The right to human decision: analyzing policies, ethics, and implementation

Written by Michael Cheng[1]

[1]Institute for Ethics in AI, University of Oxford
Sherrington Rd, Oxford, UK OX1 3QU
michael.cheng@magd.ox.ac.uk

Abstract

As automated decision-making advances, many governments have adopted AI ethics guidelines that endorse a "right to human decision" mandating that important decisions made by AI algorithms (i.e. housing loan eligibility, judicial rulings) are subject to review by human decision-makers with final decision-making authority. The African Union, Canada, China, European Union, India, United Arab Emirates, United States, Uruguay, and political entities with jurisdiction over most of the world's population have all endorsed a right to human decision, even though their proposals have received relatively little public attention. There has yet to be a systematic analysis of proposals for a right to human decision. This paper examines the major proposed forms of the right to human decision: a right to contest decisions made by AI to a human reviewer, the right to opt out from AI decision-making in favor of a human decision-maker, and the right to exclude machines entirely from particular kinds of decisions. I examine key arguments for and against a right to human decision. I argue that several practical challenges must be addressed before this proposed right is implemented. For instance, the distinction between a "human" and "AI" decision is not always clear and the advent of hybrid human-AI decision makers will only blur that line.

1 Introduction

As AI becomes increasingly embedded in decision-making throughout society, many policymakers are calling for a right to human decision, where individuals can withdraw from or appeal AI decision-making to be evaluated by a human decision-maker. Nations and international organizations with jurisdiction over most of the world's population have endorsed some form of the right to human decision in their AI ethics policy-making, even though the right to human decision has received relatively minimal public attention. In 2014, the African Union Convention on Cyber Security and Personal Data Protection declared that humans have a right to not be subject to decisions that are "based solely on automated processing of data intended to evaluate certain personal aspects" (African Union 2014; Shany 2023). In 2021, the Chinese Ministry of Science and Technology released national AI guidelines that call for humans to "have fully autonomous decision-making rights and that they have the right to accept or reject AI-provided services, the right to withdraw from AI interactions at any time, and the right to terminate AI system operations at any time" (Center for Security and Emerging Technology 2021; Chinese Ministry of Science and Technology 2021; Houweling 2021). In 2022, the US White House released a non-binding "Blueprint for an AI Bill of Rights" that called for a right to opt out from automated decision-making and "have access to a person who can quickly consider and remedy problems you encounter" (The White House 2022). Since 2018, Article 22 of the European Union's GDPR has established a right for individuals to "not be subject to a decision based solely on automated processing," although the legal implications of this right are currently being worked out in the European courts (European Union 2016; Binns and Veale 2021; Court of Justice of the European Union 2023).

Although data-driven algorithms have been used to make decisions for decades, AI decision-making represents a new challenge because AI decision-making is not fully explainable and often involves autonomous agents. As AI methods evolve, proposals for human oversight have become more common. Governments have endorsed a right to human decision in part because many forms of a right to human decision can be implemented without relying on hypothetical AI developments, providing practical value that other AI ethics proposals currently do not. For instance, when an AI algorithm for automated credit scoring is released, a right to human decision can be legally implemented by allowing individuals to appeal the AI's decision to a human reviewer, irrespective of how the algorithm was trained. In contrast, some AI ethics proposals such as explainability depend on hypothetical developments that are not found in the most powerful AI systems. For instance, it is currently impossible to precisely explain why ChatGPT releases particular outputs (Bowman 2023; Ray 2023).

Although a right to human decision has been endorsed by many policymakers, it is not yet clear what a right to human decision entails because there are few examples of the right being enforced or specified. Does a right to human decision mean that humans should have the right to contest an AI decision to a human reviewer after the decision is made, the right to opt out of AI decision-making altogether, or something else? Should a right to human decision apply when-

ever AI is used to make any type of decision or for specific, higher risk use cases?

The goal of this paper is to provide an overview of the proposed right to human decision. I describe existing policy proposals from governments and international organizations. I analyze the three major forms that proposals for the right to human decision have taken: the right to contest AI decisions to a human reviewer, the right to opt out from AI decision-making in favor of a human alternative, and the right to entirely exclude machines from certain kinds of decisions and reserve those decisions to humans. I discuss the debate over the right to human decision and provide key arguments that justify and oppose the proposed right. Finally, I argue that AI policymakers and practitioners need to address several practical challenges facing the proposed right to human decision, such as where to draw the line between a "human" and "AI" decision.

2 Existing policy proposals

Nation-states and international organizations with jurisdiction over most of the world's population have already endorsed some form of the right to human decision (Appendix). However, many of their proposals are light on the specifics of how a right to human decision would work, and there is no law that sets out an explicit procedure for a right to human decision and when it would apply. There are a few laws that call for limits on automated decision-making, but they use vague language. The African Union and European Union have passed legally binding rules that call for humans to not be subject to decisions based "solely on automated processing," although they did so through mid-2010s data processing laws that preceded recent AI developments, and it is not clear what legal obligations they create. Some scholars have proposed that these laws imply a right to human decision or explanation (Shany 2023; Goodman and Flaxman 2017; Sarra 2020; Vaccaro et al. 2021).

More recent AI ethics guidelines explicitly demand some form of a right to human decision over automated decision-making. For instance, the United Arab Emirates proposes that humans should be able to "challenge significant automated decisions concerning them and, where appropriate, be able to opt out of such decisions" and Uruguay calls for a "clearly identifiable person" responsible for the actions of an AI and declares that "when the application and / or development of AI-based solutions present ethical dilemmas, they must be addressed and resolved by human beings" (United Arab Emirates Government 2022; Government of Uruguay 2019). Nevertheless, most AI ethics proposals are vague on when human decision-making is actually required, although many imply that there is always a right to human decision whenever AI is employed. However, Singapore's AI ethics guidelines diverge from most other nations in that they state that fully automated decision-making can be ethically acceptable in some cases. Singapore suggests that the probability of AI causing harm and the severity of that potential harm should determine the level of human oversight required for a particular use case, explicitly allowing AI to make entirely autonomous decisions in "low-risk" use cases (Singapore Personal Data Protection Commission 2020).

3 Different formulations of the right to human decision

Although many nations have proposed a "right to human decision," there is no consensus over the details of the right. In this section, I describe the three major formulations of the right to human decision: the right to contest automated decisions to a human authority, the right to opt out of automated decision-making in favor of a human alternative, and the total exclusion of machines from decision-making in particular domains. It is also important to note that the exact form that a right to human decision will inevitably depend on the specific use cases involved. After all, a decision on what to wear involves lower stakes than a life or death decision.

3.1 Right to contest

The right to contest can be described simply: when AI makes a decision, humans have the right to contest that decision to a human reviewer. Decisions cannot be made by an AI algorithm with no potential for human review whatsoever (Shany 2023; Kaminski and Urban 2021).

The White House's 2022 Blueprint for an AI Bill of Rights endorses a right to "have access to timely human consideration and remedy by a fallback and escalation process if an automated system fails, it produces an error, or you would like to appeal or contest its impacts" (The White House 2022). The White House provides several examples: the customer service industry has made AI chatbots the default while maintaining teams of humans who can review more complicated cases and the government has funded "healthcare navigators" to help Americans enroll in health insurance online (The White House 2022). Nevertheless, none of these examples provide clear language for how a right to contest might be standardized or legally formulated in a manner that cannot be easily gamed. After all, is it truly meaningful to be able to appeal the decision of an AI chatbot to a human customer service agent if it requires waiting 5 hours on the phone?

3.2 Right to opt out for a human alternative

The right to opt out proposes that humans should have the right to opt out of automated decision-making and request a fully human decision maker. It allows machines to be used in decision-making if there is also a human alternative.

It is unlikely that a right to opt out could be engineered to require a fully human decision-making option for all AI use cases, since some AI uses might not have a human alternative (i.e. an AI medical treatment). Practically speaking, a right to opt out would likely apply to specific use cases. Nevertheless, in cases where there is a competition for limited resources (e.g. university admissions slots), it may be difficult to maintain fairness if some applicants are selected by human evaluators and others by AI evaluators.

3.3 Total exclusion of machines from human decision domains

Finally, the right to human decision could imply the total exclusion of machines from particular use cases, where only humans are allowed to make decisions. These kinds of decisions may be so salient to human interests that machines

cannot make them at all. Identifying specific use cases where only human decision-making is allowed could potentially allow humans to preserve some agency against the demands of large, technocratic systems. Relatively few governmental AI ethics proposals have proposed the total exclusion of machines from particular domains. However, the United Arab Emirates' AI ethics guidelines declare that "decisions related to lethal force, life and death should not be delegated to AI systems. Rules and standards should be adopted to ensure effective human control over those decisions" (United Arab Emirates Government 2022).

4 Justifying the right to human decision

In its general form, the right to human decision is the proposal that some decisions should be made by humans and not AI. This right could be implemented generally or in specific use cases—for instance, a society may decide that it would be unethical for AI to make employment decisions, and mandate a right to human decision in that use case. Nevertheless, there has been substantial normative disagreement over the proposal for the right to human decision. In this section, I review the major arguments in favor and against a right to human decision.

4.1 Arguments in favor of a right to human decision

It would be unjust to deny humans a right to appeal. When AI algorithms make decisions that impact human lives (such as whether someone qualifies for a home loan, will be hired for a job, is targeted for personalized opportunities based on automated profiling, or receives compensation for a misdeed), it could be a violation of legal due process if there were no right to appeal whatsoever (Bayamlıoğlu 2022; Ploug and Holm 2021). The ability to appeal and assert one's case is a core feature of most legal systems, and it is unfair or illegal for algorithmic decisions to be incontestable (Kaminski and Urban 2021).

Algorithms should be accountable to humans. A right to human decision ensures that algorithms can be held accountable and improved when necessary, especially when an algorithm's results are found to be undesirable. Moreover, as AI algorithms learn how to train on themselves (whether through synthetic data or through a reinforcement learning approach like the process for training AlphaGo, where AI learned by playing games against itself), the amount of human oversight required to train and release a particular algorithm might decrease, so implementing the right to human decision can promote human agency and accountability over AI (Silver et al. 2016; Mishra et al. 2022).

AI algorithms will inevitably get it wrong. Modern AI algorithms are built on mathematical methods such as maximum likelihood estimation and empirical risk minimization that attempt to generalize and search for averages across large datasets, but there will inevitably be exceptions (Cen and Raghavan 2023). Even when an AI algorithm works well for most scenarios, it might greatly underestimate a particular human being's claim to a resource, and it would be unfair for that human to have no right to appeal whatsoever (Cen and Raghavan 2023).

Existing law already implies a right to human decision. Some scholars argue that a right to human decision is already implied by existing laws. A recent paper argues that international human rights law recognizes this right on the grounds that denying the right to human decision can have a "dehumanizing" impact and treat individuals as members of algorithmically categorized groups, denying individuals their status as "human beings endowed with human dignity" (Shany 2023). The paper argues that a 1978 French law already enshrined a right to human decision, a 1995 EU directive granted individuals the right to be free from many decisions based solely on automated processing, and existing international human rights conventions already provide philosophical justifications for this right (Shany 2023). The rationale for a right to human decision is similar to many of the classic justifications for universal human rights, specifically "arbitrariness, lack of accountability, political participation, inequality and human dignity" (Shany 2023).

A right to human decision protects human dignity and agency. The right to human decision can provide humans a sense of agency over decisions that impact their lives (Henin and Le Métayer 2021). Algorithmic decision-making often feels impersonal or dehumanising, and the right to human decision ensures that humans are not simply reduced to numbers and data points (Binns et al. 2018). For instance, a human judge can exercise mercy and empathize with humans, not just treat them as data points (Huq 2020). A right to human decision helps ensure that humans are not forced to navigate a technocratic, automated universe, but are free to live their lives with a sense of human dignity.

4.2 Arguments against a right to human decision

A right to human decision might lead to worse outcomes. A right to human decision does not necessarily lead to better or more ethical decisions. For instance, several studies have found that humans are often unable to effectively evaluate the accuracy of an algorithm's decision-making, exhibit bias in their own decision-making, or do not meaningfully understand the outputs of an algorithm (Green and Chen 2019b; Lai and Tan 2019; Cabitza 2019; Wang and Yin 2021; Green and Chen 2019a). Human review might lead to worse outcomes than algorithms alone while providing only the perception of protection from poorly designed algorithms (Green 2022). The judgment of a human reviewer will often be inconsistent and might not lead to better outcomes than AI alone (Green 2022).

A human reviewer might be biased compared to an algorithm. Humans can sometimes be more "biased" than a carefully designed algorithm; people often make arbitrary decisions based on temporary feelings, whereas well-designed algorithms might be less temperamental (Abrusci and Mackenzie-Gray Scott 2023). Just because a human is reviewing an algorithm's decisions does not mean that the human will actually protect against algorithmic harms; a human might actually introduce new biases. Many studies

demonstrate that people are not good at evaluating the quality of an algorithm and determining whether an algorithm's results were fairly reached (Lai and Tan 2019; Springer, Hollis, and Whittaker 2017; Yeomans et al. 2019). With proper design, AI could be fairer and less arbitrary than government bureaucrats and other humans (Barocas, Hardt, and Narayanan 2019).

Alternatives to a right to human decision might better protect against algorithmic harms. There are many limitations to a right to human decision, such as the need to educate a large amount of human reviewers to fairly review AI outputs. The ethical goals of a right to human decision might be accomplished more effectively and with less bureaucracy by using other methods. Impact assessments or auditing of large AI systems might allow ethical considerations to be addressed before an algorithm is released to the public, without the potential bureaucracy of a right to human decision (Huq 2020).

It's not clear what a meaningful human decision looks like, and it could be impractical. It is challenging to explain what a right to human decision actually entails (Green 2022). For instance, is human review of an appeal required within a certain amount of time? Who is qualified to handle an appeal? How much consideration of an appeal is necessary for it to count? If a human simply "rubber stamps" an AI decision, that likely would not represent a meaningful appeal (Green 2022). Moreover, if AI algorithms are constantly bogged down by administrative challenges, then the potential benefits of these algorithms might be unrealized. After all, many decisions are time sensitive. Think about the fractions of a second that people expect for online recommendations or financial transactions to be completed. Allowing challenges to AI decisions might be impractical or introduce substantial delays.

A right to human decision does not necessarily provide for meaningful human review in practice. There is no guarantee that human review will be meaningful or occur in a prompt manner. Consider content moderation on social media platforms. Many social media platforms automatically moderate content and allow users to appeal decisions, but some users have complained that content moderation is arbitrary or haphazard. Even when users appeal automated decisions to remove their posts or suspend their accounts, they often report receiving no response or having to navigate substantial bureaucratic hurdles to receive a human reconsideration or response (Myers West 2018). This example emphasizes that being able to appeal automatic decisions does not always mean that satisfactory human review can be rendered.

A right to human decision could exacerbate information inequalities. Many of the people who are impacted by AI decision-making might not necessarily recognize when or how to employ the right to human decision. In use cases where AI systems distribute limited goods (for instance, when AI is used to make employment decisions or decide who qualifies for a food assistance program), allowing the right to human decision might exacerbate inequalities. There will inevitably be scenarios where people who have greater resources, more time, or are more well-connected are likelier to appeal decisions, skewing the overall distribution of goods in an unfair direction. A right to human decision could even lead to rent seeking behavior where well-connected individuals shop around or manipulate algorithms, especially when public goods are involved (Congleton, Hillman, and Konrad 2008). AI algorithms might be designed to assuage the concerns of higher-information users who are more likely to appeal, even if doing so causes worse outcomes for others.

5 Discussion

If a right to human decision is implemented, there are many practical considerations that will shape the actual exercise of such a right. For instance, most discussions of the right to human decision assume a dichotomy between "human" and "AI" decisions, but in reality there are many decisions that will be jointly made by both humans and AI. Researchers have made many proposals for effective human-AI decision making and collaboration (Lai et al. 2021; Jarrahi 2018; Ferreira and Monteiro 2021; Liu, Lai, and Tan 2021). For instance, one paper found that a hybrid human-AI team of endoscopists outperformed the diagnostic accuracy of humans alone, and that AI substantially changed the diagnosis decisions of humans in the human-AI team (Reverberi et al. 2022). If AI significantly influences a human being's decision, is that an "AI" decision that falls under the right to human decision? Would a right to human decision apply to decisions where AI is consulted in any manner, decisions where the use of AI passes a certain threshold, or only when decisions are made solely by AI? Moreover, will a human reviewing a decision be asked to refrain entirely from the influence of AI, and is that even possible to guarantee? Defining what a "human" decision and what an "AI" decision are is critical to the practical implementation of a potential right to human decision.

Moreover, if there is a right to contest AI decision-making, who decides which appeals are valid? The maker of the algorithm, a government regulator, public opinion, or another party? Will special certifications be launched for humans to qualify as "ethical" reviewers of algorithmic decisions, and if so, will the benefits outweigh the bureaucratic costs? Even if humans have the right to contest AI decision-making to a company employee, a large AI company could prevent real consideration of most appeals by relying on human reviewers who are trained in-house. But if humans appeal an AI algorithm's results to a government regulator, that regulator might not always understand how the AI algorithm works, create unnecessary bureaucracy, or make arbitrary decisions about which appeals are valid and which ones are not. Although delegating AI appeals to evaluation by experts may appear desirable, experts have their own interests. Considering the vast amount of AI decisions that could become subject to human review, implementing a right to human decision would likely create significant employment opportunities, although they would be created by a legal regime that could change as societies decide that certain decisions do or do not require a right to human decision.

Hiring and training humans to review AI decisions would require financial resources and appropriate compensation. Will companies, governments, and other actors be allowed to charge application fees to defray some of these costs, or will it be legally required that a right to human decision is free for applicants? Will the indirect economic costs of implementing a right to human decision be shifted to all who interact with an AI system, or primarily those who appeal? Moreover, a right to human decision would have minimal impact if it took too long to receive a human decision. For instance, if I am rejected for a new bank account and file an appeal that takes 13 months, or have to continuously submit unnecessary, bureaucratic paperwork, then I cannot meaningfully exercise a right to human decision. In order for a right to human decision to be meaningful, a human review must occur within a reasonable period of time.

In some cases, an AI algorithm will benefit some people while harming others; it is not always the case that AI has a uniformly positive (or negative) impact on all people. For instance, an AI algorithm that handles university admissions will advantage some people and harm others. One student who appeals the results of the AI algorithm will have different interests than another student who appeals the algorithm, and there will often be trade-offs between satisfying different appeals.

Social science research has concluded that people prefer having the ability for humans to review AI decision-making, and that they value the ability to participate, put forward their best case, and feel heard (Lyons et al. 2022; Lyons, Miller, and Velloso 2023; Aljuneidi et al. 2023). However, there are many ways to design an appeals process and many potential elements that could be contested. Would humans contest the final outcome that an algorithm generates and refer it to human review? Or could humans contest the assumptions embedded in the training data, the reasons that an explainable algorithm came to a decision, or the intermediate results that the algorithm takes? Unfortunately, many neural network algorithms are "black boxes" that are difficult or impossible to explain. Fully explainable AI is not currently possible (although there are proposals for "counter-factual explanations" that explain what would have happened if the input data had changed slightly) and might require an entirely different method for training AI beyond deep learning (Holzinger et al. 2022; Keane et al. 2021; Wachter, Mittelstadt, and Russell 2017). There is no guarantee that emerging methods for explainability will generate high-quality results that would fit into a more detailed appeals process beyond allowing the right to contest the final decision, especially given the past failures of rule-based AI "expert" systems (Wooldridge 2018).

There will inevitably be cases where there is no exact human alternative to an AI system's solution, making a right to opt out sometimes challenging to implement. It might be extremely difficult to avoid AI entirely in remote jobs, cybersecurity, or healthcare systems. What would it mean to opt out of AI-generated medical treatments? Moreover, since AI can perform many tasks more quickly than humans (e.g. automated credit scoring is nearly instant), does a full human alternative have to be as robust as the AI approach in order for a right to opt out to be meaningful? If a human alternative is inferior, then those who use AI might gain advantages in efficiency or knowledge that provide unequal opportunities. If some people opt out of AI and others choose not to, then it might be very difficult to provide the same opportunities or outcomes to both groups of people. Given the influence of AI on cultural and social norms, people who opt out of AI might face cultural isolation from those who opt into AI, and vice-versa. Therefore, the right to human decision faces many practical challenges that must be addressed.

6 Conclusion

Considering that some form of a right to human decision has already been proposed by nations and international organizations that represent most of the world's population, it is critical for future work to examine the practicalities of a right to human decision. It is especially important to define what a "human" decision is and how it differs from an "AI" decision, and clarify when a decision made by a human with the assistance of AI counts as a "human" decision and "AI" decision. Moreover, concerned parties should identify specific use cases (e.g. credit scoring, evaluating job applications, predictive medicine) and assess whether and when it will be necessary to enforce a right to contest, right to opt out, or no additional right for those use cases. Many existing AI ethics frameworks call for a broad right to be human oversight or decision-making, but the details of how these proposals would operate in practice do matter, and a one size fits all approach might not work. For instance, it is unlikely that a human review of an automated decision (or a fully human decision that excludes machines) would look the same in healthcare and financial services. Future work should examine the nuances of actually implementing a right to human decision and assess the feasibility and wisdom of implementation.

It is important to remember that the advent of AI decision-making is not necessarily incompatible with human decision-making. Humans can often improve their decision-making skills by interacting with AI. For instance, professional Go players were found to make significantly better decisions since the advent of AlphaGo (the superhuman AI that plays Go) compared to the six decades before that, since humans have been challenged to create new strategies and try out new types of decisions (Shin et al. 2023). As AI advances, it is likely that the boundaries between a strictly "human" and "AI" decision will blur, and a potential right to human decision will have to adapt accordingly. The responsible use of AI can help humanity progress, and we need both ethics and optimism as new technologies emerge.

References

Abrusci, E.; and Mackenzie-Gray Scott, R. 2023. The questionable necessity of a new human right against being subject to automated decision-making. *International Journal of Law and Information Technology*, 31(2): 114–143.

African Union. 2014. African Union Convention on Cyber Security and Personal Data Protection.

Aljuneidi, S.; Heuten, W.; Tepe, M.; and Boll, S. 2023. Did that AI just Charge me a Fine? Citizens' Perceptions of AI-

based Discretion in Public Administration. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*, 57–67.

Barocas, S.; Hardt, M.; and Narayanan, A. 2019. *Fairness and Machine Learning: Limitations and Opportunities*. fairmlbook.org. http://www.fairmlbook.org.

Bayamlıoğlu, E. 2022. The right to contest automated decisions under the General Data Protection Regulation: Beyond the so-called "right to explanation". *Regulation & Governance*, 16(4): 1058–1078.

Binns, R.; Van Kleek, M.; Veale, M.; Lyngs, U.; Zhao, J.; and Shadbolt, N. 2018. 'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 Chi conference on human factors in computing systems*, 1–14.

Binns, R.; and Veale, M. 2021. Is that your final decision? Multi-stage profiling, selective effects, and Article 22 of the GDPR. *International Data Privacy Law*, 11(4): 319–332.

Bowman, S. R. 2023. Eight things to know about large language models. *arXiv preprint arXiv:2304.00612*.

Cabitza, F. 2019. Biases affecting human decision making in AI-supported second opinion settings. In *Modeling Decisions for Artificial Intelligence: 16th International Conference, MDAI 2019, Milan, Italy, September 4–6, 2019, Proceedings 16*, 283–294. Springer.

Cen, S. H.; and Raghavan, M. 2023. The Right to Be an Exception to a Data-Driven Rule. *MIT Case Studies in Social and Ethical Responsibilities of Computing*, (Winter 2023). Https://mit-serc.pubpub.org/pub/right-to-be-exception.

Center for Security and Emerging Technology. 2021. Ethical Norms for New Generation Artificial Intelligence Released (translated from Chinese).

Chinese Ministry of Science and Technology. 2021. New Generation Artificial Intelligence Ethics Code.

Congleton, R. D.; Hillman, A. L.; and Konrad, K. A. 2008. Forty years of research on rent seeking: an overview. *The Theory of Rent Seeking: Forty Years of Research*, 1: 1–42.

Court of Justice of the European Union. 2023. Case C-634/21.

European Union. 2016. *General Data Protection Regulation*.

Ferreira, J. J.; and Monteiro, M. 2021. The human-AI relationship in decision-making: AI explanation to support people on justifying their decisions. *arXiv preprint arXiv:2102.05460*.

Goodman, B.; and Flaxman, S. 2017. European Union regulations on algorithmic decision-making and a "right to explanation". *AI magazine*, 38(3): 50–57.

Government of Canada. 2023. Responsible use of artificial intelligence (AI).

Government of Uruguay. 2019. Artificial Intelligence Strategy for the Digital Government.

Green, B. 2022. The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, 45: 105681.

Green, B.; and Chen, Y. 2019a. Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. In *Proceedings of the conference on fairness, accountability, and transparency*, 90–99.

Green, B.; and Chen, Y. 2019b. The principles and limits of algorithm-in-the-loop decision making. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 1–24.

Henin, C.; and Le Métayer, D. 2021. A framework to contest and justify algorithmic decisions. *AI and Ethics*, 1(4): 463–476.

Holzinger, A.; Saranti, A.; Molnar, C.; Biecek, P.; and Samek, W. 2022. Explainable AI methods-a brief overview. In *xxAI-Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers*, 13–38. Springer.

Houweling, E. 2021. China's new AI rulebook: humans must remain in control.

Huq, A. 2020. A right to human decision. *Virginia Law Review*, 106(3): 611–688.

Jarrahi, M. H. 2018. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business horizons*, 61(4): 577–586.

Kaminski, M. E.; and Urban, J. M. 2021. The Right to Contest AI. *Columbia Law Review*, 121(7): 1957–2048.

Keane, M. T.; Kenny, E. M.; Delaney, E.; and Smyth, B. 2021. If only we had better counterfactual explanations: Five key deficits to rectify in the evaluation of counterfactual xai techniques. *arXiv preprint arXiv:2103.01035*.

Lai, V.; Chen, C.; Liao, Q. V.; Smith-Renner, A.; and Tan, C. 2021. Towards a science of human-ai decision making: a survey of empirical studies. *arXiv preprint arXiv:2112.11471*.

Lai, V.; and Tan, C. 2019. On human predictions with explanations and predictions of machine learning models: A case study on deception detection. In *Proceedings of the conference on fairness, accountability, and transparency*, 29–38.

Liu, H.; Lai, V.; and Tan, C. 2021. Understanding the effect of out-of-distribution examples and interactive explanations on human-ai decision making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2): 1–45.

Lyons, H.; Miller, T.; and Velloso, E. 2023. Algorithmic Decisions, Desire for Control, and the Preference for Human Review over Algorithmic Review. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 764–774.

Lyons, H.; Wijenayake, S.; Miller, T.; and Velloso, E. 2022. What's the appeal? Perceptions of review processes for algorithmic decisions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–15.

Mishra, S.; Panda, R.; Phoo, C. P.; Chen, C.-F. R.; Karlinsky, L.; Saenko, K.; Saligrama, V.; and Feris, R. S. 2022. Task2sim: Towards effective pre-training and transfer from synthetic data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9194–9204.

Myers West, S. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society*, 20(11): 4366–4383.

National Institution for Transforming India, Aayog. 2021. Towards Responsible AI for All, Part 1.

Ploug, T.; and Holm, S. 2021. The right to contest AI profiling based on social media data. *The American Journal of Bioethics*, 21(7): 21–23.

Ray, P. P. 2023. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*.

Reverberi, C.; Rigon, T.; Solari, A.; Hassan, C.; Cherubini, P.; and Cherubini, A. 2022. Experimental evidence of effective human–AI collaboration in medical decision-making. *Scientific reports*, 12(1): 14952.

Sarra, C. 2020. Put dialectics into the machine: protection against automatic-decision-making through a deeper understanding of contestability by design. *Global Jurist*, 20(3): 20200003.

Shany, Y. 2023. The Case for a New Right to a Human Decision Under International Human Rights Law. *Available at SSRN 4592244*.

Shin, M.; Kim, J.; van Opheusden, B.; and Griffiths, T. L. 2023. Superhuman artificial intelligence can improve human decision-making by increasing novelty. *Proceedings of the National Academy of Sciences*, 120(12): e2214840120.

Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. 2016. Mastering the game of Go with deep neural networks and tree search. *nature*, 529(7587): 484–489.

Singapore Personal Data Protection Commission. 2020. Model Artificial Intelligence Governance Framework, Second Edition.

Springer, A.; Hollis, V.; and Whittaker, S. 2017. Dice in the black box: User experiences with an inscrutable algorithm. In *2017 AAAI Spring Symposium Series*.

The White House. 2022. Blueprint for an AI Bill of Rights. Figshare https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

United Arab Emirates Government. 2022. AI Ethics: Principles and Guidelines.

Vaccaro, K.; Xiao, Z.; Hamilton, K.; and Karahalios, K. 2021. Contestability for content moderation. *Proceedings of the ACM on human-computer interaction*, 5(CSCW2): 1–28.

Wachter, S.; Mittelstadt, B.; and Russell, C. 2017. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv. JL & Tech.*, 31: 841.

Wang, X.; and Yin, M. 2021. Are explanations helpful? a comparative study of the effects of explanations in ai-assisted decision-making. In *26th International Conference on Intelligent User Interfaces*, 318–328.

Wooldridge, M. 2018. *Artificial Intelligence: Everything you need to know about the coming AI. A Ladybird Expert Book*, volume 27. Penguin UK.

Yeomans, M.; Shah, A.; Mullainathan, S.; and Kleinberg, J. 2019. Making sense of recommendations. *Journal of Behavioral Decision Making*, 32(4): 403–414.

7 Appendix

See the following page for a table outlining textual proposals from governments for a right to human decision.

Table 1: Governments and international organizations with jurisdiction over most of the world's population have already proposed some form of a right to human decision or human oversight over AI

Entity	**Proposal**	**Year released**	**Legally binding?**
African Union	"A person shall not be subject to a decision which produces legal effects concerning him/her or significantly affects him/her to a substantial degree, and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him/her" (African Union 2014).	2014	Yes, since 2023
Canada	Responsible AI involves "explaining automated decisions to people impacted by them and providing them with opportunities to contest decisions and seek remedies, which could involve human review, where applicable" (Government of Canada 2023).	2023	No
China	"Ensure that humans have fully autonomous decision-making rights and that they have the right to accept or reject AI-provided services, the right to withdraw from AI interactions at any time, and the right to terminate AI system operations at any time. Ensure that AI is always under human control" (Center for Security and Emerging Technology 2021; Chinese Ministry of Science and Technology 2021).	2021	No
European Union	"The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. [This] shall not apply if the decision: is necessary for entering into, or performance of, a contract between the data subject and a data controller; is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or is based on the data subject's explicit consent" (European Union 2016).	2016	Yes, since 2018
India	"If the potential degree of harm for a decision is expected to be high, have appropriate mechanisms in place so stakeholders can contest and humans can get involved in the decision making process" (National Institution for Transforming India, Aayog 2021).	2021	No
Singapore	When there is a high probability of harm and harm would have "high severity," "human oversight is active and involved, with the human retaining full control and the AI only providing recommendations or input. Decisions cannot be exercised without affirmative actions by the human, such as a human command to proceed with a given decision." It is important to note that Singapore proposes that this is one of three acceptable approaches to AI decision-making and that there are cases where AI should be allowed to autonomously make decisions without human oversight. (Singapore Personal Data Protection Commission 2020).	2020	No
United Arab Emirates	"AI subjects should be able to challenge significant automated decisions concerning them and, where appropriate, be able to opt out of such decisions. AI operators using AI systems to inform significant decisions should provide procedures by which affected AI subjects can challenge a specific decision concerning them. AI operators should consider such procedures even for non-significant decisions. AI operators should make affected AI subjects aware of these procedures and should design them to be convenient and user-friendly. AI operators should consider employing human case evaluators to review any such challenges and, when appropriate, overturn the challenged decision" (United Arab Emirates Government 2022).	2022	No
United States	"You should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems you encounter. You should be able to opt out from automated systems in favor of a human alternative, where appropriate...In some cases, a human or other alternative may be required by law. You should have access to timely human consideration and remedy by a fallback and escalation process if an automated system fails, it produces an error, or you would like to appeal or contest its impacts on you. Human consideration and fallback should be accessible, equitable, effective, maintained, accompanied by appropriate operator training, and should not impose an unreasonable burden on the public" (The White House 2022).	2022	No
Uruguay	"Technological solutions based on AI must have a clearly identifiable person responsible for the actions derived from the solution actions...When the application and / or development of AI-based solutions present ethical dilemmas, they must be addressed and resolved by human beings" (Government of Uruguay 2019).	2019	No