# Extended Abstract:

# Rethinking the Right to a Human Decision-Maker: From Right to Constitutive Necessity

Blinded First Author[1]

[1] Blinded University Affiliation

**Abstract:** With the increasing encroachment of Big Data and AI governing critical domains, the legal perspective has resorted to suggesting novel human rights to protect individuals and communities and ensure meaningful human control, algorithmic accountability, and responsible AI governance. This paper explores one of such rights: the right to a human decision-maker. After a brief outline of the growing advocacy for this right, primarily following the work of Onora O'Neill, I sketch the various challenges of positing a new positive human, i.e., universal, right, including arduous allocation and questions on feasibility. Not all hope is lost, however, as I argue that framing a human decision-maker as a novel right is in fact unnecessary. Extending an argument on moral relationships by Christine Korsgaard, I argue that the structure of the claim right and correlative duty relationships necessitates human decision-makers. Here we see that an essential reciprocity of rights that requires both role-reversibility and constitutive symmetry between the right-holders and decision-makers, or duty-bearers. Importantly, while this extends Korsgaard's account of the moral law, I show why this also applies to democratic positive law. This argument about the essential moral and legal value of human intervention therefore sidesteps well-known difficulties of positive right allocation and shows that a human decision-maker is not merely a novel, desirable pursuit of responsible AI development and implementation, but is a constitutive feature of rights *tout court*.

**Keywords:** Rights, Human decision-maker, AI governance, Korsgaard

**Word Count:** 2 730

## 1 Introduction

With the growing entrenchment of Big Data and AI decision-making in critical domains—from criminal sentencing and financial lending to healthcare diagnostics—various jurisdictions have proposed novel (digital) human rights that aim to support responsible AI governance and protect individuals and communities from potential detriment arising from this increasing delegation to algorithmic decision-making. One of

these rights is the *right to a human decision-maker*. Here, as the name suggests, supporters argue that individuals subject to automated decisions with legal or moral saliency, e.g., an AI system deciding on their eligibility for a loan, ought to have the opportunity to alternatively demand human-made decisions.

In this paper, I argue that while the postulated value of a human decision-maker in the age of AI is indeed worthwhile, understanding it as a novel human right is unnecessary insofar that it fails to capture the deeper legal and moral importance of a human decision-makers in claim rights-based relationships. Through an examination of the structure of these relationships, I propose that a human decision maker is not merely a desirable pursuit, but is a constitutive feature of rights *tout court*. In making this extended argument, it not only deepens the acknowledged human-centred response to responsible AI governance by suggesting it represents fundamental requirements for legitimate moral and democratic legal authority, but also importantly side-steps recognized issues of advancing novel positive rights, such as arduous allocation and questions on feasibility.

To support this argument, I first provide a brief history of the right to a human decision-maker in Section 2. Through the lens of Onora O'Neill, Section 3 outlines primary concerns regarding the advancement of positive human rights. To address these difficulties, Section 4 revisits the framework for claiming a right and shows that through its reciprocal structure, any rights-based relationship necessitates a human ensuring its enforcement, i.e., a human duty-bearer or decision-maker. Section 5 offers the conclusion and outlines future work.

## 2 Right to a Human Decision-Maker Explained

Yuval (2023) highlights that faced with the growing concern that human understanding and control, particularly in ethically-salient decision-making contexts, is increasingly outsourced to data-driven algorithmic systems, the legal community has most commonly responded with proposals for novel human rights. Through various legal instruments, such as regulations, draft bills, and acts, this includes the right to internet access (e.g., Finland's *Communications Market Act*, Section 60 c) and the right to personal data erasure or rectification (e.g., the EU GDPR, Articles 16 and 17). However, Yuval notes that these rights may not need attribution as novel human rights because their motivating goals may already be covered by widely-accepted existing rights, such as to privacy, freedom of expression, and due process.

It is, however, the right to a human decision maker that Yuval considers the most distinct given its prima facie disconnect from existing human rights frameworks. While the formulation of this right emerged well beyond the proliferation of Big Data and AI through the 1978 French law on Data Processing, Data Files and Individual Liberties (*Loi relative à l'informatique, aux fichiers et aux libertés*), it is most commonly attributed to Article 22(1) of the EU GDPR, which establishes the general prohibition

against subjecting individuals solely to automated decisions.[1] This right has also gained traction outside of the EU, including in the US, Africa, and international treaties including the Council of Europe Convention 108+.

The shift of the right to human decision-maker toward recognition as a novel *human* right—that is, a right that is presented as universal, inalienable, and backed by state obligations—is marked by its inclusion in bills and charters of digital rights (e.g., the US *Blueprint for an AI Bill of Rights* and the EU *Declaration on Digital Rights and Principles for the Digital Decade*) as well as what Yuval argues as the backing of traditional moral justifications for human rights, including concerns about abuse of power, accountability, democratic legitimacy, and the preservation of human dignity.

This growing call to codify human intervention also emerges through similar terminology including the demand for a *human-in-the-loop* (e.g., Coeckelbergh 2020) and *meaningful human control*—the latter emerging from concerns over lethal autonomous weapons systems and their implications for international humanitarian law (Brand 2023b; Asaro 2020). Also connected is the widely-discussed epistemic right to an explanation—also advanced as an interpretation of Article 22(1) of the GDPR as well as Article 86 of the EU AI Act and Canada's proposed Artificial Intelligence and Data Act (AIDA)[2]—insofar that the moral and legal calls for explainability not only call for transparency, but also the need for intelligible socially-situated explanations and accountability of a human agent (e.g., Cowls and Floridi 2019; Vredenburgh 2022).[3] In a similar vein, Lani Watson (2021) argues for the right to know.

In sum, this section highlights that through various regulations, draft bills, acts, and varying terminology, the call for human intervention in decision-making processes where AI has already, can, or likely will, takeover is a sustained and growing effort.

## 3 Disadvantages of Positive Human Rights

Positing a novel right to a human decision-maker, however, faces difficulties that other positive human rights face, such as rights to healthcare or education. To outline these key difficulties, this section primarily relies on Onora O'Neill's (1996; 2005) assessment of positive human rights.

First, following legal philosopher Wesley Hohfeld (1919), I clarify that calls for a right to a human decision-maker, as with any right in the strictest sense, refer to interpersonal claims and correlated duties. To hold a claim-right is to be positioned to

[1] Article 22(1) of the GDPR: "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."

[2] Although the interpretation that the GDPR creates such a right is contested, see Huq (2019).

[3] For this reason, the term *explicability* is sometimes used to distinguish the deeper calls for interpretability and accountability from mere transparency requirements (e.g., Cowls and Floridi 2019).

demand a certain treatment, or duties, from another.[4] These directed duties and their corresponding claims therefore represent the same interpersonal (legal or moral) relationship from different perspectives. For example, an accused claims the right to a fair trial by recognizing judge's duty to ensure the trial follows the correct procedure.

The prerequisite to forming and sustaining such rights therefore depends on identifying correlative duties and duty-bearers—decision-makers whose judgements are constrained by duties towards the claim-holders on the receiving end of their judgements. Without this identification, one risks sliding toward ineffectual political rhetoric, lip service, or what O'Neill (1996) identifies as "manifesto rights" where "nothing can be claimed, waived, or enforced" (p. 129). This echoes Simone Weil's (1952) argument for the subordination of rights to their correlative duties: "A right is not effectual by itself, but only in relation to the obligation to which it corresponds." (p. 2).

Traditionally understood, this interpersonal duty-right relationship is bifurcated into negative (liberty) and positive rights. Liberty rights, such as to privacy and freedom of association, primarily demand non-interference and therefore can be allocated universally insofar that, in essence, the duty involves leaving one another alone. Positive rights are claims to goods and service, such as healthcare or education, and can only be realized through allocating (finite) resources, performing of services, and determining which individuals, institutions, and socio-economic systems will enforce these duties (O'Neill 2005).[5] Consider that a proposed universal right to healthcare is ineffective until first identifying who receives which services and where, who is to deliver them, and which institutional schemes, i.e., a healthcare program, will allocate and organize this right. Moreover, given that positive rights at first instance rely on the expenditure of resources, in many it cannot reasonably be allocated to *all*.

The right to a human decision maker or explanation is no different. They remain unclaimable without first looking at the various socio-economic and institutional schemes that outline who can and ought to shoulder corresponding duties and which particular duties each ought to offer. This is therefore a relatively arduous process likely resulting in a patchwork implementation of the right, i.e., *special* rights arising from specific relationships, rather than the universal approach currently proposed. Legal culture has sometimes addressed this issue with human rights by turning towards special conventions that outline these subject-specific duty-right relationships, such as the UN *Convention on the Rights of Disabled Persons* (Etinson 2013).

Moreover, despite algorithmic decision-making proliferating across domains, it remains unclear whether claiming a human decision-maker is feasible or warranted in every instance. For example, when AI is used to generate music playlists, this seemingly does not warrant a claim to demand additional resources to have a human explain why certain songs were offered; it is not unjust to receive distasteful music. Whereas it may be the case that the specific relationship between bank customers and postal banks

[4] According to Hohfeld (1919), privileges, powers, and immunities are also related to the term right, but are not rights in the strictest sense as they lack direct correlation with duties.

[5] Etinson (2013) points out that both rights give rise to both negative and positive duties. For example, forming police forces and judicial systems are necessary to enforce negative liberty rights. However, the point here is to indicate the primary duties to such rights, rather than the supporting secondary duties that may help others enforce their primary ones.

with a public service mandate warrants the demand for human-made decisions (Brand 2023a). Attempting to bestow this novel positive right *universally* therefore may prove unfair, overreaching, or unwarranted.

This section does not refute the value or possibility of forming novel positive rights—indeed, rights like education and healthcare have been implemented worldwide with varying degrees of success. Rather, it highlights the difficult road ahead for asserting such rights and the likelihood of patchwork implementation rather than achieving universal recognition.

## 4 The Right Revisited

In this last section, I argue that positing a novel right to a human decision-maker is unnecessary. Rather, a human decision maker is a constitutive element of *any* claim right relationship.

To do so, I turn to Christine Korsgaard (2018) who reminds us that for the moral law—the normative relationship of duties and corresponding claim rights—to be considered objective—that is, having authority over right-holders and duty-holders and anyone who may later find themselves in either position, it must be perceived as *reciprocal*. To illustrate what she means, Korsgaard provides the example of French laws: an individual can oblige another individual to act a certain way by using the authority of French laws. But this is only a legitimate claim on the person *if and only if* they are French. The law must apply to them. And any person claiming that another person must adhere to this duty (or law) insofar as providing them with a reason to acknowledge and respect it, *supposes that they too are bound by the same authority* (Korsgaard 2018).

I propose that the reciprocity in duty-right relationships demands two conditions: role-reversibility and constitutive symmetry. If an individual seeks a bank loan, we might expect that they have the legitimate claim that their application is reviewed on merit and therefore, that the bank has the duty to review it solely on merit. But for this duty-right relationship to be consider authoritative law, the individual must also concede that if their roles were reversed—they become the bank employee and the employee is the applicant—the bank employee (now applicant) could make the same claim.

For this role-reversibility to be feasible, both individuals must be able to consider that they may find themselves in the opposing role, either as right-holder or duty-bearer. This implies that they could indeed share the roles—it entails they must share basic constitutive qualities. Here, I propose that this quality is being a rational human being.

Lacking qualities constitutive of humans, namely our vulnerabilities (both physical and cognitive) that function alongside, shape, and sometimes override our rationality (e.g., MacIntyre 1999), AI systems, as machines, will always be in an asymmetrical relationship with us. Therefore, irrespective of any hypothetical advancements of AI systems, it will never been able to participate in the reciprocal legislation underlying all duty-right relationships. In other words, wherever humans have legitimate claims,

AI cannot take on the role as the corresponding duty-bearer, or decision-maker. Only human decision-makers can fulfill this role.

Importantly, reciprocity also constitutes an integral feature of democratic positive law, imposing role-reversibility between the relevant stakeholders of the law, i.e., the duty-bearers, or decision-makers, such as judges, juries, police, and parliamentarians, and those affected by their decisions. Recent work by Brennan-Marquez and Henderson (2019; 2023) draws from a wide-range of legal theory, including 16th century English reformer John Bradford, Rawls' theory of justice, and Scanlon's theory of fairness. They argue that there is a baseline equality infused in each legal judgement—equality in which AI cannot participate—characterized by the sentiment "there but for the grace of God." All this means is that whenever someone finds themselves on either side of the law, perhaps they are sitting in the dock rather than on the bench, or the difference between an elected politician and elector, it is ultimately due to chance: "If the winds of fortune blew different, many of our individual lives—and at some level, all of social life—would be different as well." (2019, p. 150). I also refer to Moore (2023) who links role-reversibility to foundational features of democracies like the dispensability of governments, provisionality, uncertainty, distributed authority, and open-endedness.

It is worthwhile noting that these arguments about moral and positive law refer to the ideal standards of law. There are certainly many instances, as I mentioned above, where we fail to maintain the standards of reciprocity in law; consider the systemic barriers numerous jurisdictions willfully uphold that ensure certain groups of people remain on one side of the law, rendering the vision of role-reversibility a distant ambition. We can and should always do better. But these faults are more failures of society than of the concept of democratic law. While supporters of introducing AI in legal decision-making may argue that its supposedly unbiased approach and consistency in outputs might get societies closer to equitable treatment under the law (e.g., Huq 2019), following this paper, it in fact will only push us further away from the democratic standards we wish to uphold.

Through this reminder of the integral importance of reciprocity to both the moral and positive law, we see that whenever we have claim rights, whomever the claim is directed, whose decisions and how they act are constrained by our claims, must share with us the possibility of role-reversal and therefore basic constitutive features—they must also be human.

# 5 Conclusion

This paper proposes a deeper argument supporting the moral and legal value of human decision-makers within responsible AI governance frameworks. Rather than attempting to posit a novel human right, we see that human decision-making amounts to a constitutive feature of how we already structure claim right relationships. This account is both simpler, insofar as it sidesteps the initial allocation problems of positive human rights, and more demanding because it establishes human decision-making as the ideal wherever moral and legal claim rights are found.

Given this paper's scope, two key points are left to answer in future work. First, because this argument hinges on robust human decision-making, i.e., that humans maintain genuine control, more must be said about what constitutes genuine control in decision-making processes, especially when AI-assisted tools may remain part of the procedure. This will involve a survey of cognitive vulnerabilities, such as well-documented complacency, over-reliance, diffidence, automation bias, and placing unexamined trust in their decisions—called *loafing* (Zerilli et al. 2022), as well as how courts currently approach on determining the principle decision-maker in Human-AI teaming contexts (e.g., SCHUFA Holding (Scoring), C-634/21, 2023). As this paper argues from the Kantian tradition, this answer will also likely look to the traditional understanding of human agency where the agent is expected sustain epistemic control wherein they can understand, adopt, and communicate the salient motivating reasons that guide the decision. Explainable AI (XAI), with its potential for providing reasoned explanations, may prove useful in AI-assisted decision making (Barredo Arrieta et al., 2020; Baum et al., 2022).

Second, while this augment addresses the arduous allocation process of asserting a novel positive human right, it does not answer how the human decision maker ought to explain themselves to other humans. This will require an interdisciplinary examination drawing from philosophy, sociology, psychology and the cognitive sciences that takes into account the social contexts and expectations of explanations (e.g., Miller 2019).

**Acknowledgments.** Blinded

**Disclosure of Interests.** No competing interests to declare.

## References

1. Asaro, Peter. "Autonomous Weapons and the Ethics of Artificial Intelligence." In *Ethics of Artificial Intelligence*, edited by S. Matthew Liao, 212-236. Oxford: Oxford University Press, 2020.https://doi.org/10.1093/oso/9780 190905033.003.0008
2. Barredo Arrieta, Alejandro et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." *Information Fusion* 58 (2020): 82-115. https://doi.org/10.1016/j.inf fus.2019.12.012
3. Baum, Kevin et al. "From Responsibility to Reason-Giving Explainable Artificial Intelligence." *Philosophy & Technology* 35, no. 12 (2022). https://doi.org/ 10.1007/s13347-022-00510-w
4. Brand, Joshua. "The Duty to Implement Explainable Artificial Intelligence: A Case Study of Public Service Postal Banks." *Canadian Journal of Practical Philosophy* 9, no. 1 (2023a). https://ojs.uwindsor.ca/index.php/cjpp/article/view/8146
5. Brand, Joshua. "Why reciprocity prohibits autonomous weapons systems in war." *AI and Ethics* 3 (2023b): 619-624. https://doi.org/10.1007/s43681-022-00193-1
6. Brennan-Marquez, Kiel and Stephen Henderson. "Artificial Intelligence and Role- Reversible Judgment." *Journal of Criminal Law and Criminology* 109, no. 2 (2019): 137-164. https://dx.doi.org/10.2139/ssrn.3224549

7. Coeckelbergh, Mark. “Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability.” *Science and Engineering Ethics* 26, no. 4 (2020): 2051-2068. https://doi.org/10.1007/s11948-019-00146-8
8. Cowls, Josh and Luciano Floridi. “A Unified Framework of Five Principles for AI in Society.” *Harvard Data Science Review* 1, no. 1 (2019): 1-15. https://doi.org/10.1162/99608f92.8cd550d1
9. Etinson, Adam. “Human Rights, Claimability and the Uses of Abstraction” *Utilitas* 25, no. 4 (2013) : 463-486. https://doi.org/10.1017/S0953820813000101
10. Hohfeld, Wesley Newcombe. *Fundamental Legal Conceptions as Applied in Judicial Reasoning*. Edited by W.W. Cooke. New Haven: Yale University Press, 1919.
11. Huq, Aziz. Z. (2020). A Right to a Human Decision. *Virginia Law Review, 106*(3) (2022) : 611–688.
12. Korsgaard, Christine. *Fellow Creatures: Our Obligations to the Other Animals*. Oxford: Oxford University Press, 2018. https://doi.org/10.1093/oso/ 9780198753858.001.0001
13. MacIntyre, Alasdair. *Dependent Rational Animals: Why Human Beings Need the Virtues*. Open Court, 1999.
14. Miller, Tim. “Explanation in Artificial Intelligence: Insights from the Social Sciences.” *Artificial Intelligence* 267 (2019): 1-38. https://doi.org/10.1016/ j.artint.2018.07.007
15. Moore, Alfred. *Reversibility and Democracy: The Epistemic Functions of Political Competition*. In M. Mackenzie, M. Setala, & S. Kyllonen (Eds.), *Democracy and the Future Future-Regarding Governance in Democratic Systems* (pp. 55-76). Edinburgh University Press, 2023.
16. O’Neill, Onora. “The Dark Side of Human Rights.” *International Affairs (Royal Institute of International Affairs 1944-)* 81, no. 2 (2005): 427-439. https://doi.org/10.1111/
17. O’Neill, Onora*. Towards Justice and Virtue: A Constructive Account of Practical Reasoning*. Cambridge: Cambridge University Press, 1996.
18. Vrendenburgh, Kate. “The Right to Explanation.” *The Journal of Political Philosophy* 30, no. 2 (2022): 209-229. https://doi.org/10.1111/jopp.12262
19. Watson, Lani. *The Right to Know: Epistemic Rights and Why We Need Them*. Abington, New York: Routledge, 2021.
20. Weil, Simone. *The Need for Roots: Prelude to a Declaration of Duties to Mankind*. 1952. Translated by Arthur Wills. London and New York: Routledge, 2005.
21. Yuval, Shany. “The Case for a New Right to a Human Decision Under International Human Rights Law.” Working Paper (2023), available at https://dx.doi.org/10.2139/ssrn.4592244
22. Zerilli, John, Bhatt, Umang, & Weller, Adrian. “How transparency modulates trust in artificial intelligence. ” *Patterns, 3*(4), Article 100455 (2022). https://doi.org/10.1016/j.patter.2022.100455