# Discover Artificial Intelligence

Discover

**Research**

# Public sector AI transparency standard: UK Government seeks to lead by example

Nigel Kingsman[1] · Emre Kazim[1] · Ali Chaudhry[2] · Airlie Hilliard[2] · Adriano Koshiyama[2] · Roseline Polle[2] · Giles Pavey[3] · Umar Mohammed[2]

## Abstract

In releasing the Algorithmic Transparency Standard, the UK government has reiterated its commitment to greater algorithmic transparency in the public sector. The Standard signals that the UK government is both pushing forward with the AI standards agenda and ensuring that those standards benefit from empirical practitioner-led experience, enabling coherent, widespread adoption. The two-tier approach of the Algorithmic Transparency Standard encourages transparency inclusivity across distinct audiences, facilitating trust across algorithm stakeholders. Moreover, it can be understood that implementation of the Standard within the UK's public sector will inform standards more widely, influencing best practice in the private sector. This article provides a summary and commentary of the text.

## 1 Introduction

The policy, legal and governance debate regarding managing the risks and opportunities of automated systems (we use 'artificial intelligence' (AI) in the broadest sense of algorithmic systems) continues to grow at pace. Indeed, with respect to both UK institutions and international bodies, the turn has moved from abstract concerns of harm and impact (which we read the early literature of 'AI Ethics' as encompassing) to concretisation via proposed regulations and standards (which is very much where we are now [1]). In this paper, we offer our initial comments on the recently published Algorithmic Transparency Standard [2]. We read the text as a step forward by the UK government in both driving the AI standards agenda and ensuring that the standards are informed by practice—via the piloting—(thereby that they will be both practical and meaningful)—see [3] for discussion and [4, 5] for advocacy of the approach. We provide an overview of the proposals and offer initial commentary on various standouts. Our main takeaways are:

- *Signal:* we read the text as a signal that this is a priority for the government. Notwithstanding this, we believe that it is critical that this moves forward through testing and an encouragement for industry to follow suit.
- *Method:* a public sector approach to driving through standards; here we recognise (implicitly) the power of government procurement in driving standards through industry, and also comment that not codifying too early (through general legislation) will ground eventual (potential) regulation and standards in an empirically informed manner.

✉ Nigel Kingsman, nigel.kingsman@holisticai.com | [1]UCL, Holistic AI, London, UK. [2]Holistic AI, London, UK. [3]Unilever, London, UK.

Springer

- *Public transparency:* the two-tier approach is a reasonable approach for catering to distinct audiences and, more generally, we recognise that there will always be the problem of communicating to different stakeholders. Here we also note the difficulty in explaining algorithmic systems in a legible manner.
- *Avoids data protection debate:* recognizing that the data protection debate is still very active (in terms of where the UK will move viz. potentially liberalizing on this front), with finalization of the UK's data protection framework not a prerequisite for proceeding with the Algorithmic Transparency Standard.

We conclude with comments on what we take to be a *transparency first approach* to algorithmic accountability, and with a note welcoming initiatives to pilot the Transparency Standard and make those *pilots public*. We do this by emphasizing the importance of empirically informed standards, evidenced through case-study and practitioner engagement. In addition, we highlight some of the omissions and areas for future development. Whilst it is potentially right to start with practitioner pilots, true transparency requires explainability to citizens and this is an area yet to be fully dealt with and developed (here concerns regarding the digital divide surface).

## 2 Overview

In this section, we provide an overview of the Algorithmic Transparency Standard.

The Standard can be seen as a continuation of the discussion around the UK National Data Strategy [6] and the subsequent UK National AI Strategy [7]. In particular, the Standard responds to the suggestion, itself in response to the National Data Strategy consultation, that for 'enabling accountability in practice, one concrete recommendation is the use of algorithmic registers' [8] which cites the cities of Amsterdam and Helsinki as having "launched algorithmic registries detailing how city governments use algorithms to deliver public services". Moreover, the Standard is a concrete example of the 'public sector as an exemplar' [9], 'using the public sector as both an adopter and innovator of AI' [9], and using the public sector to 'set an example for the safe and ethical deployment of AI' [10] (see also [11]). Briefly, according to our reading, the UK is taking steps in establishing an ecosystem of trust (governance)—potentially an alternative to the EU's regime—that is underpinned by an innovation and opportunity agenda with a view to driving positive geopolitical and economic outcomes [9].

Whilst this section reproduces the Standard, it is presented as a reading of the text within which the overview, together with commentary, is interwoven. We have divided the overview into five sections, namely:

- Transparency tiers
- Methodology and purpose
- What exactly is the algorithmic part doing?
- Risk assessment
- Ownership

For each, we reproduce the original text and then offer a comment below.

### 2.1 Transparency tiers

*You must fill in both sections of the template. For the:*
- *Tier 1 information, provide a short non-technical description of your algorithmic tool, and an overview of what the tool is and why the tool's being used*
- *Tier 2 information, provide more detailed technical information, such as specific details on how your tool works and the data the tool uses*

We see the transparency tiers as addressing the **reporting** part of the AI accountability debate, which includes providing evidence that AI risk management has occurred. According to our view, there are two dimensions to reporting: what needs to be reported and to whom.

- **Level of reporting:** this aligns directly with the risk level of the system. Quite simply, the higher the risk level the more extensive the reporting demand.
- **To whom:** there are a number of critical stakeholders to whom the reporting is directed towards.

**Regulators:** sector-level and cross-sector regulatory bodies having the capacity for censure for compliance failures;
**Persons directly impacted:** end clients and potential customers that might be disadvantaged by the decisions informed by the system;
**Public:** here we include citizens, other end clients and potential customers;
**Other third parties:** journalists, activist groups, academics, researchers;
**Internal use:** system owners, internal control and compliance functions, management vertical.

In particular, the transparency tiers bifurcate the algorithmic transparency by audience. The more concise, non-technical tier 1 information is most suited for the 'persons directly impacted' and 'public' categories, and the provision of the detailed technical tier 2 information, in addition to the tier 1 information, is most suited for the 'regulators', 'internal use' and 'other third parties' categories. In general, we endorse the view that transparency should take into account the nature of the audience and hence think that this two-tier approach is an appropriate starting point in recognising this and structuring communication. We recognise that this approach deviates from that proposed by the EU [1] which specifies an information requirement for 'high-risk AI systems' only, and makes no provision for distinct audiences. Similarly, the 'Algorithmic Justice and Online Platform Transparency Act' [12] and the 'Data Accountability and Transparency Act of 2020' [13], both of the US, make no such bifurcation. In the Asia Pacific region, China's 'International Information Service Algorithm Recommendation Management Regulations' [14], although "demanding more transparency over how the algorithms function" [15], make no mention of distinct transparency provision by audience, and Australia, whilst not having specific laws regulating algorithmic decision making [16], nonetheless addresses transparency through disclosure within 'Australia's AI Ethics Principles' [17], but again does not explicitly seek differentiation by audience. Similarly, the existing algorithmic registers for the cities of Amsterdam and Helsinki [18] and the Government of Canada [19], do not use a tiered approach by audience.

## 2.2  Methodology and purpose

### 2.2.1  Tier 1 information

Explain:
- How your tool works
- What problem you're aiming to solve using the tool, and how it's solving the problem
- Your justification or rationale for using the tool

### 2.2.2  Tier 2 information

Describe the scope of the tool, what it's been designed for and what it's not intended for.
Expand your justification for using the tool, for example:
- Describe what the tool has been designed for and not designed for, including purposes people may wrongly think the tool will be used for (2.1)
- Provide a list of benefits—value for money, efficiency or ease for the individual (2.2)
- List non-algorithmic alternatives you considered, if this applies to your project, or a description of your decision-making process before you introduced the tool (2.3)

List the tool's technical specifications, including:
- The type of model, for example an expert system or deep neural network (2.4)
- How regularly the tool is used—for example the number of decisions made per month, or number of citizens interacting with the tool (2.5)
- The phase—whether the tool is in the idea, design, development, production, or retired stage including the date and time it was created and any updates (2.6)
- The maintenance and review schedule, for example specific detail on when and how a human reviews or checks the automated decision (2.7)
- The system architecture (2.8)

We see the information requirements demanding the 'what?' and 'why?' of the system as addressing the **risk mapping** part of the AI accountability debate, centred upon what exactly needs to be identified in terms of risks and how those risks can be assessed in terms of level of risk.

- **Identifying relevant vertical(s):** Regarding what needs to be identified, according to our view, the verticals of bias, explainability, robustness and privacy correspond in some sense to the technical/engineering aspects of a system, whereas the verticals of financial, reputational and (forthcoming/existing) compliance risk correspond in some sense to non-technical governance concerns [20]. *A strong view of ours is that this demarcation of technical and non-technical is dissolved when AI assurance as a whole is considered—the mediation of this, we see, in terms of reporting (see below).*
- **Risk level:** it is critical that the potential harm that a system may pose is determined. We see this in terms of how the system is used in the sense of end purpose—a system used to determine the social welfare provisions of a person is clearly high risk, whereas a system used to detect spam is not. Similarly, an autonomous system taking the final decision might be considered more risky compared to an AI system that enables more informed decision making by the 'human in the loop'.

According to our view, both the intended use of the system and the methodologies and technologies employed by the system are key drivers of which technical risk verticals deserve most attention. For example, if the end users for the system are intended to be the general public, bias might come to the fore with the focus on thorough data and code analysis, whereas if the system relies on deep learning architectures, explainability might come to the fore with the focus on understanding the tool's decisions by a range of stakeholders. However, most critically, it is the intended use that will determine the level of risk and hence the extent to which transparency will be required.

## 2.3  What exactly is the algorithmic part doing?

### 2.3.1  Tier 1 information

*Explain:*
● *How your tool is incorporated into your decision making process*

### 2.3.2  Tier 2 information

*Explain how the tool is integrated into the process, and what influence the tool has on the decision making process. (3.1)*
*Explain how humans have oversight of the tool, including:*
● *How much information the tool provides to the decision maker, and what the information is (3.2)*
● *The decisions humans take in the overall process, including options for humans reviewing the tool (3.3)*
● *Training that people deploying and using the tool must take, if this applies to your project (3.4)*
*Explain your appeal and review process. Describe how you're letting members of the public review or appeal a decision. (3.5)*

The transparency information required around the position of the system within the decision making process centres around the **human involvement** part of the AI accountability debate, whilst also touching upon risk mapping, as already discussed.

Broadly, there are three human-algorithm relationships that a system can adhere to:

- *Human in the loop:* No system decisions or recommendations are acted on in the absence of explicit, pro-active action on the part of a human user;
- *Human on the loop:* All system decisions and recommendations are acted on automatically but are supervised by a human operator who is able to intervene in the process at any time for whatever reason;
- *Fully autonomous:* Decisions and recommendations generated by the system are acted on automatically with no real-time human oversight.

In general, it is understood that, as the system moves from 'human in the loop' to 'fully autonomous', algorithm risk increases as the human component is decreased. However, for 'human in the loop' systems, the efficacy of human control can be negatively impacted by high rates of throughput, whilst the system performance might demand speed of action that makes continuous human intervention impossible (e.g. high frequency trading). Indeed, although we welcome the mapping of human oversight (with a view to identifying points of decision making and levels of autonomy in the system), we have concerns that increased human oversight will become the de facto mechanism of mitigating risk (elsewhere we argue that human oversight can often reintroduce bias [21], whilst [22] discusses the human reasoning's lack of transparency).

## 2.4  Risk assessment

### 2.4.1  Tier 2 information

*Data*
*List and describe:*
- *The datasets you've used to train the model*
- *The datasets the model is or will be deployed on*

*Add links to the datasets if you can*
*Include:*
- *The name of the datasets you used, if applicable (4.1)*
- *An overview of the data used to train and run the tool, including a description of which categories were used for training, testing or operating the model—for example 'age', 'address' and so on (4.2)*
- *The URL for the datasets you've used, if available (4.3)*
- *How and why you collect data, or how and why data was originally collected by someone else (4.4)*
- *The data sharing agreements you have in place (4.5)*
- *Details on who has or will have access to this data and how long the data's stored for, and under what circumstances (4.6)*

*Impact assessments*
*List the impact assessments you've done, for example:*
- *Data Protection Impact Assessment*
- *Algorithmic impact assessment*
- *Ethical assessment*
- *Equality impact assessment*

*For each assessment, add:*
- *The assessment name (5.1)*
- *A description of the impact assessment conducted (5.2)*
- *The date you completed the assessment (5.3)*
- *A link to the assessment or a summary of the assessment, if available (5.4)*

*Risks*
*Provide a detailed description of common risks for your tool, including:*
- *The names of the common risks (5.5)*
- *A description of each identified risk (5.6)*

*For example:*
- *Potential harm from the tool being used in a way it was not meant or built for*
- *Creation of biased results, including through training data that is not representative or contains biases*
- *Arbitrariness and functionality, such as the tool providing unfair or incorrect decisions*

*Mitigations*
*Provide a detailed description of actions you've taken to mitigate the risks in your 'Risks' Sect. (5.7)*

We see the information requirements, requested under tier 2 only, concerning 'Data' and 'Impact assessments' as providing some granularity around the risks identified in the risk mapping. The 'Data' piece highlights any risks to

the rights and freedoms of the end users in regards to the protection of their data, together with surfacing whether sensitive features are used in the building of the algorithmic system or if they lead to any kind of discriminatory behaviour on the part of the system.

Meanwhile, the 'Impact assessments' (for a pertinent example of a equality impact assessment in the context of AI see [23]), 'Risks' and 'Mitigations' pieces surface any concerns thus far found by the system owner, including with respect to privacy, ethics, equality and the algorithm itself, together with any mitigations for such that have been put in place. Here, we also note that model owners might observe and communicate tension between data (privacy) and ethics (fairness) [24, 25], and tension between different concepts of fairness [21, 26–28]. The majority of the information detailed here can be surfaced as part of a third-party technical audit (something we note has been included in recent revisions to the proposed EU AI act [1]).

### 2.5 Ownership

#### 2.5.1 Tier 1 information

---

*Explain:*

● *How people can find out more about the tool or ask a question—including offline options and a contact email address of the responsible organisation, team or contact person*

---

#### 2.5.2 Tier 2 information

---

*List who's accountable for deploying your tool, including:*

● *Your organisation (1.1)*

● *The team responsible for the tool (1.2)*

● *The senior responsible owner (1.3)*

● *Your external supplier or any third parties involved, if the tool has been developed externally (1.4)*

● *The Companies House number of your external supplier (1.5)*

● *The role of the external supplier (1.6)*

● *The terms of their access to any government data (1.7)*

---

For completeness, we also include those parts of the text that provide the audiences for both tiers with the necessary contact and model ownership information in order to pursue any queries as needed. Of particular note for the tier 2 audience, the information requirement includes details of the terms of any third-party access to government data, which we view as being a necessary transparency provision needed to encourage trust in the use of government (or state) data.

## 3 Commentary

In this section we offer focused commentary in a thematic fashion.

### 3.1 Signal

The Algorithmic Transparency Standard is part of the National Data Strategy [6] and follows on from the National AI Strategy [7], both of which commit to greater algorithmic transparency in the public sector. Indeed, the National AI Strategy includes transparency as a feature of AI requiring 'a unique policy response' [29] and details a National AI Research and Innovation Programme to support, inter alia, development of new capabilities around 'transparency of AI technologies' [29].

In releasing the Standard, the UK Government has continued to reiterate this commitment which, together with its expediency (coming only two months after the National AI Strategy), communicates the importance that the government

attaches to this part of the AI framework. However, as we note below, in order for this to move beyond a positive signal, we call for the piloting to be accelerated and for there to be voluntary participation by industry i.e. that industry players can also submit.

### 3.2 Method

We support the approach taken by the UK Government with the Algorithmic Transparency Standard insofar as it can be read as implicitly stimulating the wider industry standards debate and agenda. Three particular points:

- The Algorithmic Transparency Standard has been designed as a tool for UK public sector organisations. Although the Standard isn't explicitly aimed at the private sector, it can be understood that implementation within the UK's sizable public sector will inform standards more widely and influence best practice in the private sector with the adoption of an identical or near-identical framework for algorithmic transparency. We see this approach as an example of 'public sector as an exemplar' [9–11] in line with the National AI Strategy [7].
- The Algorithmic Transparency Standard is being subjected to a pilot allowing scope for iteration and evolution. This approach benefits from practitioner-led, empirically-informed feedback to be fed back into the Standard before finalisation, which we anticipate will both lead to the Standard being more practicable and facilitate better long-term uptake of the Standard.
- We see the pilot approach as contrasting with that being taken by both the EU [1] and the US [12] who are taking a legislator-led approach, but sharing some similarities to China, who are using 'AI pilot zones' to innovate, inter alia., practical and effective policy tools [30], and Canada, who are encouraging user feedback for their 'Algorithmic Impact Assessment Tool' [19], both using a practitioner-led approach.

### 3.3 Public transparency

We view the tiered approach as providing a means of communicating appropriate algorithmic transparency information to distinct audiences. In particular, tier 1 information delivers insight to a wide non-technical audience, including the public, that is both accessible and concise. We see such accessibility as being a prerequisite for widely-held trust in the algorithm on the part of the public.

The tier 2 information requirements ensure that more technical audiences are satisfied as to the purpose, design, use and risk of an algorithmic tool. Pertinently, the tier 2 information provides such technical audiences with a level of detail that encourages the model owner to ensure that best practices are followed, using a 'processes' approach [31].

We presently see this audience tiering methodology as being unique in the context of AI standards-setting and, in particular, with respect to algorithmic registers [18, 32].

### 3.4 Data protection

The Algorithm Transparency Standard does not make or encourage any assumptions as to the data protection framework under which the algorithm owner operates. In particular, we understand this as allowing the Standard to be immediately implementable whilst UK data standards continue to be a subject of regulatory debate [33].

## 4 Conclusion

We conclude by commenting both on the *transparency first approach* to algorithmic accountability and on the initiative to pilot the transparency standard, with such *pilot being made public*. We do this by emphasizing the importance of empirically informed standards, evidenced through case-study and practitioner engagement.

- **Transparency first:** The Transparency Standard requires algorithm owners to publish, to the Algorithmic Transparency Standard collection, details concerning algorithm purpose, ownership, design, and use, together with details over data use and risk assessment and mitigation. We view the exposure of the algorithm owner's diligence in this way as creating a pressure for that

owner to exhibit that best practice has been followed. In particular, the owner is required to show that a robust process has been undertaken to ensure that risks and concerns around the algorithm have been well considered and investigated with efforts at mitigation where necessary. Moreover, we envisage that some element of competitiveness between contributors to the collection might lead to a beneficial ratchet effect for best practice.

- **Pilot and Empiricism:** Developing standards through the use of pilots to empirically inform practices has already been noted as a priority to ensure appropriate AI governance [3]. We especially note that standards that have been tested and evolved through implementation by a range of practitioners prove to be more practicable, itself leading to better adherence to those standards once finalised. Moreover, such practicality facilitates increased interpretability and standardisation with respect to model owner's reporting to the Algorithmic Transparency Standard collection. In making the transparency pilot public, all interested parties are able to gain insight as to the evolution of the pilot, facilitating feedback both to contributors and UK governmental and regulatory bodies. Moreover, providing transparency to the standards-making process itself also engenders public trust in algorithms.

In closing, we note that the template does not deal with how citizens will properly be able to understand the algorithm and its impact on them. There is a digital divide here which has not been tackled. Although we do see the template improving transparency, it has yet to fully deal with some of the complexity involved in doing so.

**Data availability**  Not applicable.

**Code availability**  Not applicable.

**Declarations**

**Competing interests**  The authors declare no competing interests.

# References

1. European Commission. Proposal for a regulation of the European Parliament and of the council: Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. 2021. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206. Accessed 07 Dec 2021.
2. Central Digital and Data Office. Algorithmic transparency standard. 2021. https://www.gov.uk/government/collections/algorithmic-transparency-standard. Accessed 07 Dec 2021.
3. Polle R, Kazim E, Carvalho G, Koshiyama A, Inness C, Knight A, Gorski C, Barber D, Lomas E, Yilmaz E, Thompson G. Towards AI standards: thought-leadership in AI legal, ethical and safety specifications through experimentation. SSRN Electron J. 2021. https://doi.org/10.2139/ssrn.3935987.
4. Bandy J. Problematic machine behavior: a systematic literature review of algorithm audits. Proc ACM Hum-Comput Interact. 2021; 5: 1–34. Doi: https://doi.org/10.1145/3449148.
5. Pagallo U, Aurucci P, Casanovas P, Chatila R, Chazerand P, Dignum V, et al. AI4People—on good AI governance: 14 priority actions, a S.M.A.R.T. model of governance, and a regulatory toolbox/ Social Science Research Network. 2019. https://papers.ssrn.com/abstract=3486508
6. Department for Digital, Culture, Media & Sport. UK national data strategy. 2021. https://www.gov.uk/government/publications/uk-national-data-strategy. Accessed 07 Dec 2021.
7. Office for Artificial Intelligence. National AI strategy. 2021. https://www.gov.uk/government/publications/national-ai-strategy. Accessed 07 Dec 2021.

8.  Open Data Institute. Getting data right: perspectives on the UK National Data Strategy 2020. 2020. http://theodi.org/wp-content/uploads/2021/01/Getting-data-right_-perspectives-on-the-UK-National-Data-Strategy-2020-1.pdf. Accessed 06 Jan 2022.

9.  Kazim E, Almeida D, Kingsman N, Kerrigan C, Koshiyama A, Lomas E, Hilliard A. Innovation and opportunity: review of the UK's National AI Strategy. Discov Artif Intell. 2021;1:14. https://doi.org/10.1007/s44163-021-00014-0.

10.  Zapisetskaya B, De Silva S. National AI strategy: step change for the AI economy in the UK. 2021. https://www.cms-lawnow.com/ealerts/2021/10/national-ai-strategy-step-change-for-the-ai-economy-in-the-uk. Accessed 24 Jan 2022.

11.  Yaros O, Prinsley M A, Vanryckeghem V, Randall R, Hajda O, and Hepworth E. UK Government publishes National Artificial Intelligence Strategy. 2021. https://www.mayerbrown.com/en/perspectives-events/publications/2021/10/uk-government-publishes-national-artificial-intelligence-strategy. Accessed 24 Jan 2022.

12.  Markey E. J. Algorithmic justice and online platform transparency act. 2021. https://www.congress.gov/bill/117th-congress/senate-bill/1896/text. Accessed 06 Jan 2022.

13.  Brown S. Data accountability and transparency act. 2020. https://www.banking.senate.gov/imo/media/doc/Brown%20-%20DATA%202020%20Discussion%20Draft.pdf. Accessed 06 Jan 2022

14.  Cyberspace Administration of China. Provisions on the administration of algorithm recommendations for internet information services. 2022. http://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm. Accessed 24 Jan 2022.

15.  China Briefing News. China passes sweeping recommendation algorithm regulations. 2022. https://www.china-briefing.com/news/china-passes-sweeping-recommendation-algorithm-regulations/ Accessed 24 Jan 2022

16.  Cox J, Lewih A, and Halforty I. AI, machine learning & big data laws and regulations. 2021. https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/australia. Accessed 24 Jan 2022.

17.  Department of Industry, Science, Energy and Resources. Australia's AI ethics principles. 2021. https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles. Accessed 24 Jan 2022.

18.  Haataja M, van de Fliert L, and Rautio P. Public AI registers: realising AI transparency and civic participation in government use of AI. 2020. https://algoritmeregister.amsterdam.nl/wp-content/uploads/White-Paper.pdf. Accessed 06 Jan 2022.

19.  Government of Canada. Algorithmic Impact Assessment Tool. 2021. https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html. Accessed 24 Jan 2022.

20.  Koshiyama A, Kazim E, Treleaven P, Rai P, Szpruch, L, Pavey, G, et al. Towards algorithm auditing: a survey on managing legal, ethical and technological risks of AI, ML and associated algorithms. SSRN Electron J. 2021. Doi: https://doi.org/10.2139/ssrn.3778998

21.  Kazim E, Barnett J, Koshiyama A. Automation and fairness: assessing the automation of fairness in cases of reasonable pluralism and considering the blackbox of human judgment. SSRN Electron J. 2020. https://doi.org/10.2139/ssrn.3698404.

22.  Bernini M. The opacity of fictional minds: Transparency, interpretive cognition and the exceptionality thesis. In: The Cognitive Humanities, Peter Garratt, Ed. Palgrave Macmillan. p. 35–54.

23.  Institute for the Future of Work. Building a systematic framework of accountability for algorithmic decision making. 2021. https://www.ifow.org/publications/policy-briefing-building-a-systematic-framework-of-accountability-for-algorithmic-decision-making. Accessed 07 Dec 2021.

24.  Kazim E, Koshiyama A. The interrelation between data and AI ethics in the context of impact assessments. AI Ethics. 2021;1:219–25. https://doi.org/10.1007/s43681-020-00029-w.

25.  Agarwal S. Trade-Offs between fairness and privacy in machine learning. IJCAI 2021 Workshop on AI for Social Good. 2021. https://crcs.seas.harvard.edu/publications/trade-offs-between-fairness-and-privacy-machine-learning. Accessed 24 Jan 2022.

26.  Kleinberg J, Mullainathan S, Raghavan M. Inherent trade-offs in the fair determination of risk scores. 2016. arXiv:1609.05807.

27.  Chouldechova A. Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. Big Data. 2017;5:153–63. https://doi.org/10.1089/big.2016.0047.

28.  Berk R, Heidari H, Jabbari S, Kearns M, Roth A. Fairness in criminal justice risk assessments: the state of the art. Sociol Methods Res. 2021;50:3–44. https://doi.org/10.1177/0049124118782533.

29.  Keeling E. The UK's National AI Strategy: Setting a 10-year agenda to make the UK a "global AI superpower". 2021. https://www.allenovery.com/en-gb/global/blogs/digital-hub/the-uks-national-ai-strategy---setting-a-10-year-agenda-to-make-the-uk-a-global-ai-superpower. Accessed 06 Jan 2022.

30.  The State Council of The People's Republic of China. China promotes local AI pilot zones. http://english.www.gov.cn/statecouncil/ministries/202005/09/content_WS5eb66f29c6d0b3f0e9497457.html Accessed 24 Jan 2022.

31.  Kazim E, Koshiyama AS. A high-level overview of AI ethics. Patterns. 2021;2: 100314. https://doi.org/10.1016/j.patter.2021.100314.

32.  The Law Society. Algorithm use in the criminal justice system report. 2019. https://www.lawsociety.org.uk/topics/research/algorithm-use-in-the-criminal-justice-system-report. Accessed 06 Jan 2022

33.  Almeida D, Shmarko K, Lomas E. The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of US, EU, and UK regulatory frameworks. AI Ethics. 2021. https://doi.org/10.1007/s43681-021-00077-w.