

Automation Bias and Procedural Fairness

A short guide for the public sector

John Zerilli, Iñaki Goñi, Matilde Masetti Placci



This fellowship ran from January-May 2023 as part of BRAID.

BRAID is a UK-wide programme dedicated to integrating Arts and Humanities research more fully into the Responsible AI ecosystem, as well as bridging the divides between academic, industry, policy and regulatory work on responsible AI. Funded by the Arts and Humanities Research Council (AHRC), BRAID represents AHRC's major investment in enabling responsible AI in the UK. The Programme runs from 2022 to 2028. Working in partnership with the Ada Lovelace Institute and BBC, BRAID supports a network of interdisciplinary researchers and partnering organisations through the delivery of funding calls, community building events, and a series of programmed activities. Funding reference: Arts and Humanities Research Council grant number AH/X007146/1.

Learn more at www.braiduk.org

This research was supported via UK Research and Innovation by the R&D Science and Analysis Programme at the Department for Culture, Media & Sport. Any primary research, subsequent findings or recommendations do not represent Government views or policy and are produced according to research ethics, quality assurance, and academic independence.

To request an alternative format of this report please email braid@ed.ac.uk.

Table of Contents

<i>Preface</i>	3
<i>Key Points</i>	3
<i>Executive Summary</i>	4
<i>Background</i>	4
<i>Research Questions</i>	5
<i>Methodology: Human Factors</i>	5
<i>Methodology: Law</i>	8
<i>Findings: Human Factors</i>	9
<i>Findings: Law</i>	11
<i>References</i>	14
<i>Appendix</i>	20



● ● ● Preface

This research was supported via UK Research and Innovation (UKRI) by the (former) Department for Digital, Culture, Media and Sport (DCMS) Science and Analysis R&D Programme. It was developed and produced according to UKRI's initial hypotheses and output requests. Any primary research, subsequent findings or recommendations do not represent DCMS views or policy and are produced according to academic ethics, quality assurance, and independence.

The DCMS Science and Analysis R&D Programme funds researchers to provide critical evidence for the department's policy areas. Led by the Chief Scientific Adviser and Director of Analysis, a key objective is to make a step change in how the department develops its evidence base; improving its policymaking and facilitating knowledge exchange between government and independent experts. In 2022/23, DCMS partnered with UKRI research councils (the Engineering and Physical Sciences Research Council and the Arts and Humanities Research Council) to deliver research across four topic areas.

● ● ● Key Points

- Automation bias is an attitude of blind trust in automation which users of sophisticated technology can often fall into.
- Automation bias not only has tangible consequences for the people affected by poor decision-making, it potentially has legal consequences for the government department relying on automation.
- If public sector officials succumb to automation bias they are in danger of unlawfully fettering or improperly delegating their discretion.
- The easiest way to mitigate the risks of automation bias and its legal consequences is for those responsible for procurement decisions to adhere to a simple checklist that ensures that the pitfalls of automation are avoided as much as possible.

● ● ● Executive Summary

The use of advanced artificial intelligence (AI) and data-driven automation in the public sector poses several organisational, practical, and ethical challenges. One that is easy to underestimate is automation bias, which, in turn, has underappreciated legal consequences. Automation bias is an attitude in which the operator of an autonomous system will defer to its outputs to the point where they overlook or ignore evidence that the system is failing. The legal problem arises when statutory office-holders (or their employees) either fetter their discretion to in-house algorithms or improperly delegate their discretion to third-party software developers – something automation bias may facilitate. A synthesis of previous research suggests an easy way to mitigate the risks of automation bias and its potential legal ramifications is for those responsible for procurement decisions to adhere to a simple checklist that ensures that the pitfalls of automation are avoided as much as possible.

● ● ● Background

The danger of human operators overtrusting technical systems has been recognised for many years (explored within the field of ‘human factors’ engineering and human-computer interaction). One well-established result of research over the past four decades is that as the quality of automation improves, and the human operator’s role becomes progressively less demanding, the operator succumbs to ‘automation bias’. This is an attitude in which an operator will defer to a system’s outputs to the point where the operator overlooks or ignores evidence that the system is failing. Decades of research confirm that this and related problems (such as diminished attention or ‘situation awareness’ and the gradual loss of legacy and manual control skills) are both insidious and potentially intractable. These problems seem to worsen as automation improves and, somewhat alarmingly, afflict experts as much as novices – despite often rigorous training regimes put in place to prevent them.

Automation bias in the public sector has underappreciated legal consequences.

Administrative law principles in the UK jealously guard the repository of statutory

discretion. A power conferred upon a minister of the Crown, for example, must only be exercised by the minister (or their departmental employees). The exception to this principle occurs when legislation conferring power to the decision-maker permits delegation to third parties (either expressly or by implication). Likewise, administrative law prohibits an authorised decision-maker from ‘fettering’ their discretion, for instance, by blindly following departmental policy without turning their mind diligently to the decision at hand. Unthinking, uncritical reliance on algorithmic decision tools in the public sector – such as may occur through automation bias – probably amounts to a breach of these principles. Accordingly, increasingly ‘smart’ data-driven and other AI solutions in the public sector need to be vetted for their tendency to induce automation bias.

● ● ● Research Questions

We wanted to get a sense of what the growing literature on the use of AI in public-sector settings was like, what sorts of questions were being answered, whether and how automation bias featured as a concern, and whether any practical recommendations could be formulated on the basis of this literature. We also wanted to take soundings from other jurisdictions on the legal issues posed by automation bias: were improper delegation and related issues being flagged at all, and what were the legal/regulatory responses, if any?

● ● ● Methodology: Human Factors

For the human factors aspect of the investigation, we conducted an ‘umbrella review’ of AI in government using keywords related to AI, government, and systematic reviews. Figure 1 depicts the umbrella review selection process that started with 514 records. These were found using keywords entered in the two most prestigious databases of academic literature, as well as through a flexible search (grey literature). After sequential steps of filtering, we produced a final selection of 35 systematic

reviews. A small selection of non-review articles was also examined. The final list is included in the bibliography.



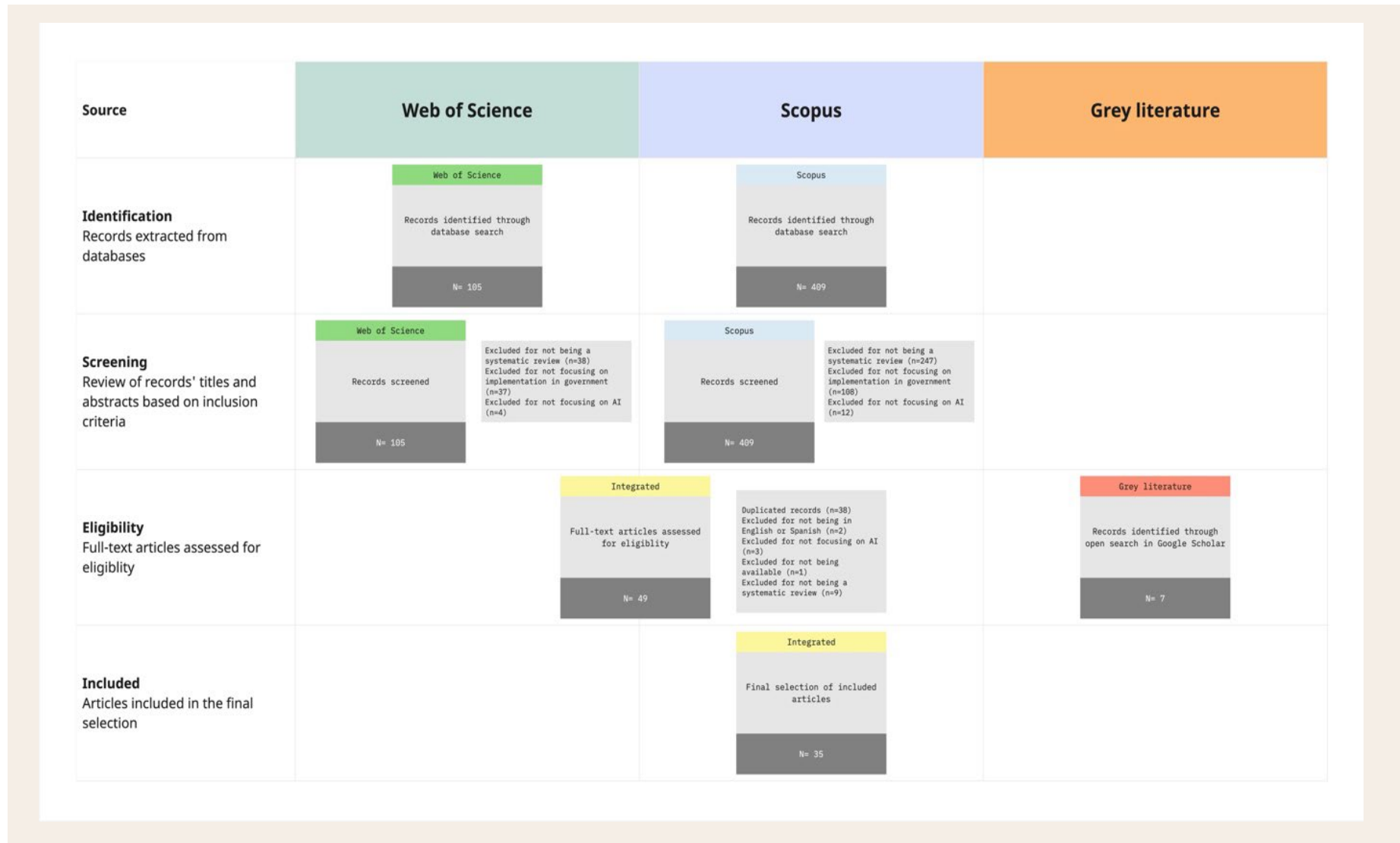


Figure 1. Selection process for the umbrella review

● ● ● Methodology: Law

For the legal aspect of the investigation, we limited ourselves to six countries, and for each country we examined the use of AI in four sectors of public administration. For each of these, in turn, we asked two broad legal questions, as indicated in Table 1.

Countries	Sector	Questions
<ul style="list-style-type: none"> • France • Germany • US • India • Australia • Canada 	<ul style="list-style-type: none"> • Policing • Immigration & Asylum • Social Security/Welfare • Healthcare 	<ul style="list-style-type: none"> • Have due process concerns arisen? • How, if at all, have they been dealt with?

Table 1. Countries, sectors, and questions examined

The countries selected span both civil and common law jurisdictions, as well as one (India) which is in the unique position of undergoing digitalisation against the background of a relatively old bureaucratic regime.

The four sectors were chosen for three reasons. First, they are sectors in which the government is most likely to be responsible for a large number of individuals, thereby incentivising the introduction of data-driven AI. Second, some of these sectors were among the first to experience digitalisation, while others (such as healthcare) were late in coming. This chronological variety provides a potentially wider sampling of ways in which AI can be incorporated into government services. Sectors with longer memories of digitalisation would presumably have had much more of their automatable functions automated by now, whereas sectors relatively new to digitalisation are likely to have AI and data-driven innovations running atop of legacy systems (including manual procedures). This variety introduces interoperability and compatibility factors that may be important in getting a sense of the relevant legal issues. Third, the four sectors

are highly sensitive political areas where the fettering of ministerial discretion would carry especially significant weight in the public mind.

● ● ● Findings: Human Factors

Surprisingly, not much in the literature on the automation of public administration was concerned with automation bias. We did, however, note a few trends:

- There appears to be some evidence of confirmation bias among some bureaucrats, indicating selective adherence to algorithmic recommendations when the algorithm confirms the user's suspicions (we call this the 'just like I thought' phenomenon).
- In some cases, bureaucrats' misgivings about an algorithm don't translate to non-use. In other cases, bureaucrats' trust in an algorithm doesn't translate to using it. This highlights a difference between behaviour and attitudes towards algorithms: while it's often assumed that when a user declines to follow an algorithm, it is because they don't trust it (and vice versa), this need not be the case.
- To reduce overreliance on AI systems, cognitive forcing has been proposed as an alternative to simple 'explainable AI' approaches. With this approach, the design of systems actively induces users to pay attention and remain alert. However, there may be a trade-off between reducing overreliance and subjective ratings of the design. In other words, the more the systems force attention by public servants, the more likely they will be dissatisfied with it.
- There is emerging evidence from simulated trials that warnings to judges about the relatively low accuracy of criminal risk assessment tools may be effective in counteracting some of the effects of automation bias.
- Incentive structures, including a variety of accountability mechanisms, also seem to counteract some of the effects of automation bias. Making decision-makers accountable for their decisions may then elicit a more discriminating

use of algorithms insofar as decision-makers will be less likely to take algorithmic outputs at face value when they 'have skin in the game'.

A synthesis of more general reviews on the human factors of AI (i.e. extending beyond the public sector) suggests that human performance can be enhanced when automation augments rather than replaces human skills. But this need not always be the case. When a system performs better than a human or human-AI team on key metrics, these systems might be able to replace the human user/s altogether and be left to operate autonomously. In general, human-AI teams should be set up so as to allow human users to focus on those aspects of a task better suited to human rather than autonomous execution – matters involving empathy, abstraction, conceptualisation, and the exercise of judgment, as opposed to those requiring calculation, analysis, speed, and iterative processing at scale.

At the same time, when human and machine performance in some task is comparable, the allocation of functions should be flexible enough to support dynamic interaction, with hand-over and hand-back (as occurs when a driver disengages cruise control and thereby resumes manual control of acceleration). This can assist in keeping users 'on their toes'. It's important to stress, however, that dynamism may be inappropriate (and even dangerous) when there is hand-over between agents that are ill-matched in their competencies (e.g. where the human is far inferior, or superior, to an autonomous system). Dynamism should therefore only be built into a workflow when the human and machine are nearly equivalently proficient at the task in question. When there is a clear difference in competencies, the division of labour should be stark. Human operators should be effectively barred from interfering with the machine's outputs, since human interference with a system that performs substantially better than humans can degrade performance. Breaking the task up in this way also increases the chances of finding optimally reliable software to handle the automated parts of a task/decision.

By far the prevailing themes of the public sector literature coalesced around two sets of challenges unrelated to human factors, one practical, the other ethical. The practical

challenges included issues around harmonisation of new technologies with existing work systems, digital literacy and skills shortages, budgetary constraints affecting procurement choices, and general distrust of new technologies. Ethical challenges related to the social legitimacy and overall beneficence of AI (implicating concerns around fairness, transparency, equity, safety, data protection, non-discrimination, surveillance, accountability, respect for autonomy and dignity, job displacement, and citizen engagement). The *Appendix* provides representations of the benefits, ethical challenges, and practical challenges of AI implementation in government found through our analysis of systematic reviews.

● ● ● Findings: Law

Overall, France has the strongest data privacy laws and has been the slowest country to integrate AI within any of the four sectors we examined. The United States and Canada have comparable levels of AI uptake in government, though Canada places more onerous obligations on ministers. The UK, Australia, and Germany fall somewhere in between France and the US/Canada as far as uptake is concerned. In none of the jurisdictions had procedural fairness/due process questions arising from automation bias been a major concern. However, some of these had been very briefly flagged in some grey literature (particularly in Australia). Furthermore, we expect litigation around the legality of public sector algorithms to increase in coming years. Australia's experience with what has come to be known there as 'Robodebt' (a debt assessment algorithm using a spurious method of 'income-averaging' to calculate sums allegedly owed to the government by welfare recipients) now serves as the clearest international example of how *not* to automate public functions. Robodebt was found illegal by the courts in 2021 and resulted in the Australian government paying compensation to many victims of the scheme. It was the subject of a Royal Commission (which concluded on 18 April 2023). The now-infamous Post Office-Horizon scandal is the UK's own version of automated administration gone horribly wrong.

Recommendations: An AI uptake 'checklist'

The following six principles can serve as a framework both for assessing the viability of any human–AI team as well as guiding their design and implementation.

- **Division of labour.** Tasks with automatable subtasks should reflect a clear allocation of responsibilities between the human- and computer-operated parts of the task.
- **Complementarity.** The allocation of responsibilities should proceed in such a way that those subtasks better suited for human handling are not automated, and those better suited for computer handling are not manually controlled. Human operators should be effectively barred from interfering with the machine's outputs.
- **Dynamism.** The allocation of responsibilities should incorporate hand-over and hand-back protocols where this flexibility contributes to optimal performance. This assumes that some tasks or subtasks can be safely handled either by humans or computers insofar as humans and computers have shared competencies with respect to the task/subtask.
- **Co-evolution.** User requirements co-evolve over time, and decision-support tools should reflect this. Human–AI teams should be designed with adaptability and change in mind. This means designers should not over-specify how an automated system will work, but allow its users to tailor the system so that it best meets their particular needs.
- **Pragmatism.** Automated systems should yield to existing practices within an organisation, even if they appear archaic. (When mobile phones first appeared, people didn't automatically dispense with their hardcopy telephone and address books. Manual and analogue systems held on for a while longer until the arrival of smart phones put the internet in people's pockets.)

- Context-sensitivity. Each automated system, situated within its own unique decision context, may prioritise these principles and negotiate their various trade-offs differently.

The last point is particularly important. It is important to appreciate that the above is a high-level framework that is applicable across domains (e.g. sentencing, welfare administration, logistics, etc.) as well as tasks (e.g. the sentencing task may have more in common with some types of welfare determination, from the point of view of human factors, even though sentencing and welfare administration are very different domains). Accordingly, adaptation to context – especially task context – is paramount.


● ● ● References

1. Administrative Review Council. (2004). *Automated assistance in administrative decision making*. Barton, ACT: Commonwealth of Australia.
2. Agostino, D., Saliterer, I., & Steccolini, I. (2022). Digitalization, accounting and accountability: A literature review and reflections on future research in public services. *Financial Accountability & Management* 38: 152–176.
3. Almeida, P. G. R. de, dos Santos, C. D., & Farias, J. S. (2021). Artificial intelligence regulation: A framework for governance. *Ethics and Information Technology* 23: 505–525.
4. Almutawa, M., & Rashid, H. (2020). Comprehensive review on the challenges that impact artificial intelligence applications in the public sector. *Proceedings of the 5th NA International Conference on Industrial Engineering and Operations Management*.
5. Bundesministerium für Wirtschaft und Klimaschutz. (2020). *Strategie Künstliche Intelligenz der Bundesregierung*. Available at: <https://www.bmwk.de/Redaktion/DE/Publikationen/Technologie/strategie-kuenstliche-intelligenz-fortschreibung-2020.pdf%20for%20updated%20post-COVID-19>, accessed 20.05.2024.
6. Canadian Association of Radiologists Artificial Intelligence Working Group. (2019). White paper on ethical and legal issues related to artificial intelligence in radiology. *Canadian Association of Radiologists Journal* 70: 107.
7. Citron, D. K. (2008). Technological due process. *Washington University Law Review* 85: 1249–1313.
8. Damij, N., & Bhattacharya, S. (2022). The role of AI chatbots in mental health related public services in a (post)pandemic world: A review and future research agenda. *2022 IEEE Technology and Engineering Management Conference (TEMSCON EUROPE)*: 152–159.
9. Dreyling, R. M., Tammet, T., & Pappel, I. (2022). Artificial intelligence use in e-government services: A systematic interdisciplinary literature review. In: *Future data and security engineering*, eds. T. K. Dang, J. Küng, & T. M. Chung, 547–559. Springer.

10. El Asri, H., & Benhlima, L. (2020, December). Artificial intelligence based knowledge management in public administration: A mapping study. *Proceedings of 21st European Conference on Knowledge Management*.
11. European Migration Network. (2020). The use of digitalisation and artificial intelligence in migration management. EMN–OECD Inform.
12. European Parliament. (2020). The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. Study by the Panel for the Future of Science and Technology. Available at: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf), accessed 20.05.2024.
13. Freeman Engstrom, D., & Ho, D. E. (2021). Artificially intelligent government: A review and agenda. In: *Research handbook on big data law*, ed. R. Vogl, 57–86. Edward Elgar.
14. Haug, K. B. (2022). Structuring the scattered literature on algorithmic profiling in the case of unemployment through a systematic literature review. *International Journal of Sociology and Social Policy*.
15. Hildreth, R. (2019). Birmingham Police break down newly released crime numbers. *WBRC Fox 6 News* (13 November 2019).
16. Hoekstra, M., van Veenstra, A. F., & Chideock, C. (2021). A typology for applications of public sector AI. *EGOV-CeDEM-EP*. 121–128.
17. Kalampokis, E., Karacapilidis, N., Tsakalidis, D., & Tarabanis, K. (2022). Artificial intelligence and blockchain technologies in the public sector: A research projects perspective. In: *Electronic government*, eds. M. Janssen, C. Csáki, I. Lindgren, E. Loukis, U. Melin, G. Viale Pereira, M. P. Rodríguez Bolívar, & E. Tambouris. *EGOV2022*: 323–335.
18. Koulis, R. & Evans, K. (2021). Pushing with impunity: The legacy of risk classification assessment in immigration detention. *Georgetown Immigration Law Journal* 36(1).
19. Leão, H. A. T., Canedo, E. D., & Souza, J. C. F. (2018). Digitization of government services: Digitization process mapping. In: *On the move to meaningful internet systems*, eds. H. Panetto, C. Debruyne, H. A. Proper, C. A. Ardagna, D. Roman, & R. Meersman. *OTM2018*: 3–20.

20. Lyra, M. S., Damásio, B., Pinheiro, F. L., & Bacao, F. (2022). Fraud, corruption, and collusion in public procurement activities, a systematic literature review on data-driven methods. *Applied Network Science* 7: 83.
21. Madan, R., & Ashok, M. (2023). AI adoption and diffusion in public administration: A systematic literature review and future research agenda. *Government Information Quarterly* 40: 101774.
22. Mohamad, I., Hughes, L., Dwivedi, Y. K., & Alalwan, A. A. (2022). AI technologies for delivering government services to citizens: Benefits and challenges. In: *The role of digital technologies in shaping the post-pandemic world*, eds. I. O. Papagiannidis, S. Alamanos, E. Gupta, S. Dwivedi, Y.K. Mäntymäki, & M. Pappas, 38–57. Springer.
23. New South Wales Parliamentary Research Service. (2020). The use of artificial intelligence by government: Parliamentary and legal issues. *E-brief* (September 2020).
24. O'Donovan, D. (2023). 'Amateurish, rushed and disastrous': Royal commission exposes robodebt as ethically indefensible policy targeting vulnerable people. *The Conversation* (10 March 2023).
25. Office of the Privacy Commissioner of Canada. (2021). Joint investigation of Clearview AI, Inc. by the Office of the Privacy Commissioner of Canada, the Commission d'accès à l'information du Québec, the Information and Privacy Commissioner for British Columbia, and the Information Privacy Commissioner of Alberta. Available at: <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001/#toc1>, accessed 20.05.2024.
26. Ojo, A., Mellouli, S., & Ahmadi Zeleti, F. (2019). A realist perspective on AI-era public management. *Proceedings of the 20th Annual International Conference on Digital Government Research*. 159–170.
27. Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: Framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A* 376: 1–20.

28. Pechtor, V., & Basl, J. (2022). Analysis of suitable frameworks for artificial intelligence adoption in the public sector. In: *Digitalization of society, business and management in a pandemic: 30th Interdisciplinary Information Management Talks*, eds. C. Gerhard, D. Petr, & O. Václav, *IDIMT-2022*.
 29. Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2020). Big data and AI – A transformational shift for government: So, what next for research? *Public Policy and Administration* 35: 24–44.
 30. Pi, Y. (2021). Machine learning in governments: Benefits, challenges and future directions. *JeDEM EJournal of EDemocracy and Open Government* 13: 203–219.
 31. Plantinga, P. (2022). Digital discretion and public administration in Africa: Implications for the use of artificial intelligence. *Information Development* 026666692211175.
 32. Reis, J., Santo, P. E., & Melão, N. (2019). Artificial intelligence in government services: A systematic literature review. In: *New knowledge in information systems and technologies*, eds. Á. Rocha, H. Adeli, L. P. Reis, & S. Costanzo. *WorldCIST'192019*: 241–252.
 33. Reis, J., Santo, P. E., & Melao, N. (2019). Impacts of artificial intelligence on public administration: A systematic literature review. *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)* 1–7.
 34. Reis, J., Santo, P. E., & Melão, N. (2021). Influence of artificial intelligence on public employment and its impact on politics: A systematic literature review. *Brazilian Journal of Operations & Production Management* 18: 1–22.
 35. Sánchez-Céspedes, J.-M., Rodríguez-Miranda, J.-P., & Salcedo-Parra, O.-J. (2022). Aplicación de la inteligencia artificial en la formulación de políticas públicas relacionadas con la vocación agrícola de las regiones. *Revista Científica* 44: 172–187.
 36. Sarker, M. N. I., Wu, M., & Hossin, M. A. (2018). Smart governance through bigdata: Digital transformation of public agencies. *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)* 62–70.
 37. Savaget, P., Chiarini, T., & Evans, S. (2019). Empowering political participation through artificial intelligence. *Science and Public Policy* 46: 369–380.
- 

-
38. Sharma, G. D., Yadav, A., & Chopra, R. (2020). Artificial intelligence and effective governance: A review, critique and research agenda. *Sustainable Futures* 2: 100004.
39. Soomro, K., Bhutta, M. N. M., Khan, Z., & Tahir, M. A. (2019). Smart city big data analytics: An advanced review. *WIREs Data Mining and Knowledge Discovery* 9(5).
40. Sousa, W. G. de, Melo, E. R. P. de, Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly* 36: 101392.
41. Ubaldi, B., Le Fevre, E. M., Petrucci, E., Marchionni, P., Biancalana, C., Hiltunen, N., Intravaia, D. M., & Yang, C. (2019). State of the art in the use of emerging technologies in the public sector. In *OECD Working Papers on Public Governance* (Issue 31). OECD Publishing.
42. Valle-Cruz, D., & Gil-García, J. R. (2022). Tecnologías emergentes en gobiernos locales: Una revisión sistemática de literatura con la metodología PRISMA. *Revista Mexicana de Análisis Político y Administración Pública* 11: 9–28.
43. Valle-Cruz, D., Ruvalcaba-Gomez, E. A., Sandoval-Almazan, R., & Ignacio Criado, J. (2019). A review of artificial intelligence in government and its potential from a public policy perspective. *Proceedings of the 20th Annual International Conference on Digital Government Research*. 91–99.
44. van Noordt, C., Medaglia, R., & Misuraca, G. (2020). Stimulating the uptake of AI in public administrations: Overview and comparison of AI strategies of European member states. *Proceedings of Ongoing Research, Practitioners, Workshops, Posters, and Projects of the International Conference EGOV-CeDEM-E2020*: 269–277.
45. Whiteford, P. (2023). Why robodebt's use of 'income averaging' lacked basic common sense. *The Conversation* (16 March 2023).
46. Wirtz, B. W., Langer, P. F., & Fenner, C. (2021). Artificial intelligence in the public sector: A research agenda. *International Journal of Public Administration* 44: 1103–1128.
- 

-
47. Yusriadi, Y., Rusnaedi, R., Siregar, N. A., Megawati, S., & Sakkir, G. (2023). Implementation of artificial intelligence in Indonesia. *International Journal of Data and Network Science* 7: 283–294.
 48. Zerilli, J. (2020). Algorithmic sentencing: Drawing lessons from Human Factors research. In: *Sentencing and artificial intelligence*, ed. J. Ryberg & J. Roberts, 165–183. Oxford University Press.
 49. Zerilli, J., Bhatt, U., & Weller, A. (2022). How transparency modulates trust in artificial intelligence. *Patterns* 3: 1–10.
 50. Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Algorithmic decision-making and the control problem. *Minds and Machines* 29: 555–578.
 51. Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly* 38: 101577.
- 

All concepts found in ethical challenges:



All concepts in ethical challenges with more than one mention:





This fellowship ran from January-May 2023 as part of BRAID.

BRAID is a UK-wide programme dedicated to integrating Arts and Humanities research more fully into the Responsible AI ecosystem, as well as bridging the divides between academic, industry, policy and regulatory work on responsible AI. Funded by the Arts and Humanities Research Council (AHRC), BRAID represents AHRC's major investment in enabling responsible AI in the UK. The Programme runs from 2022 to 2028. Working in partnership with the Ada Lovelace Institute and BBC, BRAID supports a network of interdisciplinary researchers and partnering organisations through the delivery of funding calls, community building events, and a series of programmed activities. Funding reference: Arts and Humanities Research Council grant number AH/X007146/1.

Learn more at www.braiduk.org

This research was supported via UK Research and Innovation by the R&D Science and Analysis Programme at the Department for Culture, Media & Sport. Any primary research, subsequent findings or recommendations do not represent Government views or policy and are produced according to research ethics, quality assurance, and academic independence.

To request an alternative format of this report please email braid@ed.ac.uk.

