access to important public services, law enforcement, and systems involved in the management of migration and border control.

However, for high-risk AI systems used in law enforcement, migration, asylum and border control management, the requirement to register systems on the database differs. Information regarding those systems will not be made public, and instead will be on a non-public section of the database, where visibility is strictly limited to the Commission and market surveillance authorities with regard to their national section of that database [58].

Public authorities and decision-makers using or developing low-risk AI systems have fewer obligations. They must inform users of the system, offering them the option to opt out. They also have to indicate if the system generates or manipulates content, such as images, videos, or text, and provide information about the data sources and the methods used.

---

[58] EU AI Act, Recital 69.

Reported effectiveness of requirements

The EU AI Act is a very new piece of regulation, so its effectiveness will depend on how it is implemented and enforced in practice in each Member State (and how this is overseen by the new EU AI Office). However, the Act has been subject to scrutiny throughout the negotiations and some commentators have criticised the transparency requirements for not going far enough.

The Act exempts AI tools designed for military, defence or national security use and does not apply to systems designed for use in scientific research and innovation. The exemptions may raise some concerns about the potential misuse or abuse of AI systems that are not subject to transparency requirements, and the impact on the accountability and oversight of public decisions that rely on such systems.

There have been calls from civil society organisations for the requirement to register high-risk AI systems in the public-facing EU database to be extended to those used in law enforcement, migration, and border control [59]. It has been argued that the use of AI by police, security and migration authorities requires "more not less transparency due to the vast consequences for human and procedural rights." [60].

Civil society organisations have expressed concerns about the threshold for the requirement for public authorities to register the use of AI systems within the EU database and argue that any uses of AI systems used by or on behalf of public authorities, regardless of their level of risk, should be registered within the database [61]. They contend that because of the unique role and responsibility of public authorities, the sensitive personal data they have access to, and the consequential effects their decisions have on individuals they should be subject to more stringent transparency requirements when using AI systems.

---

[59] EDRi and Access Now, Civil society statement, 'EU lawmakers must regulate the harmful use of tech by law enforcement in the AI Act' (20 September 2023)

[60] EDRi, EU AI Act Trilogues: Status of Fundamental Rights Recommendations (16 November 2023) https://edri.org/our-work/eu-ai-act-trilogues-status-of-fundamental-rights-recommendations/.

[61] AlgorithmWatch, EU Artificial Intelligence Act – recommendations on public transparency, Ensure consistent and meaningful public transparency (April 2022) https://algorithmwatch.org/en/wp-content/uploads/2022/04/Database-issue-paperApril2022.pdf.

France

Overview

At present, France has no specific legislation which governs AI. However, a series of legislation addresses the use of ADM by public bodies.

The three key principles which regulate AI and ADM in France are fairness, [62] transparency and a right to an explanation. Public administrations are required to provide reasons for their decisions ('principe de motivation des décisions'), a principle that has existed in French administrative law since 1979.

When public administrations collect and use personal data, they must comply with the Data Protection Law (adopted in 1978 and revised since GDPR adoption).

The use of ADM systems by public bodies is further regulated by the Digital Republic Act (2016), which mandates transparency and a right to an explanation in an effort to implement the transparency principle.

A decision by the Constitutional Council ('Conseil constitutionnel') in charge of the review of the constitutionality of legislation, has stated that algorithms capable of revising the rules they apply themselves, without human involvement and validation (such as machine-learning based algorithms), may not be used as the sole basis for an individual administrative decision as this would not allow public bodies to explain to those affected by the decision, the factors that had been taken into account when reaching the decision [63].

---

[62] The principle of fairness ("loyauté") was formulated by the Council of State in its 2014 annual study on "Le numérique et les droits fondamentaux" ("Digital technology and fundamental rights")

[63] Décision n° 2018-765 DC du 12 juin 2018
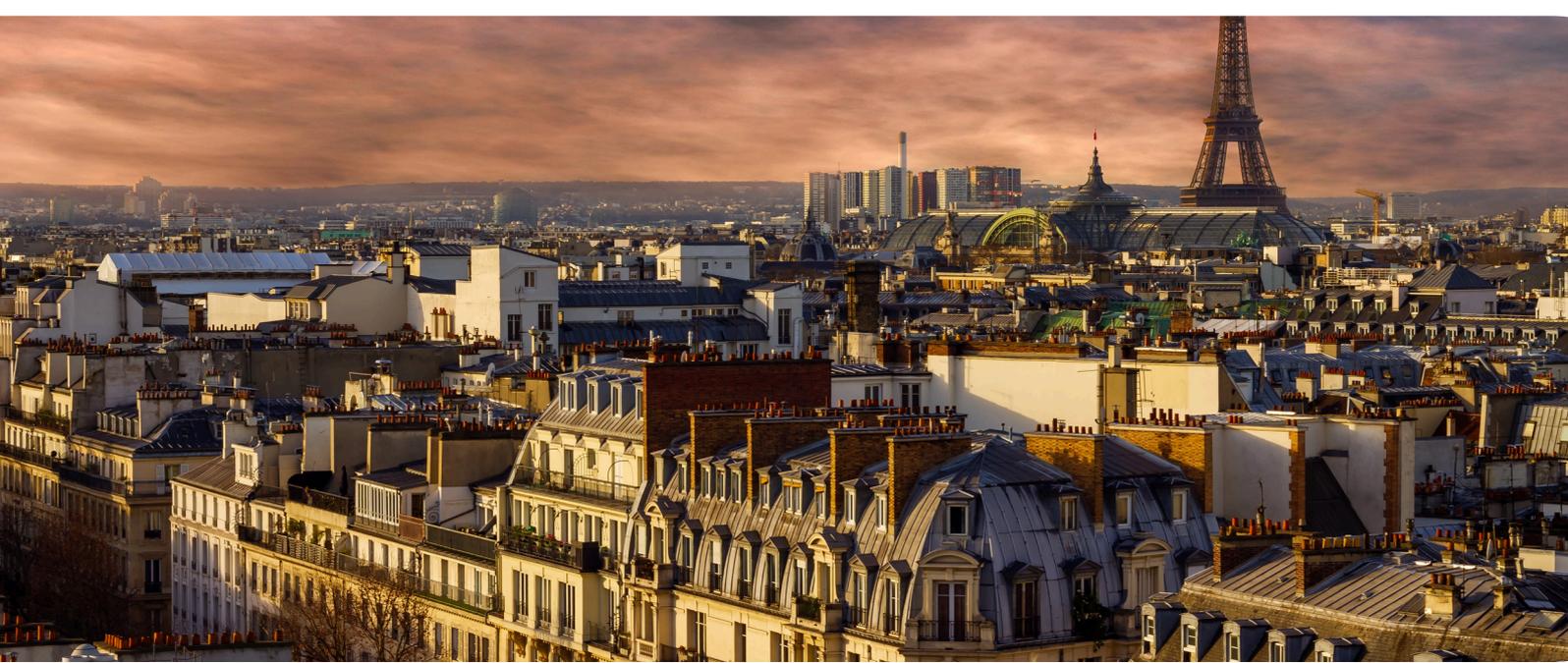
Transparency requirement(s)
France

Statutory duty

Loi pour une République numérique / Digital Republic Act (2016) introduced the principle of transparency for public algorithms when they are used to take individual administrative decisions. In practice, this means offering individuals (natural and legal persons alike) new rights.

The text amends the Code des Relations entre le public et l'administration / Code of relations between the public and the Administration ('CRPA')) and lays down three obligations for public bodies:

- *Explicit mention:* to inform a decision subject that an algorithm has been used in a decision making process and what their rights are,
- *General information:* to publish the operating principles of the main processing operations when they are the basis of individual administrative decisions,
- *Individual information:* to provide the individual concerned with a detailed set of information about the algorithm, its functioning and the data processed for the individual's specific case on request.

Compared to the GDPR, the obligations introduced by the Digital Republic Act are broader, as they cover both solely automated decisions and cases where algorithms are only decision support tools [64].

The legislation also codifies the 'principe de motivations des décisions' that public bodies should provide reasons for their decisions [65].

Individuals may refer a matter to the Commission for Access to Administrative Documents ('CADA') where they have been refused access to a public body decision (Article L342 CRPA) and CADA may impose sanctions on a public body (Article L326-1 CRPA).

These sanctions include:

- An order prohibiting the public body from reusing public information for a certain period; or
- A fine.

Separately, legal action may be brought before an administrative court to challenge the legality of a public body decision under the CRPA (Article L411). Contentious appeals against administrative decisions under the CRPA are brought before the ordinary administrative courts (Article L431 and the Code of Administrative Justice).

These duties apply to public administrations that practice automated individual decision making, and to AI as well as less complex forms of automated decision-making and to content, goods and services created by means of computer algorithms.

---

[64] "With great power comes great responsibility": keeping public sector algorithms accountable, Etalab working paper, Chignard S. and Penicaud S. 2019

[65] Article L.211-2, CRPA

What the requirements mean

Under the CRPA, public bodies must publish online the rules defining algorithmic processing used in the performance of their tasks when such algorithms are the basis of individual decisions [66]. In the case of administrative decision-making, the decision subject must be explicitly informed in writing [67], and the public body must communicate, upon request, the degree and method of contribution of the algorithmic processing to the decision made; the processed data and their sources; the processing parameters the variables of the algorithm and their weighting in the case of the person concerned; and, finally, the tasks performed by the algorithm [68].

Reported effectiveness of requirements

The effectiveness of the requirements is still debated in France. Still, the consensus is that public administrations fall short of their transparency obligations, particularly with the new obligations introduced by the Digital Republic Bill.

Among the reasons identified are **lack of sanctions for non-compliance, a lack of awareness of these obligations and, more broadly, a lack of a shared definition of what ADM systems are**.

Public administrations are prone to say that they don't use 'fully' automated decision-making systems and that all systems are under human supervision. Recent research questions the extent of human supervision and its effectiveness in ensuring a fair and explainable decision [69].

[66] Article L. 312-1-1 and L. 312-1-3 of the CRPA.

[67] Article L. 311- 3-1 of the CRPA

[68] Article R. 311-3-1-2 of the CRPA

[69] Winston Maxwell. Lz contrôle humain des systèmes algorithmiques - un regard critique sur l'exigence d'un "humain dans la boucle". Université Paris 1 Panthéon- Sorbonne, 2022.
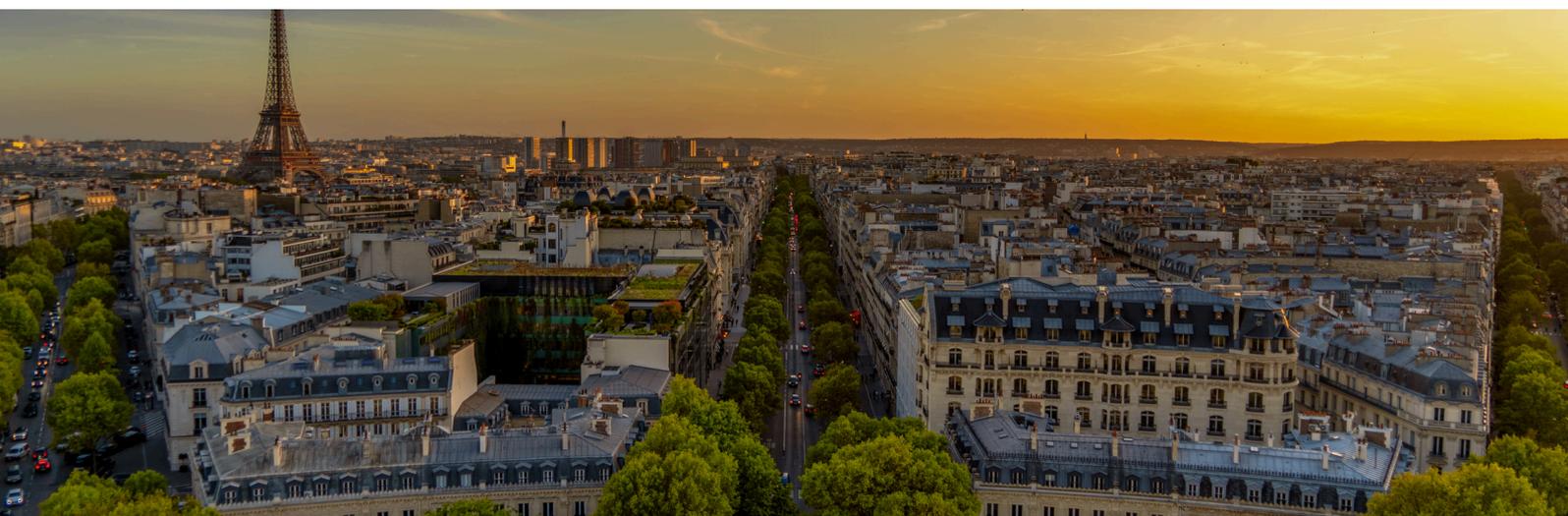
The requirements have been tested in some cases [70]. For example, in a 2023 case, the Defenseur des droits held that a school had breached the CRPA as it could not show that a decision about a student's academic results was not fully automated. It recommended that the Minister of National Education and Youth take steps to ensure that no individual decision was taken in a fully automated manner, and that the Minister should ensure compliance with the transparency obligations provided for in the CRPA.

However, the exceptions to the law exclude several high-stakes systems from its purview, including:

- **When systems are used to "search for infractions":** CADA (Commission d'accès aux documents administratifs), France's regulatory body on access-to-information law has considered that the source code of the risk-scoring algorithm used by CNAF (France's family branch of social security) to identify fraud didn't have to be made public, as it contributes to finding breaches of the law (in this case, fraud), and publishing the parameters could help people and organised groups to game the system. However, it established that past models are communicable [71].

- **To protect the security of information systems:** in 2023, CADA considered that publishing the source code of Parcoursup (the system for allocating higher education places) would expose the security of the IT systems to vulnerability. Higher education institutions using Parcoursup are also exempted from most transparency obligations by law, to **"protect the secret of pedagogical deliberations"** [72].

---

[70] See for instance Défenseur des droits, décision 2023-140 du 26 juin 2023 relative à l'absence de prise en compte des résultats scolaires d'une élève dans le processus d'affectation Affelnet.

[71] Avis n° 20226179 du 15 décembre 2022, see also CADA's 2022-2023 activity report, p.43 : https://www.cada.fr/sites/default/files/CADA_RAPPORT_2022_2023.pdf

[72] Article L.612-3 of the Code de l'éducation

Japan

Overview

Japan takes a risk-based and soft-law approach to regulation of AI. Its government has positioned its flexible process of "agile governance" as a fundamental policy for a digitalised society. It argues that given the speed of evolution of AI technologies, AI governance methods need to be agile and continuously evaluated and updated. There is no distinction made between the use of AI by public bodies and private actors within this approach.

The AI Governance in Japan Ver. 1.1 report, states that "legally binding horizontal requirements for AI systems are deemed unnecessary at the moment." [73]. This is in-line with the notion of "agile governance", which is promoted in Japan.

Japan's AI regulatory policy is based on The Social Principles of Human-Centric AI, which were introduced in 2019 [74].

---

[73] Expert Group on How AI Principles should be implements, AI Governance in Japan Ver 1.1: Report from the Expert Group on How AI Principles Should be Implemented, 9 July 2021 20210709_8.pdf (meti.go.jp)

[74] humancentricai.pdf (cas.go.jp)

Transparency requirement(s)
Japan

Non-statutory duty

Japan's AI regulatory policy is based on the **'Social Principles of Human-Centric AI (2019)'** which promote three basic philosophies:

- **Dignity** - a society in which human dignity is respected;
- **Diversity and Inclusion** - a society in which people with diverse backgrounds can pursue their own well-being; and
- **Sustainability** - a sustainable society [75].

There are seven Social Principles:

1. Human-centric;
2. Promote education/literacy;
3. Protect privacy;
4. Ensure security;
5. Maintain fair competition;
6. Ensure fairness, accountability and transparency; and
7. Promote collaborative innovation.
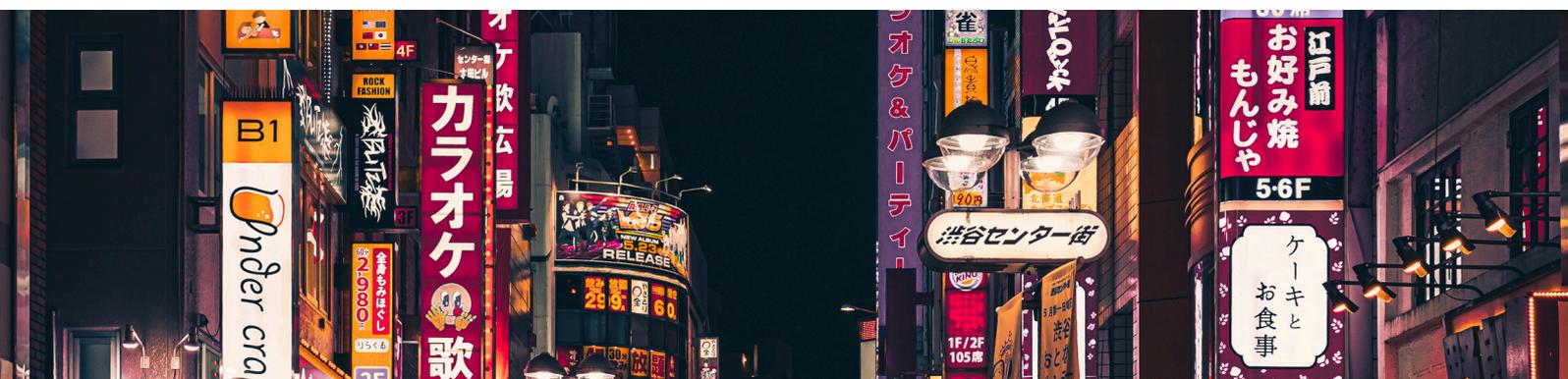
Compliance with these principles is voluntary.

Non-statutory duty

The AI Governance in Japan Ver. 1.1, 9 July 2021, report comprehensively describes Japan's approach to AI regulatory policy. It sets out the Japanese Government's viewpoint on regulation, that a prescriptive, static, and detailed regulation in this context could stifle innovation.

In early 2022, Japan published **'The Governance Guidelines for Implementation of AI Principles'** to present action targets, support the implementation of the Social Principles of AI and provide practical examples of how to achieve them [76].

---

[75] https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf

[76] https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_2.pdf

What the requirements mean

The Social Principles apply to society as a whole. It is noted that state legislative and administrative bodies should pay special attention to these Principles and the social frameworks that the Principles refer to should be implemented across Japanese society including national and local governments as well as in multilateral frameworks.

Based on the Social Principles, companies involved in AI business, typically as a developer and operator, should establish and comply with the goals to be implemented according to the purpose and method of their AI business. The government is said to respect companies' voluntary efforts for AI governance, whilst providing non-binding guidance to support and guide such efforts.

The Social Principles group together **fairness, accountability, and transparency,** stating that it is 'necessary to ensure fairness and transparency in decision-making, appropriate accountability for the results, and trust in the technology, so that people who use AI are not subject to undue discrimination with regard to personal background, or to unfair treatment in terms of human dignity' [77].

In relation to transparency, the Social Principles state that appropriate explanations should be given on a case-by-case basis depending on the application of AI and each particular situation, including such information as when AI is being used, how the AI data is obtained and used, and what measures have been taken to ensure the appropriateness of results obtained from AI operations. They also give value to people being able to understand AI's proposals and make judgments on them, and suggest there should be appropriate opportunities for an open dialogue, as required, regarding the use, adoption, and operation of AI.

Reported effectiveness of requirements

It is difficult to measure the success of the Japanese approach due to common issues with non-binding guidelines, including implementation and monitoring the status of compliance. It has been reported that the rules are ambiguous whilst proponents say this approach offers a pragmatic response to the rapidly evolving AI landscape [78].

---

[77] https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf, page 10.

[78] https://gammalaw.com/japans-new-ai-guidelines-technological-triumph-or-regulatory-riddle/

The United States of America (USA)

Overview

This section focuses on measures at the federal level in the USA, and does not consider proposed or existing provisions for the regulation of AI and the use of algorithms and ADM that have been introduced at the State level [79].

The USA does not yet have specific regulations governing the use of AI in public decision making. However, it is a focus of legislative development, with a number of measures recently proposed [80].

In 2020, the Office of Management and Budget (OMB) released Guidance for Regulation of Artificial Intelligence Applications, a memorandum that provides guidance to federal agencies on how to approach the regulation of AI applications, including considerations for transparency.

In 2022 the White House released a "Blueprint for an AI Bill of Rights", proposing a rights-based regulatory framework.

---

[79] While the current paper focusses on federal measures only, some states are issuing their own Executive Order on AI – see for example Governor Newsom's Executive Order to "Prepare California fopr the Progress of Artificial Intelligence": https://www.gov.ca.gov/2023/09/06/governor-newsom-signs-executive-order-to-prepare-california-for-the-progress-of-artificial-intelligence/

[80] There are a number of Bills that have been introduced (proposed) that are in very early stages of the legislative process. This section will not cover all in detail, as it is unclear whether all Bills will get the support necessary to progress to an Act. For further information see: the Transparent Automated Governance Act 2023 https://www.congress.gov/bill/118th-congress/house-bill/6886/text; the Artificial Intelligence Research, Innovation, and Accountability Act 2023 https://www.congress.gov/bill/118th-congress/senate-bill/3312/text?s=8&r=33&q=%7B%22search%22%3A%22AI+in+the+Federal+Government+Act%22%7D#id46b47245ae1d441788e1d25db9109f3b; the AI Disclosure Act 2023 https://www.congress.gov/bill/118th-congress/house-bill/3831/text.

The White House has introduced two Executive Orders on the use of AI. Following the first in 2020, in October 2023 the White House introduced an Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence [81]. Principle 4 states that AI policies must be consistent with the advancement of equity and civil rights.

On 28 March 2024, the OMB announced that it was releasing the first government wide policy designed to mitigate AI risk [82]. By 1 December 2024, Federal agencies will be required to implement concrete safeguards when using AI in a way that could impact civil rights or safety, including requirements for transparency around the use of AI.

---

[81] https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

[82] https://www.whitehouse.gov/briefing-room/statements-releases/2024/03/28/fact-sheet-vice-president-harris-announces-omb-policy-to-advance-governance-innovation-and-risk-management-in-federal-agencies-use-of-artificial-intelligence/#:~:text=The%20policy%20released%20today%20requires,is%20addressing%20the%20relevant%20risks.

Transparency requirement(s)
USA

Non-statutory duty

In October 2022, the White House released the **Blueprint for an AI Bill of Rights.** In this, the White House Office of Science and Technology Policy identifies five principles that should guide the design, use, and deployment of automated systems to "protect the American public in the age of artificial intelligence." One principle is Notice and Explanation articulated as, "[y]ou should know that an automated system is being used and understand how and why it contributes to outcomes that impact you."

Executive Orders

An Executive Order introduced in 2020 titled **Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government** sought to encourage the design and use of AI in a manner that fosters public trust and confidence while also protecting civil liberties and rights.

Section 3 provides for the following principles for the use of AI in government: [83]

- Lawful and respectful of our Nation's values
- Purposeful and performance-driven
- Accurate, reliable, and effective
- Safe, secure, and resilient
- Understandable
- Responsible and traceable
- Regularly monitored
- Transparency
- Accountable

---

[83] Executive Order 13960, Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government (3 December 2020) https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government.

Section 5 of the 2020 EO required the Federal Chief Information Officers Council (CIO Council) to work with agencies, as appropriate, to provide case inventories of AI use, including current and planned use. Guidance on the provision of case inventories was published by the CIO in 2021 then again in 2023 [84].

In October 2023, an Executive Order titled **Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence** was introduced. It directs agencies with the responsibility of generating, implementing and/or overseeing standards and guidance with respect to AI-related risks [85]. In March 2024, the OMB released a management memo to accompany this Executive Order titled Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence [86]. The memo provides direction to federal agencies on how they can and cannot use AI systems to make decisions about individuals. It goes beyond the Executive Order and details concrete actions that federal agencies are required to take [87].

---

[84] https://www.cio.gov/assets/resources/2023-Guidance-for-AI-Use-Case-Inventories.pdf

[85]Executive Order 14110, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (30 October 2023) https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

[86] Office of Management and Budget, Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence (March 2024) https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

[87] Tech Policy Press, The OMB Memo Shows That AI Can Be Governed (March 2024) https://www.techpolicy.press/the-omb-memo-shows-that-ai-can-be-governed/.
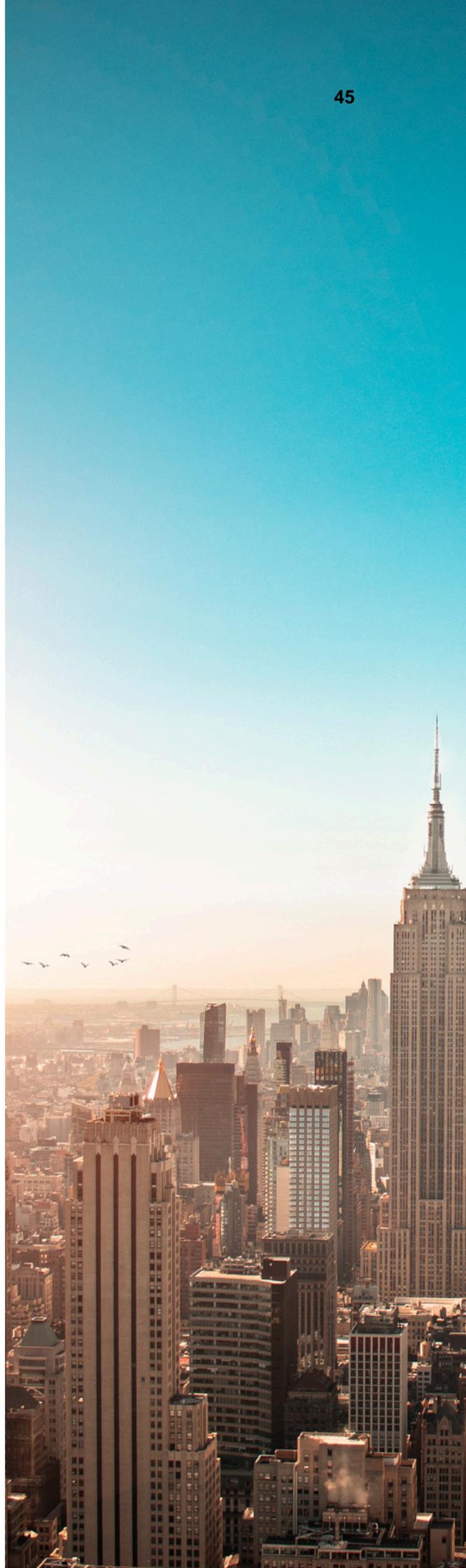
What the requirements mean

The Blueprint for an AI Bill of Rights does not constitute binding guidance for agencies and therefore does not legally require compliance with the principles. It has a two-part test to determine what systems are in scope. It applies to (1) automated systems that (2) have the potential to meaningfully impact the American public's rights, opportunities, or access to critical resources or services.

Executive orders are directives from the President of the United States that manage operations of the federal Government [88]. Although they are not legislation, they have the same effect as laws created through the legislative process and can create binding obligations in relation to federal agencies. While neither the 2020 nor 2023 Executive Orders on AI introduce sanctions for non-compliance with the directions and requirements, if a federal agency fails to carry out the mandate of an Executive Order without good reason, it can follow that the head of that agency will be removed by the President. Political pressures can therefore encourage compliance.

The Executive Order of 2020 applies to AI designed, developed, acquired, or used specifically to advance the execution of agencies' missions, enhance decision making, or provide the public with a specified benefit, and applies to both new and existing uses of AI [89]. It provides principles for the Federal Government to adhere to when designing, developing, acquiring, and using AI. The transparency principle requires agencies to be transparent in disclosing relevant information regarding their use of AI to appropriate stakeholders, including Congress and the public, to the extent practicable and in accordance with applicable laws and policies [90]. This includes making agencies' case inventories of AI use publicly available, to the extent practicable [91]. The Order excludes the Department of Defence and those agencies and agency components with functions that lie wholly within the Intelligence Community.

---

[88] "What is an Executive Order?". Insights on Law and Society. Vol. 17, no. 1. American Bar Association. Fall 2016. ISSN 1531-2461

[89] Section 9(b-c), Executive Order 13960.

[90] Section 3, Principle H, Executive Order 13960.

[91] Section 5 of Executive Order 13960 of December 3, 2020

The Executive Order of 2023 applies to executive departments and agencies and requires them to adhere to principles 'as appropriate and consistent with applicable law' [92]. The Order defines AI in broad terms as including any machine-based system that can make predictions, recommendations, or decisions. Section 10(e) of the Executive Order focusses on transparency and requires the Director of the OMB, on an annual basis, to issue instructions to agencies for the collection, reporting, and publication of agency AI use cases. Through these instructions, the Director is able to expand agencies' reporting on how they are managing risks from their AI use cases and update or replace the guidance originally established in the 2020 Executive Order. Essentially, this requires each agency (except for the Department of Defence and the Intelligence Community) to annually submit an inventory of its AI use cases to OMB and subsequently post a public version on the agency's website [93].

The OMB evaluates the effectiveness of agency programs, policies and procedures, assesses competing funding demands across agencies and sets funding priorities. If agencies do not meet a deadline set by the OMB, this may raise questions as to the effectiveness of such agency's programs, policies and procedures and can impact their funding.

The memo released by the OMB in March 2024 includes the requirement for agencies to provide "public notice and plain-language documentation" through the maintenance of an AI use case inventory [94]. The information submitted about the AI systems functionality to the use case inventory must be adequately detailed and generally accessible so that it provides public notice of the AI to its users and the general public. The memo also suggests that agencies must provide reasonable and timely notice about the use of AI and a means to directly access any public documentation about it in the use case inventory [95].

---

[92] Section 2, Executive Order 14110.

[93] Section 225 of the Advancing American AI Act, https://www.whitehouse.gov/wp-content/uploads/2023/11/AI-in-Government-Memo-draft-for-public-review.pdf

[94] Office of Management and Budget, Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence (March 2024), page 21 https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

[95] Office of Management and Budget, Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence (March 2024), page 21 https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

Reported effectiveness of requirements

The Blueprint has received some criticism for not sufficiently addressing important issues, such as those surrounding educational access, worker surveillance, and most uses of AI in law enforcement. Further, it has been noted that the White House has yet to effectively coordinate and facilitate AI regulation [96].
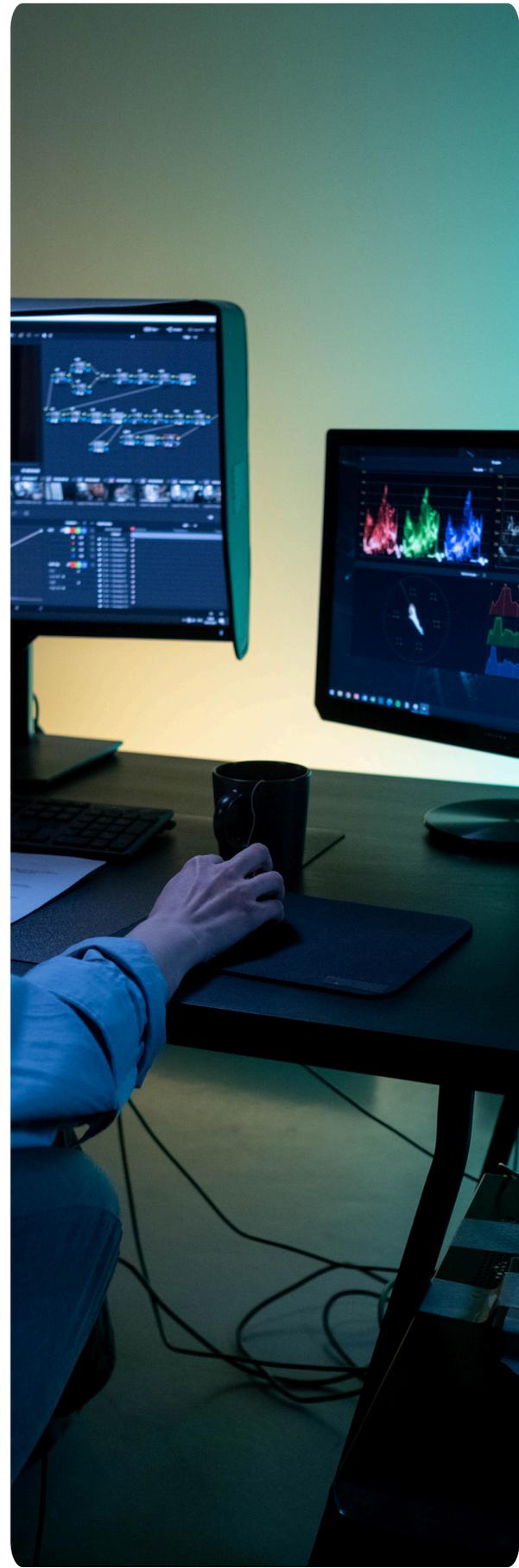
The 2020 Executive Order only provides guidance to policy makers and does not create strict legal duties for agencies using AI in decision making and therefore it is difficult to determine its effectiveness. The requirement in Section 5 to provide case inventories appears to have limited compliance, with a study in 2023 showing nearly half of agencies failed to publicly issue AI use case inventories, even where there were known uses of machine learning [97].

The 2023 Executive Order and OMB memo have been said to 'set the federal government up to be a model for accountable AI' and that the regulation of its own use is 'significant' [98]. The EO places a number of obligations on executive departments and agencies to monitor and publish details of AI use.

[96] The AI Bill of Rights makes uneven progress on algorithmic protections | Brookings

[97] Christie Lawrence, Isaac Cui, and Daniel Ho. 2023. The Bureaucratic Challenge to AI Governance: An Empirical Assessment of Implementation at U.S. Federal Agencies. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES '23). Association for Computing Machinery, New York, NY, USA, 606–652. https://doi.org/10.1145/3600211.3604701

[98] https://www.brookings.edu/articles/how-the-ai-executive-order-and-omb-memo-introduce-accountability-for-artificial-intelligence/

While the Executive Order itself was broad, the guidance provided by the OMB memo (intended to be read together with the EO) provides more detailed directions, including for agencies to proactively monitor AI, regularly report on use cases and make that information detailed and accessible. If adhered to, this requirement is said to be a key transparency and accountability mechanism [99].

However, commentators have identified problems with the previous practices of the AI use case inventory, including both that agencies left known AI uses off their inventory and that the reporting requirements were minimal and did not include testing and bias assessment results [100]. Effectiveness of the AI use case inventory as a transparency and accountability mechanism is said to depend on whether existing loopholes and under-reporting concerns are addressed through the OMB process to come [101]. The March 2024 memo references draft guidance on AI reporting for agencies, which provides information on the criteria, formal and mechanisms for federal agencies to standardise the submission of use case inventories, which may go some way to address these concerns [102].

---

[99] https://www.brookings.edu/articles/how-the-ai-executive-order-and-omb-memo-introduce-accountability-for-artificial-intelligence/

[100] Rebecca Heilweil and Madison Alder, The government is struggling to track its AI. And that's a problem, Feedscoop (3 August 2023) https://fedscoop.com/the-government-is-struggling-to-track-its-ai-and-thats-a-problem/; Stanford University, Human-Centred Artificial Intelligence, Im plementation Challenges to Three Pillars of America's AI Strategy (December 2022) https://hai.stanford.edu/white-paper-implementation-challenges-three-pillars-americas-ai-strategy.

[101] https://www.brookings.edu/articles/how-the-ai-executive-order-and-omb-memo-introduce-accountability-for-artificial-intelligence/

[102] Office of Management and Budget, Draft guidance for 2024 agency artificial intelligence reporting per EO 14110 (March 2024) https://www.whitehouse.gov/wp-content/uploads/2024/03/DRAFT-Guidance-for-Agency-Artificial-Intelligence-Reporting-per-EO14110.pdf.
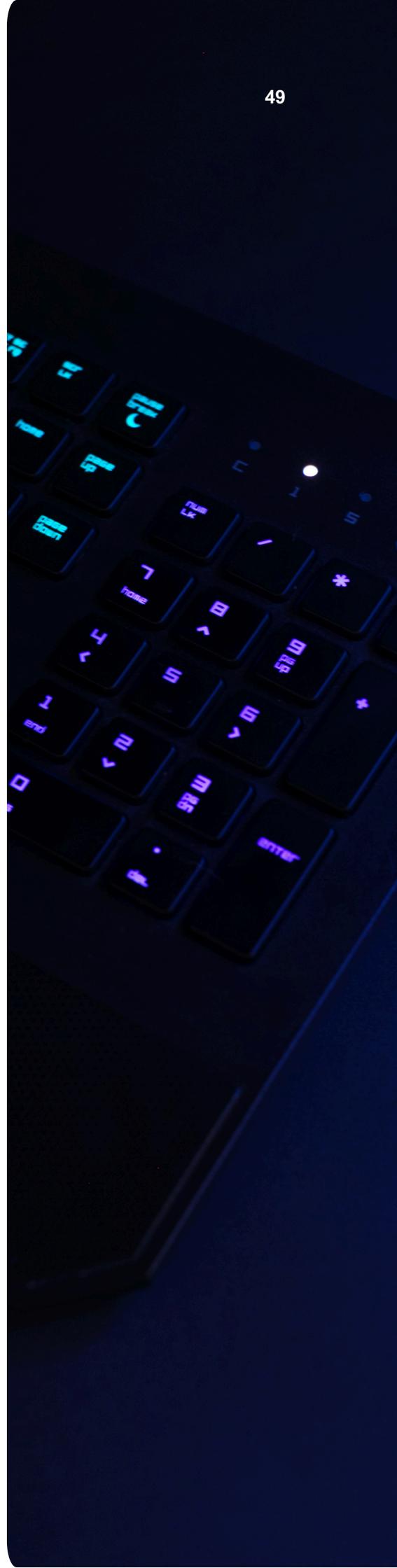
Comparative analysis of transparency requirements

The jurisdictions selected for this paper are representative of a broad range of approaches to regulating AI and ADM, spanning the spectrum of approaches from comprehensive and hard-law to sector specific and soft-law. This section will assess the strategies and requirements from the five jurisdictions against the type of transparency they aim to secure on two levels:

1) **Individual** – transparency for decision subjects and those affected by an AI or ADM system. Transparency on this level should allow individuals to understand whether automation and/or AI has been used in the process of making a decision about them, how the tool or system operates and the role it played in the decision reached.

2) **Systemic** – transparency to the broader public about the use of AI and ADM by public authorities. Transparency on this level should allow for wide-ranging sections of the public to be able to understand the landscape of public authority use of AI and ADM. It should allow for meaningful understanding of which bodies are using AI and ADM, in which policy areas such tools and systems are being rolled out, the role it plays in public administration and decision-making, the reason behind its introduction, and impact on processes.

Individual level transparency

All of the jurisdictions within scope of this paper, other than **Japan**, have put in place specific requirements that aim to secure transparency on the individual level. Whilst one of the Social Principles of Japan's regulatory policy is **'fairness, accountability and transparency'**, it is not accompanied by specific requirements to secure transparency from AI developers or users. Instead, the Social Principles encourage the availability and recognise the value of appropriate explanations on a case-by-case basis. The absence of specific and clear requirements means there is no guarantee that individuals in **Japan** are able to secure transparency around the use of AI or ADM in decisions that affect them. Public authorities are also under no obligation to make the relevant information available or to make efforts to communicate it effectively to individuals. With such broad scope for interpretation of the principles and guidelines, the extent of implementation is left to the discretion of authorities. Individuals would not have certainty regarding the information that should be available to them, their ability to request it or the ability to seek redress if the principles are not met.

Across the other four jurisdictions explored within this paper, there is some consistency across the types of requirements introduced regarding individual level transparency.

The requirement to explicitly inform decision subjects that they have been subject to AI or algorithmic processing when they are used to take individual administrative decisions is present within the **French framework**. Within the **Canadian regulatory framework**, individuals would similarly be notified that they have been subject to a decision which was reached with the use of AI or algorithmic processing, but through the provision of an explanation of the decision itself which would serve as a notification by informing the individual of the presence of AI or algorithmic processing. Under the regulatory framework in **Canada**, public authorities must provide meaningful explanations of their AI and algorithmic informed decisions to affected individuals. The requirements oblige authorities to provide information about how the decision was made and why the decision was made. The circumstances in which explanations will be provided (and the circumstances in which all other transparency requirements apply under the DADM and Policy on Service and Digital) are also limited because provisions apply only to decisions made by the federal government. This means there is no requirement to notify affected individuals or provide explanations regarding tools or systems used by provincial governments, municipalities, or provincial agencies such as police services, child welfare agencies and/or many other important public institutions [103].

[103] Law Commission of Ontario and The Chair on Accountable AI in a Global Context, Comparing European and Canadian AI Regulation (November 2021) https://www.lco-cdo.org/wp-content/uploads/2021/12/Comparing-European-and-Canadian-AI-Regulation-Final-November-2021.pdf.

Although not required proactively, explanations are available on request under **France's** regulatory framework. The explanation should include information regarding the degree and method of contribution of the algorithmic process in the decision-making process, the processed data and its sources, the criteria and factors as applied to the individual's situation, and the operations carried out by the processing. All this information should be conveyed in an intelligible form and should not infringe upon legally protected secrets.

These provisions, both those requiring notification and those requiring the provision of an explanation, go further than what is in place in the **UK** as they apply beyond solely ADM and require explicit notification in instances where AI, algorithms and automation assist in decision-making.

The principle of **"Notice and Explanation"** within the **US Blueprint for an AI Bill of Rights** would seek to ensure that individuals understand how and why an automated system contributes to outcomes that impact them. Achieving this would likely require the provision of tailored explanations, either proactively or on request. As noted above, it is difficult to assess how effective the Blueprint has been, however it is conceivable that public authorities might interpret such a broad principle and stated desired outcome differently. Therefore, affected individuals might not be guaranteed to receive explanations and even where they do, they might differ in detail and information provided, which would affect the extent to which they are meaningful.

Taking the above requirements and guidance from **Canada, France and the USA** together, it is evident that there is a cross-jurisdictional drive towards securing individual level transparency through regulation and legislation around the use of AI and ADM systems. This trend goes beyond existing requirements to provide information directly to individuals where solely automated decision-making has occurred, found within legal frameworks such as the **UK and EU GDPR**. This is of particular importance because most decisions made by public authorities that utilise AI or algorithms do not rely on the tool or system to make the entire or final decision. It is more common that ADM systems are intended to act as decision-support rather than decision-making tools, with a 'human in the loop' that takes the final decision. This leaves a gap within transparency requirements intended to secure individual transparency, leaving the more common contemporary usage outside of the scope. The requirement to notify individuals of the presence of this technology even in the process of decision-making and the requirement to provide explanations of how the tool interacts with the decision-making process, both proactively and on request, are effective mechanisms for securing some level of understanding for those subjected to and affected by decisions made with the assistance of AI or similar technologies.

The sole notification of the presence of AI in a decision-making process will not alone provide meaningful transparency. To know that AI or a similar technology was used, puts the individual in a better position to ask further questions about the process or the role of the technology in reaching the specific decision, but that requires an individual to take active interest in the matter and have the time and capacity to pursue correspondence with the public authority about this specific element of the decision they received. This type of disclosure, as opposed to proactive transparency, can place "a tremendous burden on individuals to seek out information about a system, to interpret that information, and determine its significance" [104].

A similar burden is likely to occur where explanations are only available upon request by the individual. Transparency will not be meaningfully achieved if individuals need to know what to ask for and how to ask for it. There is an asymmetry in power between data controllers, AI and ADM systems operators, and those subject to public authority decisions. Public authorities already have the relevant information more or less to hand and have an established line of communication with the individual due to the need to notify them of the decision made.

Contrastingly, in these circumstances a decision subject would either need to submit further questions to the public authority following the notification of the presence of technology in the decision-making process or request an explanation of how the decision was reached. In both of these instances, some of the burden for securing transparency is shifted to the individual and the level of disclosure rests on their capacity and appetite for further enquiries.

---

[104] Ananny, M. and Crawford, K., Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability, New Media and Society, 2016, 7 https://journals.sagepub.com/doi/10.1177/1461444816676645
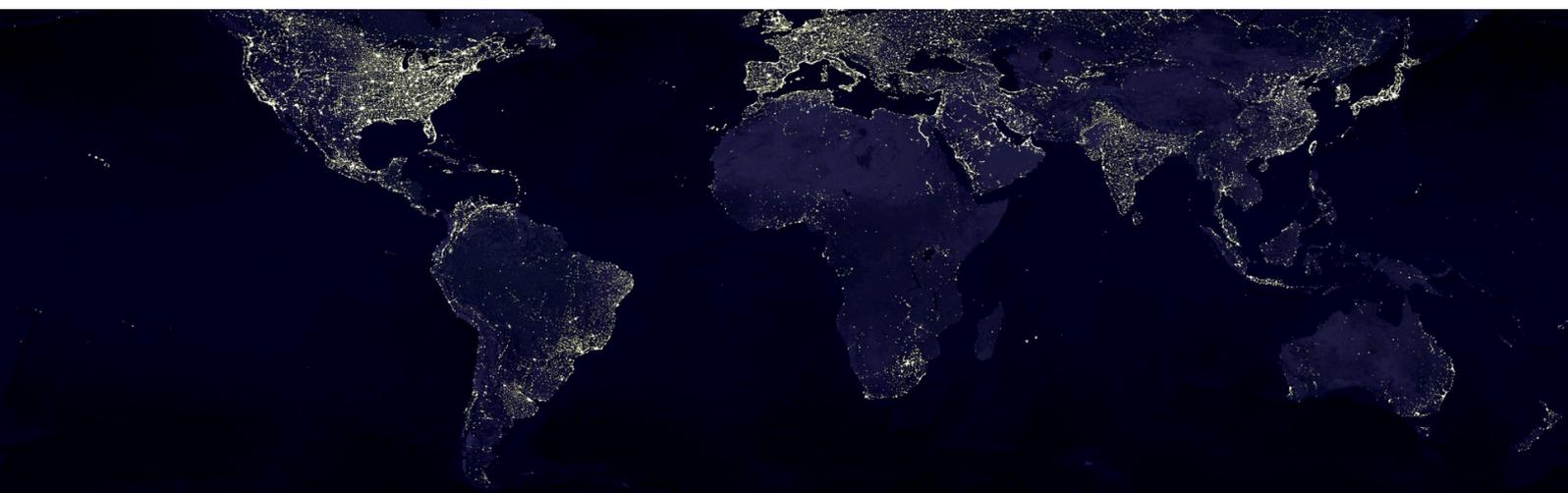
Having the ability to request an explanation, or the requirement for public authorities to provide one on request, gives the illusion that explanations are widely requested and provided. Edwards and Veale highlight the "notice and choice fallacy" of requiring individual consent to share data via privacy policies that are largely not read or properly understood [105]. They underscore that the availability of an explanation could similarly become an empty formality or a "transparency fallacy". Where requirements on public authorities such as those in France do not require proactive disclosure of an explanation, it is possible that the requirement to provide or right to request an explanation is at risk of perpetuating a transparency fallacy where methods for securing disclosure are available but not properly engaged with, and information is not further disclosed.

The risk of transparency fallacy is mitigated by the requirement for public authorities to proactively provide explanations. The requirement on public authorities in Canada to provide an explanation, which includes certain specified categories of information is likely to support consistent levels of disclosure.

For such explanations to be consistent and meaningful, it is more effective for the categories of information that must be included within the explanation to be specified so as to set a baseline for the quality and usefulness of explanation provided. The vague nature of the US principle and requirement would likely result in inconsistent provision of explanations, varying levels of quality and uncertainty in the extent to which the explanation received is able to help an individual to understand that an AI tool or system was used and the effect it had on the decision they received.

---

[105] Edwards, Lilian and Veale, Michael, Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'? (2018). IEEE Security & Privacy (2018) 16(3), pp. 46-54, 7 https://ssrn.com/abstract=3052831.

Systemic level transparency

There is some commonality between methods for achieving systemic level transparency across the jurisdictions with robust transparency requirements for public authorities.

Four out of the five jurisdictions have put in place provisions to either encourage or mandate public authorities to make accessible to the general public information about their use of AI or algorithmic tools. Japan is the only jurisdiction included within this paper that does not have a requirement of this type.

The details of the requirement vary across the jurisdictions in terms of both their legal status and the specific requirement they place on public authorities. In France, the requirement is a statutory one and therefore places a legal requirement on public authorities to comply with it. However, the requirement is less prescriptive than that in other jurisdictions. The CRPA requires public administrations to publish online the rules defining the algorithmic processing used in the performance of their tasks when such requirements are the basis of individual decisions. In making this information available, public authorities are increasing the level of systemic transparency by publishing online information regarding their use of AI and algorithmic tools.

However, there is no evidence that this information is collated in to one specific online location or logged in a public database, inventory, or hub, which is likely to make it difficult for individuals to find the information. As noted in the section on France at page 34, public authorities are struggling to fulfil these requirements partly because there is a lack of guidance on how algorithms should be inventoried, what information must be included and how the information needs to be presented [106].

Other jurisdictions have taken a more prescriptive approach, requiring in various degrees the publication of information. This includes prescribing specific categories of information which are then collated and published online in repositories of information about algorithmic systems, which emphasise public access. Requirements to engage with systems of this kind are present in Canada, the EU and the US.

---

[106] Open Government Partnership, France: Transparency of Public Algorithms (FR0035) opengovpartnership.org/members/france/commitments/FR0035/

In Canada, this takes the form of a collection of completed AIAs on the Government of Canada website [107]. There is an assessment questionnaire of around eighty questions to be completed by departments and agencies using ADM, designed to help assess and mitigate the impacts associated with deploying an automated decision system. The questionnaire is broad and prompts the submission of information across a range of categories of information, such as who is accountable for the project, the reasons for introducing automation into the decision-making process, details of the algorithm itself, whether the process is fully or partially automated and data sources. Both the completion and release of the questionnaire is required by the DADM.

A final version of the AIA is then required to be publicly posted on Government of Canada websites or the Canadian 'Open Government Portal'. It is the release, or publication, of completed AIAs and the collation of these documents online that allows for public access to information about AI or algorithmic systems. The requirement to complete a formulaic questionnaire means that the categories of information contained in these public facing documents are consistent if completed in full. The questionnaire also provides certainty for departments and agencies who are required to engage with the progress by being upfront about what information needs to be collected and setting a benchmark for disclosure.

The Canadian DADM does not have the legal status of a statute nor is it voluntary; it falls somewhere in between. Professor Teresa Scassa notes the importance of the Directive within the federal government, and emphasises that there are accountability frameworks to ensure compliance, but the requirements to comply with directives are internal to government, as are the sanctions and therefore there are no actionable rights for individuals or organisations [108]. The practical effect of the lack of meaningful compliance mechanism can be seen in the low number of publicly available AIAs on the open government portal. As mentioned at <u>page 25</u> 21 tools have been disclosed through the completion and publication of AIAs whereas 303 have been logged in the independent Canadian TAG Register, 95% of which have been used at the federal level and therefore would be subject to requirements under the DADM. This is a stark difference that demonstrates two things. First, the number of systems and tools that are in use in opacity at the federal level in Canada. Second, that the requirement to complete and publish AIAs is lacking strength and bite, resulting in many federal agencies opting not to comply but facing no sanction.

---

[107] Canada, Directive on Automated Decision-Making, (2019), section 6.1.

[108] Teresa Scassa, Administrative Law and the Governance of Automated Decision-Making: A Critical Look at Canada's Directive on Automated Decision-Making (October 30, 2020) https://ssrn.com/abstract=3722192 at 6-7.

In the EU, the requirement to register AI systems in the EU database only applies to high-risk AI systems and the requirement is placed on providers (or developers) of AI systems rather than users or deployers. In the context of securing transparency for public sector use of AI, the framing of this requirement might limit the straight-forward registration of AI systems by public authorities. However, public authorities deploying high-risk AI systems are required to register themselves in the database and select systems that they envisage to use. Whilst helpful in linking public authority use to the high-risk systems registered in the database, it is somewhat counteractive to the aim of transparency that high-risk AI systems used in high stakes environments such as law enforcement, migration, asylum and border control management, will not be included in the public section of the database [109]. It will be necessary to see how the requirement to register high-risk AI systems plays out through the implementation of the EU AI Act, but it is likely that many public sector uses of AI will not be publicly logged on the EU database either due to falling outside the scope of high-risk, not being registered in the first place by providers, or being stored in the non-public section of the database.

For those systems that are registered in the database, there is a determined list of information to be submitted by the providers of high-risk AI systems [110]. To highlight a few categories of information from the list, providers are required to submit information regarding accountable or responsible individuals, the intended purpose of the AI system, a basic description of the information used by the system and its operating logic and can choose to submit a URL for additional information. Deployers (likely to be public authorities) of high-risk AI systems are required to submit, alongside other information, a summary of the findings of the fundamental rights impact assessment conducted as required under Article 29a [111].

Despite the likely narrow scope of the requirements in the EU, the establishment of a defined hub in the form of a public database that holds a collation of information on AI systems on the market and in deployment is a positive example of a clear and prescriptive attempt to obtain information that can be displayed to the public. The legal status of the requirements and clear determined list of categories of information to be submitted will likely mitigate the risk of vagueness that may disincentivise engagement.

---

[109] EU AI Act, Recital 69.

[110] EU AI Act, Annex VIII : Information to be Submitted upon the Registration of High-Risk AI Systems in Accordance with Article 51, Section A.

[111] EU AI Act, Annex VIII : Information to be Submitted upon the Registration of High-Risk AI Systems in Accordance with Article 51, Section B; EU AI Act, Chapter 3: Chapter 3: Obligations of Providers and Deployers of High-Risk AI Systems and Other Parties, Article 29a: Fundamental Rights Impact Assessment for High-Risk AI Systems.

Requirements in the USA under the October 2023 Executive Order go some way to achieving repositories of information about algorithmic systems. As set out in more detail at page 41 on requirements in the USA, each governmental agency (other than the Department of Defence and the Intelligence Community) is required to annually submit an inventory of its AI use cases to the Office of Management and Budget (OMB) and post a public version of the inventory on the agency's website [112].

Under this requirement, there would be a separate inventory of AI or ADM systems used by each government agency which would provide coherent oversight of specific agency use, but not of public sector AI usage more generally. Whilst this undoubtably increases levels of transparency, it misses the opportunity for these submissions to be put into a centralised repository. There is value in having a centralised repository of all public sector AI and ADM use as it allows for a picture to be built up of the landscape of AI usage across government and insight into where such systems are being deployed and for what purposes. However, the likelihood of individuals being able to locate this information is increased in the US due to the suggestion that agencies should include use case inventory documentation or a link to it in contexts where people will interact or be impacted by the AI.

This would usefully alert those who will be subject to the AI or ADM system to not only its existence but also to the more detailed information about the particular system and the availability of further information on the agencies use of AI more generally.

The US requirement might also be limited in effectiveness by only requiring the submission of AI use cases annually. By setting the benchmark for submission as once a year, it is unlikely that government agencies will submit an inventory of AI use cases more regularly meaning that information will be up to a year old before updated. There are no specific requirements or guidelines of what information should be submitted by government agencies regarding the AI systems, other than that information on its functionality must be adequately detailed and generally accessible. Contrasting this to the clear requirements in both the EU and Canada, it is likely that the use case inventories in the US will differ between agencies and even AI systems and tools meaning the level of transparency secured through this requirement will likely not be consistent.

---

[112] Section 225 of the Advancing American AI Act, https://www.whitehouse.gov/wp-content/uploads/2023/11/AI-in-Government-Memo-draft-for-public-review.pdf

Conclusion and recommendations

This section sets out four recommendations for the UK Government. The recommendations are informed by the overview of transparency requirements across the five jurisdictions within the scope of this paper and the analysis of their effectiveness for securing transparency around the use of AI and ADM by the public sector.

If adopted, these four recommendations would allow the UK to better implement the principle of 'appropriate transparency and explainability' as set out in the AI Regulation White Paper.

Individual level transparency

In the development of the UK's AI regulatory framework, greater consideration must be given to the need for individual level transparency. Transparency on this level should allow individuals to understand whether automation or AI has been used in the process of making a decision about them, how the tool or system operates and the role it played in the decision reached. We see from the analysis in the previous section a variety of methods for securing individual level transparency. The two recommendations that follow take into account this analysis, particularly the need for proactive transparency to be led by public authorities.

1) Public authorities should notify individuals of the presence of an AI, algorithmic or automated tool or system.

- Public authorities should be required to explicitly inform decision subjects or those affected by a decision or action taken by a public authority about the use of an AI, algorithmic or automated tool or system when communicating the decision to them.
- This should apply in instances where the tool or system has been used to partially make or support a decision-making process as well as to solely make a decision.

2) Explanations should be provided proactively to individuals.

- Alongside the notification of the presence of an AI, algorithmic or automated tool or system, public authorities should be required to proactively provide explanations to affected individuals, providing information on how and why the decision was reached. The proactive provision of an explanation avoids placing the burden of requesting an explanation, or specific information, on the individual.
- The requirement for public authorities should include specific categories of information to be included in the explanation, such as tailored information on the contribution of the AI, algorithmic or automated tool or system in the decision-making process, including the tasks performed by the system.
- The explanation should make clear that individuals may request further information specific to the decision they received, such as how the data is obtained and used, provider details, purpose, accuracy, measures taken to ensure the appropriateness of results, and the presence of human oversight and intervention mechanisms.

Systemic level transparency

The UK is in an encouraging position when it comes to the development of provisions to secure systemic level transparency. Transparency on this level should allow for wide-ranging sections of the public to be able to understand the landscape of public authority use of AI and ADM. It should allow for meaningful understanding of which bodies are using AI and ADM, in which policy areas such tools and systems are being rolled out, the role it plays in public administration and decision-making, the reason behind its introduction, and impact on processes.

The UK's Algorithmic Transparency Recording Standard (ATRS) if properly engaged with could put the UK in a leading position when it comes to systemic level transparency. The analysis in the previous sections shows commonality in attempts to devise online repositories of information about algorithmic systems, which emphasise public access.

The limitations of these methods in other jurisdictions centre around a lack of clarity in terms of the requirement or what categories of information to provide. The report templates under the ATRS enable clarity and provide specific and clear guidelines to authorities regarding the information to be provided.

The ATRS Hub is also a positive element of the UK's methods for securing systemic level transparency. If public authorities in the UK submitted reports regarding all uses of AI and ADM that fall within its scope, the Hub would act as a centralised repository of public sector AI and ADM use. This would allow for a picture to be built up of the landscape of AI usage across government and provide insight into where such systems are being deployed and for what purposes.

The **two recommendations** that follow take into account this position and would allow for the UK to properly implement the ATRS and allow it to reach its full potential.

3) Statutory requirement for submission of reports to the Algorithmic Transparency Recording Standard (ATRS) Hub.

- The recent commitment made in the Government's response to the AI Regulation White Paper to make the ATRS a requirement for all government departments is a step in the right direction. To ensure full engagement with the requirements under the ATRS, compliance should be mandated through a legally enforceable requirement.

4) In contexts where people will interact with or be impacted by an AI, algorithmic or automated tool or system, the ATRS should be mentioned and a link to the ATRS Hub should be provided.

- This will effectively notify individuals of the existence of the ATRS and direct them toward the information.
- This would allow individuals and the general public to understand in advance how a decision will be made and inform them of the presence of AI or ADM within the process, rather than only after the decision has been made and they have been affected by it.

Acknowledgements

Reviewers

Methods

The recommendations presented in this paper are grounded in rigorous comparative research of regulatory requirements in diverse jurisdictions, where the approach to transparency requirements vary.

The conclusions of this policy paper have been informed by a range of evidence, including:

- Public Law Project's desk-based research.

- Two stakeholder roundtables.

- Independent comparative legal research commissioned by Public Law Project and carried out by Fieldfisher.

- Expert insights on the requirements and their effectiveness across the different jurisdictions.