JOURNAL OF
CROSS-DISCIPLINARY RESEARCH
IN COMPUTATIONAL LAW [CRCL]

The Unbearable Likeness of Being: How Artificial Intelligence Challenges the Social Ontology of International Human Rights Law

Sue Anne Teo*

Abstract

This paper examines how the social ontology that underpins the international human rights framework is being challenged by the affordances of AI/ML systems. To set the stage, this paper adopts a socially situated understanding of human rights – acknowledging the socially embedded nature of individuals within societies. Drawing upon Gould's theory on the social ontology of human rights, the individual is not only socially embedded but it is this social situatedness that enables the exercise of positive agency (including moral and political agency). The role of human rights is then to preserve conditions that enable the exercise of such capacities. While the ubiquity of computational technologies such as AI systems may prima facie seem to embrace and operationalise sociality, the paper highlights three pressure points, arguing that they lead towards the structural atomisation of individuals in ways that are in tension with the normative aims of international human rights law. Data points that group, infer and construct individuals through their likeness, atomise individuals as means to an end through AI/ML systems. Further, the efficiency-driven framing of AI/ML reliant on computational tractability means that individuals risk instrumentalisation through optimisation. Finally, the AI/ML mediated shaping of epistemic and enabling conditions can lead to contextual atomisation – threatening the condition antecedent for socially situated exercise of moral agency and with it, human rights. In diagnosing these structural challenges, this paper provides a deeper mapping of the problem space to inform AI/ML and human rights scholars and practitioners of better accounting for the social ontology of human rights in our computational environments.

Keywords: human rights, social ontology, atomisation, artificial intelligence, international human rights law

Replier: Wayne Wobcke • Associate Professor, UNSW Sydney, rw.wobcke@unsw.edu.au.

Journal of Cross-disciplinary Research in Computational Law

DOI: pending

www.journalcrcl.org

* Postdoctoral Research Fellow, Raoul Wallenberg Institute of Human Rights and Humanitarian Law, Lund University, sue_anne.teo@rwi.lu.se

Introduction

The increasing use of emerging technologies, such as artificial intelligence have raised discrete human rights issues, such as the right to privacy, non-discrimination, freedom of expression and data protection. Less explored, however, are the ways in which artificial intelligence/machine learning ('AI/ML') systems are challenging the very core conceptions that sustain the edifice of the contemporary human rights framework. Policy makers and human rights practitioners assume the sufficiency of the human rights regime in tackling such new challenges, operating on a 'normative equivalency' paradigm that claims to be able to accommodate the novelty of modalities and harms brought forth by emerging technologies such as AI/ML.[1] The normative equivalency paradigm takes as its starting point the understanding that the international human rights framework – set in motion through the 1948 Universal Declaration of Human Rights (UDHR) and expanded through subsequent legal instruments – is sufficient to address challenges posed by new technologies. The online space and offline space are normatively equivalent so to speak. On the other hand, support for this paradigm is waning, as computational technologies such as artificial intelligence – mainly, though not exclusively, through machine learning – introduce novel challenges. Despite contested histories and normative justifications on why we even have human rights,[2] the current international human rights framework was oriented to respond to qualitatively different sites of concern. This can be traced primarily to the necessity of securing a response to the atrocities surrounding the Second World War.[3] which had 'outraged the conscience of mankind' and undermined the 'inherent dignity' of the individual. [4] This paper highlights that one key element of the framework, the social ontology that underpins human rights, is being challenged by the affordances of AI/ML systems.

The conceptual framework adopted by this paper is that of a socially situated understanding of human rights drawing from the theory of social ontology of human rights by Carol Gould. This acknowledges the socially situated and embedded nature of individuals within societies wherein positive agency can prevail. Two reasons motivate this: First, the framework is inductively motivated by tracing existing human rights instruments, case law and academic scholarship, which this paper argues indeed reflects that the theory and practice of human rights is intimately bound up with sociality. Secondly, it is deductively motivated in the sense that the intuitive claims that AI/ML systems are displacing the human and undermining human agency are given more concrete shape and form through the 'structural atomisation' problematique offered in this paper. The first motivation serves a ground-clearing purpose while the latter serves a functional one. Additionally, this work also engages in deciphering why the international human rights framework has increasingly been criticised as inadequate[5] and aims to address the under-examination and under-theorisation of how AI/ML systems impact the foundational assumptions of human rights.

[1] Dafna Dror-Shpoliansky and Yuval Shany, 'It's the end of the (Offline) World as we know it: From Human rights to Digital Human rights – a proposed typology' (2021) 32(4) European Journal of International Law 1249.

[2] Michael Freeman, *Human rights* (Third edition, Polity Press, Cambridge 2017) ch 2; See also Allen Buchanan, 'Why International Legal Human Rights?' in Rowan Cruft, S Matthew Liao, and Massimo Renzo (eds), *Philosophical Foundations of Human Rights* (Oxford University Press 2015); Mary Ann Glendon, *A World Made New: Eleanor Roosevelt and the Universal Declaration of Human Rights* (First edition, Random House Trade Paperbacks 2002); On the contested histories and normative justifications of human rights, see amongst others Makau Mutua, 'The Ideology of Human Rights' (1996) 36 Virginia Journal of International Law 589; Ratna Kapur, *Gender, Alterity and Human Rights : Liberal Freedom in a Fishbowl* (Edward Elgar Publishing 2018); John Tasioulas, 'On the foundations of human rights' in Rowan Cruft, S Matthew Liao, and Massimo Renzo (eds), *Philosophical Foundations of Human Rights* (Oxford University Press 2015); Costas Douzinas, *The end of human rights: Critical thought at the turn of the century* (Bloomsbury Publishing 2000).

[3] Johannes Morsink, 'World war two and the universal declaration' (1993) 15 Human Rights Quarterly 357.

[4] Preamble, Universal Declaration of Human Rights 1948.

[5] See for example Hin-Yan Liu, 'The digital disruption of human rights foundations' in *Human Rights, Digital Society and the Law* (Routledge 2019); Hin-Yan Liu, 'AI challenges and the inadequacy of human rights protections' (2021) 40(1) Criminal Justice Ethics 2; Sue Anne Teo, 'How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework' (2022) 40 Nordic Journal of Human Rights 216; Evelyn Douek, 'The limits of international law in content moderation' (2021) 6 UC Irvine Journal of International, Transnational & Comparative Law 37; Sue Anne Teo, 'Human dignity and AI: Mapping the contours and utility of human dignity in addressing challenges presented by AI' (2023) 15(1) Law, Innovation and Technology 241.

This social ontology of human rights can be intuitively contrasted with the popularised (western) liberal 'individualistic' account that animated the history and theoretical foundation of human rights.[6] In turn, the vernacular of human rights protection is oriented around individual rights.[7] This paper, alongside contemporary case law and scholarship, however, argues that social embeddedness and capacities for positive freedoms within this sociality were themselves implicit normative aims within the rights in the UDHR. This conveys individuals in relations with others in communities, situated within organised societies with corresponding institutions positioned to protect these sets of rights.[8] Juxtaposed against the socially situated individual, AI/ML systems subject individuals to *structural atomisation* in ways that are fundamentally misaligned with the social ontology within international human rights law.

The words '*structural atomisation*' need unpacking. '*Structural*' here denotes a macro-level condition[9] that is neither fleeting nor discrete, operating in ways that reinforce or introduce forms of injustices; for example, the historical and societal injustices seen through the lens of AI as sociotechnical systems. Its structural quality means that it is not easily amenable to discrete and targeted means of problem resolution. '*Atomisation*' in turn is used to express how an individual is *othered*, untethered to embodied and socially embedded contexts that matter for her exercise of positive agency.

Additionally, the phenomenon under exploration, artificial intelligence, itself needs unpacking. The term itself is indeterminate, with points of contention centering around the nature of machine intelligence, as differentiated from biological intelligence[10] and the focus upon its capacity for autonomous actions. The paper adopts a wide conception of the term AI/ML, defining it as computational systems with varying degrees of autonomy that can generate outputs such as predictions, recommendations, decisions and actions that can affect, interface with or learn from physical or virtual environments.[11] Machine learning (ML) is in turn a sub-set of AI and denotes the data-driven self-learning capacities of the abovementioned computational systems. Bringing these two concepts together (AI/ML), this paper seeks to demonstrate that the structural atomisation engendered by AI/ML systems is antithetical to a socially situated individual bearing capacities to exercise positive freedoms. This occurs in three distinct but related ways.

First, data points that group, infer and construct individuals through their likeness, instrumentally atomise individuals as means to ends in ways that are not of their own choosing, through AI/ML systems in which the situated individual has little or no say. Second, individuals risk instrumentalisation through optimisation, wherein the efficiency-driven framing of AI/ML tends to encourage problem solving in ways that respond to computational tractability. Thus, it is not the case that individuals are

[6] The individualistic account is traceable to theories by John Locke and Thomas Hobbes on the 'rights of man' and the American and French revolutions, all of which features as a popularised account of the historical emergence of the modern language of human rights.

[7] Buchanan (n 2) 247. Buchanan argues that the individual emerged as the subject of concern for the international human rights law framework as a response towards radical group inegalitariansim which disregarded the inherent worth of the individual person – *qua* individual; Bart Van der Sloot, 'Is the Human Rights Framework Still Fit for the Big Data Era? A Discussion of the ECtHR's Case Law on Privacy Violations Arising from Surveillance Activities' in Serge Gutwirth, Ronald Leenes, and Paul De Hert (eds), *Data Protection on the Move: Current Developments in ICT and rivacy/Data Protection* (Springer 2016) (where the author highlights that despite the increase in human rights litigation, such as within mass surveillance cases, which do not require the demonstration of impinged individual interest or individual harm, the latter is still the primary requirement which needs to be satisfied in order to make a human rights claim).

[8] Samantha Besson, 'The law in human rights theory' (2013) 7 Zeitschrift für Menschenrechte-Journal for Human Rights 120, 139 (where Besson notes that: 'Human rights are rights individuals have against the political community, *i.e.* against themselves collectively. They generate duties on the part of public authorities not only to protect equal individual interests, but also individuals' political status qua equal political actors').

[9] The definition of structural used here is borrowed from the structural racism discourse. See for example NIMHD, 'Structural Racism and Discrimination' (2022) ⟨https://web.archive.org/web/20220628032157/https://www.nimhd.nih.gov/resources/understanding-health-disparities/srd.html⟩.

[10] Nils J Nilsson, *The quest for artificial intelligence: A History of Ideas and Achievements* (Cambridge University Press 2010) xiii.

[11] This definition draws its inspiration from the OECD definition of AI which was defined as: 'a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.' See OECD, 'The OECD Artificial Intelligence (AI) Principles' ⟨https://oecd.ai/en/ai-principles⟩; UNESCO, 'Recommendation on the Ethics of Artificial Intelligence' (*UNESCO General Conference*, 2021) ⟨https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14⟩ which adopts a similar definition.

instrumentalised by AI/ML systems, it is that they *cannot help but be* instrumentalised when the objective is optimisation. Representation about social and physical phenomena is necessarily flattened, replaced with questions of fair data representation and fairness of AI/ML. This new form of knowledge representation replaces the necessarily contestable spaces of justice and fairness and compacts incommensurable values into computational optimisations. Rights are however, not (straightforwardly) about optimisations. Third, the contextual atomisation through AI/ML mediated shaping of epistemic and enabling conditions can threaten the condition antecedent for socially situated exercise of moral agency and with it, human rights. Such precarity exposes the inadequacy of human rights responses that focus upon harms through its exogenous (as perceivable and observable) typology instead of structural conditions as potential *enablers* of harm.[12]

This paper contributes to the literature by highlighting these three pressure points that *structurally atomise* the individual in ways that are in tension with the social ontology of the human rights law framework. Importantly, this goes beyond the focus of many scholars who highlight the insufficiencies of the individualistic framework of human rights only to foreground the societal impacts of AI/ML, as if these are the only two facets of concern. In diagnosing these structural challenges, it invites reflection on how the design of our computational environments can take the social ontology of human rights more seriously.

The paper consists of two main sections: 'A Social Ontology of Human Rights?' sets the stage of analysis, situating the individual within the contemporary human rights framework as one that is in socially embedded contexts and in relations with others in society with forms of political, economic and social institutionalisation that facilitate the exercise of positive freedoms. 'The Social Ontology of Human Rights and Structural Atomisation Affordances of AI/ML systems' elaborates on the three pressure points posed by the affordances of AI/ML systems that can lead to the *structural atomisation* of individuals, thereby undermining the social ontology of international human rights law. The final section concludes.

A Social Ontology of Human Rights?

To set the stage, the paper departs from the popularised dominant individualist orientation of international human rights law.[13] It argues that a more situated account of the individual in relations with others in society, within communities structured by polity and economic and social institutions, better accounts for existing human rights discourse. To be clear these are not binary positions. As mentioned, the grounds and justifications of human rights have and continue to be, animated by moral, political and philosophical contestation. In fact, it is only within the space afforded by such contestations that the social ontological position, as adopted in this paper, can enter into the conversation at all.

While a deeper analysis of the contested origins[14] of the international human rights law framework is beyond the scope of this paper, the framework can be said to contain key conceptual foundations that define how it is operationalised. Scholarship traces the international institutionalisation of the framework as a response to horrors of the Second World War and the Holocaust.[15] The human rights framework aimed to pushback against the (then) state-based power differential by recognising (and gradually legally bestowing) rights possessed by individuals against the state. In turn, the individual rights orientation of international human rights law has been said to reflect an understanding of a western liberal individual – an autonomous being capable of steering the course of their own life.[16] The role of the state is thus to respect this autonomous individual by not directly interfering and preventing others from interfering in the exercise of his freedoms. This vision draws from the negative liberty tradition

[12] Teo, 'How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework' (n 5).
[13] Buchanan (n 2); Glendon (n 2).
[14] Mutua (n 2); Kapur (n 2); Douzinas (n 2).
[15] Morsink (n 3); See however Samuel Moyn, *The last utopia: human rights in history* (Harvard University Press 2012).
[16] Mutua (n 2).

of political philosophy which can be traced to Berlin,[17] Hobbes and Locke.

Gould's theory on the social ontology provides an alternative account, namely of the socially infused and embedded nature of human rights and its normative aims.[18] Building on her political philosophy in redefining the nature of democracy and freedom, Gould criticised the understanding of political freedom as the mere freedom to act, as an understanding that is empty in substance and fails to take into account the sociality of human life. Freedom requires more than bare unencumbered agential capacities. According to Gould, freedom to act is meaningless without the necessary conditions to realise those very actions.[19]

Thus, as contrasted with theories of negative freedom, Gould's account of freedom is one of positive freedom intertwined in various ways with sociality. This encompasses two key elements. First, in addition to non-interference, the exercise of freedom requires the ensuring of social and material means in which to secure those freedoms. Gould calls these the 'enabling *conditions* of actions.'[20] Secondly, instead of individuals being theorised in the abstract, Gould's account situates an individual's exercise of freedoms as part of self-development. This individual is in turn situated within societies and plays an active role in defining, developing and realising their own self-development. It is argued that 'projects, whether individual or social, normally involve not simply internal self-transformation in a subjective sense but also objective changes in the world in which the agents act to give effect to their purposes.'[21] In turn, the sociality element is pervasive, as individuals' 'self-development depends on these social relations...'[22] This includes the provision of education, employment and other endeavours. The individual however, is neither entirely dictated nor constituted by social elements,[23] but exists in mutually constitutive relationships – thus, the individuality in question is better expressed as one of 'social individuality.'[24] Thus, Gould's theory of the social ontology of human rights is 'in important ways based on sociality and [human rights] are themselves fundamentally social or relational conceptions.'[25]

In addition to the ontological element of sociality, the explication of Gould's theory also bears normative stripes, reminiscent of how human rights take their point of departure in the facticity of human life alongside the normative standards that the framework *aims* to protect and achieve.[26] Although the explication of the theory emphasised the examples of education, employment (including direct decision making by workers in this regard), health and the other human rights by now already recognised within the international framework, her underlying point centers on *conditions* that enable the exercise of positive agency. Gould's social ontology of human rights treats

[17] Isaiah Berlin, 'Two Concepts of Liberty' in *Reading Political Philosophy* (Routledge 2014).

[18] Carol C Gould, 'A Social Ontology of Human Rights' in Rowan Cruft, S Matthew Liao, and Massimo Renzo (eds), *Philosophical Foundations of Human Rights* (Oxford University Press 2015).

[19] Gould's work on political philosophy based upon social ontology builds upon her engagement with Marxist social ontology. However, while Marx emphasised the structural constraints on human action, Gould's work explicates the purposiveness and potentiality of human agency in engaging with, changing and shaping structural conditions of possibilities. See Carol C Gould, *Marx's Social Ontology Individuality and Community in Marx's Theory of Social Reality* (MIT Press 1980).

[20] Carol C Gould, *Rethinking democracy: Freedom and social co-operation in politics, economy, and society* (Cambridge University Press 1990) 35 (emphasis own). Further, according to Gould in page 37, '[o]ne may be free of external constraint by others and yet not be free to realize one's chosen purposes because the necessary conditions or means are not available.' Within human rights and development literature, parallels can be drawn between Gould's position and those of human capabilities scholarship from Amartya Sen and Martha Nussbaum. See Martha C Nussbaum, *Creating capabilities: The human development approach* (Harvard University Press 2011); See also Amartya Sen, 'Human rights and capabilities' (2005) 6(2) Journal of human development 151.

[21] Gould, *Rethinking democracy: Freedom and social co-operation in politics, economy, and society* (n 20) 48.

[22] ibid 49.

[23] ibid 208. In addition to liberal theories grounded upon the 'nature of individuals alone' discounting the social, Gould also discounts 'holism' theories which are said to express 'universal structures of society or on the requirements of the common good.'

[24] ibid 209.

[25] Gould, 'A Social Ontology of Human Rights' (n 18) 177.

[26] Jack Donnelly, *The concept of human rights* (Cornell University Press 2013) 16. Donnelly notes that '[h]uman rights are at once a utopian ideal and a realistic practice for implementing that ideal.'

the exercise of human agency as one of 'individuals-in-relations' and emphasises the 'positive availability… enabling conditions of action [(both material and social)]'.[27] As this paper will demonstrate, where these conditions are disrupted or come under threat by AI/ML systems, this can impact the positive agency that fundamentally intersects with sociality.

In turn, the sociality of human rights is implicitly recognised within international human rights instruments themselves. In the Preamble of the UDHR, there is a recognition of the 'human family', 'to promote social progress' and it also calls for 'every individual and every organ in society' to not only keep the Declaration in mind but also to promote respect for the rights and freedoms therein. Article 1 UDHR speaks of acting 'towards one another in the spirit of brotherhood.' Other articles draw upon social relationships within a societal context and institutional ordering that afford the protection of rights in order to guarantee positive freedoms. Key contemporary rights that come to mind include labour rights, family-related rights, children's rights, political participation and self-determination.[28] According to Beitz: '[t]aken together, these rights are not best interpreted as minimum conditions for any kind of life at all. The rights of the Declaration and the covenants bear on *nearly every dimension of a society's basic institutional structure*, from protections against the misuse of state power to requirements for the political process, health and welfare policy, and levels of compensation for work.'[29]

Further, the recognition of the import of sociality alongside the means to attain forms of self-development in the Gouldian sense is implicit within the right to privacy, even as a classical interpretation of the content of the right treats it as freedom from interference – a right to be let alone.[30] Article 8 of the European Convention on Human Rights (ECHR) has been interpreted as encompassing the meaningful exercise of individual autonomy, including through and with others in a community. In *Breyer*, the European Court of Human Rights (ECtHR) observed that:

> [p]rivate life is a broad term not susceptible to exhaustive definition. Article 8 protects, inter alia, the right to identity and personal development and the right to establish and develop relationships with other human beings and the outside world.[31]

This was similarly held in *A.M-V v. Finland* whereby Article 8 was said to concern 'rights of central importance to the individual's identity, self-determination, physical and moral integrity, maintenance of relationships with others and a settled and secure place in the community.'[32] Private life also includes the right of the individual to a 'private social life,'[33] which cannot be divorced from a social context. It is through the process of dynamic interaction with others, yet with requisite room for seclusion and reflection that an individual's personality is gradually formed (and changes). Cohen in turn notes that a conception of privacy premised upon the separation of self and society is mistaken. She argues that:

> processes of self-development do not conform to the idealized theoretical models preferred by liberal legal theorists […], [s]elfhood is a product of both social shaping and embodied experience. People are born into networks of relationships, practices, and beliefs, and those networks profoundly shape the processes of self-articulation.[34]

[27] Gould, 'A Social Ontology of Human Rights' (n 18) 181.

[28] See for example Article 23 UDHR and Article 11 ECHR on trade unions, Article 12 UDHR and Article 8 ECHR on private and family life, Article 16 UDHR and Article 11 ECHR on the right to found a family and to marry, the Convention on the Rights of the Child 1989, Article 25 ICCPR and Article 3 Protocol 1 ECHR on political participation, Article 27 UDHR on the participation in cultural life, Article 1 ICCPR on the peoples' right to self-determination.

[29] Charles Beitz, 'What human rights mean' (2003) 132(1) Daedalus 36, 39 (emphasis own).

[30] Louis Brandeis and Samuel Warren, 'The Right to Privacy' (1890) IV Harvard Law Review 193.

[31] *Breyer v. Germany,* no. 500001/12, § 73, 30 January 2020. See also *S and Marper v. The United Kingdom* [GC], nos. 30562/04 and 30566/04, §66, ECHR 2008.

[32] A.-M.V. v. Finland, no. 53251/13, § 76, 23 March 2017.

[33] *Bigaeva v. Greece*, no. 26713/05, §§ 22-25 28 May 2009; López Ribalda and Others v. Spain [GC], nos. 1874/13 and 8567/13, §§ 87-88, 17 October 2019.

[34] Julie E Cohen, 'Turning privacy inside out' (2019) 20(1) Theoretical inquiries in law 1, 12.

In turn, identities do not come fully formed but go through socially embodied and socially situated forms of identity building. Since those boundaries constantly shift, along with the constructed subject, the content of what is protected by the *right* to privacy may always fall short of the *concept* of privacy that it seeks to protect.[35]

The social ontology of human rights, premised upon the situatedness of the individual within community, societal and institutional contexts that affords conditions for positive agency, offers a richer understanding of many rights – not as one of contestation but as one of constant negotiation and construction, including through technology.[36]

The Social Ontology of Human Rights and Structural Atomisation Affordances of AI/ML systems

Having accounted for the sociality of human rights within its theory and practice, this section will first address the interplay between the social ontology of human rights and AI/ML systems before highlighting some concerns that arise from AI/ML systems that challenge this account. It identifies three pressure points that the affordances of AI/ML pose towards the social ontology of international human rights law.

First, it goes without saying that technological change brings about social change. Langdon Winner observed that: '[i]ndividual habits, perceptions, concepts of self, ideas of space and time, social relationships and moral and political boundaries have all been powerfully restructured in the course of modern technological development.'[37] At first glance, the liberalisation and democratisation of technology, through its low (and increasingly falling) barriers to accessibility, seem to enable social communication and connection throughout the world. In fact, Facebook's (now META) CEO Mark Zuckerberg acknowledged that the aim of the platform was to make the world 'more open and connected'[38] – thus embracing humanity's needs for bonds of sociality through their instinct to connect. Correspondingly, algorithmically mediated connections and engagements online draw us closer to those who share our interests and afford contextual conditions for self-development in the digital sphere. Communities thus expanded from those situated in close physical proximity to us to encompass bonds of relations with communities around the world. At a higher level, data-driven AI/ML systems that parse through numerosity of data aim precisely to find relevance and connection. Relevance and connection in terms of those who are 'like' us, thus acknowledging at a high level how the individual is embedded within communities and structures of social relationships. Thus, it appears that the values underpinning the social ontology of human rights seem to be reflected and operationalised in the digital context and attuned to the material affordances of AI/ML.

While computational technologies have facilitated connections in seamless and unprecedented ways, the next section discusses the three pressure points posed by the design and deployment of AI/ML systems towards the social ontology of human rights.

The Instrumental Atomisation of the Individual

First, the structural atomisation of AI/ML instrumentally atomises the individual. The data-driven capacity of AI/ML systems rely upon data, both personal or non-personal, to make predictions. In turn, data that is gathered reveals insights through correlations and inferences gleaned from machine learning that may not be obvious or observable to fallible human faculties. Where these predictions are then personalised to the individual, the process happens through a paradox – in that, 'in order to apprehend users as individuals, platforms must first assemble

[35] Mireille Hildebrandt, 'Privacy and Identity' in Erik Claes, Antony Duff, and Serge Gutwirth (eds), *Privacy and the Criminal Law* (Intersentia 2006) 4.
[36] Sheila Jasanoff, *Science at the bar: Law, science, and technology in America* (vol 9, Harvard University Press 1997).
[37] Langdon Winner, *The whale and the reactor: A search for limits in an age of high technology* (University of Chicago Press 2020) 9–10.
[38] Mark Zuckerberg, 'Bringing the World Closer Together' (2021) ⟨https://www.facebook.com/notes/393134628500376/⟩.
[39] Thao Phan and Scott Wark (2021) ⟨https://culturemachine.net/vol-20-machine-intelligences/what-personalisation-can-do-for-you-or-how-to-do-racial-discrimination-without-race-thao-phan-scott-wark/⟩.

them into groups based on their likenesses with other individuals.'[39] This process of assembling into groups through data aggregation can be specified as a feature selection within a model, but can also occur outside of the knowledge of not only the individual, but also of those deploying or designing the AI/ML system.[40] These data groupings and correlations are considered advantages of AI/ML systems – revealing actionable insights that have driven advances within fields as diverse as science, healthcare, transportation and smart cities.[41]

However, as Van der Sloot writes, 'these groups are not stable, but fluid and not unique or sparse, but omnipresent and widespread. Group profiles may be created in a split second, they may be used by all kinds of organisations and institutions and they may change by altering the determinants and criteria according to new insights or needs, so that who is part of a group profile and who is not may change every day, or even more often.'[42] The individual can be grouped in ways where they might disagree with, in a process in which they have no say. Thus, the deployment of AI/ML systems within areas tightly bound to moral considerations of autonomy should give pause, for it displaces the very subject it seeks to know. The concept of autonomy, as shown through the expansive interpretation undertaken by ECHR in the Article 8 case law, is concerned with the meaningful *exercise* of autonomy. Furthermore, the exercise of this autonomy is situational and contextual, embedded within social contexts and relationships.[43] The exercise of autonomy involves the individual having a say in the contextual environment, in terms of how one is read and modulated, even and especially when mediation and modulation of the digital environment is increasingly pervasive and persistent.[44] The political economy of the commodification of data and human experiences can further entrench the displacement of these individual capacities.

Thus, for example, the deployment of problematic forms of 'emotional AI' which ostensibly gauge emotions from faces and the increasing use of surveillance technologies such as facial recognition systems, displace and instrumentalise the individual, in addition to potential chilling effects they can engender.[45] The instrumental atomisation impacts the ways in which an individual can 'self-present', referring to 'the ability of individuals to present multifaceted versions of themselves and thus behave differently depending on the circumstances.'[46] The contextual boundaries allow for an individual to tailor behaviours according to the accepted social and behavioural norms bounded within those contexts. The modulation of the individual from correlated data, gleaned from one context into another context, alters these boundaries and with it, the self-presentation that is an exercise of one's autonomy.[47]

Additionally, instrumental atomisation of the individual also occurs through personalised choice architectures, for example, through social media platforms that display dy-

[40] Sandra Wachter, 'The theory of artificial immutability: Protecting algorithmic groups under anti-discrimination law' (2022) 97 Tulane Law Revuew 149.

[41] Oscar H Gandy, 'Statistical surveillance: Remote sensing in the digital age' in David Lyon, Kevin D Haggerty, and Kirstie Ball (eds), *Routledge handbook of surveillance studies* (Routledge 2012); Rose Yu and others, 'Deep learning: A generic approach for extreme condition traffic forecasting' [2017] Proceedings of the 2017 SIAM international Conference on Data Mining 777.

[42] Bart van der Sloot, 'Do groups have a right to protect their group interest in privacy and should they? Peeling the onion of rights and interests protected under article 8 ECHR' in Linnet Taylor, Luciano Floridi, and Bart Van der Sloot (eds), *Group privacy: new challenges of data technologies* (Springer 2017) 220.

[43] Cohen, 'Turning privacy inside out' (n 34); Julie E Cohen, 'What privacy is for' (2012) 126 Harvard Law Review 1904.

[44] Mireille Hildebrandt, *Smart technologies and the end (s) of law: novel entanglements of law and technology* (Edward Elgar Publishing 2015) 102–103.

[45] See for example Michal Kosinski, 'Facial recognition technology can expose political orientation from naturalistic facial images' (2021) 11(1) Scientific reports 100; See also Douglas Heaven, 'Why faces don't always tell the truth about feelings' (2020) 578(7796) Nature 502; Rajeev Ranjan and others, 'Deep learning for understanding faces: Machines may be just as good, or better, than humans' (2018) 35(1) IEEE Signal Processing Magazine 66.

[46] Peggy Valcke, Damian Clifford, and Viltė Kristina Dessers, 'Constitutional Challenges in the Emotional AI Era' in Amnon Reichman and others (eds), *Constitutional Challenges in the Algorithmic Society* (Cambridge University Press 2021); See also Julie E Cohen, *Configuring the Networked Self: Law, Code, and the Play of Everyday Practice* (Yale University Press 2012); Erving Goffman, *The Presentation of Self in Everyday Life* (Anchor Books 1990).

[47] See generally Helen Nissenbaum, 'Privacy as contextual integrity' (2004) 79 Washington Law Review 119; Maria Brincker, 'Privacy in public and the contextual conditions of agency' in *Privacy in public space* (Edward Elgar Publishing 2017).

[48] Taina Bucher, 'Want to be on the top? Algorithmic power and the threat of invisibility on Facebook' (2012) 14(7) New media & society 1164.

namic content tailored to each individual but also through default options.[48] Sætra observes that:

> Technology is constantly used to rearrange the 'doors' we encounter online, and to manipulate us into changing our behaviour when we choose which products to purchase, ads to click, new articles to read, and people to befriend. The doors of our lives are constantly rearranged both by algorithms set free to maximize certain variables, or by people intentionally rearranging them to elicit a particular behaviour through, for example using default options, framing, or providing strategic anchors.[49]

Behavioural psychology studies have established that framing options affect behaviours.[50] In the parlance of human rights and in considering whether a harm suffered amounts to a violation, the dynamic creation of such personalised choice architectures also make it difficult for the individual to mount a claim of harm when comparator experiences are not available.[51] Using content moderation in social media platforms as an example, it can be difficult to tell if content removed on one account is also removed on another and if so, whether contextualisation mattered in the equation. This is due to the fact that such experiences take place in personalised silos. Thus, measures of deviations are rendered much more onerous or impossible, essentially neutralising the utility of human rights. Manipulation has been raised as a possible objection as nudges through personalised choice architectures do not engage the rational agency possessed by individuals, but instead manipulate the sub-conscious into action.[52] Others argue that the economic logic of big data relies on emotional manipulation.[53] However, at the same time, Sætra argues that psychological interference is relational and situational and an appeal to protections based upon the 'average' person will likely not suffice.[54]

This first step shows that by instrumentally atomising the individual, the latter is deprived of a say in the exercise of positive agency that underpins a socially embedded existence. A counterargument could be offered however. Instead of instrumental atomisation, it may be that the data-driven mediation that characterises modern digital lives is thoroughly social as big data is socially produced, relational and anything but individual.[55] In turn, the social infusion comes from the application of data that AI/ML systems adjudge to be of relevance or likeness to us – it is sociality and relationality made possible and in certain cases, imposed, albeit through computational means. Does the difference of ontological substrates (i.e. biological versus computational) matter?

The normative expectations protected through human rights however are not only concerned with sociality *tout court*. Were this the case, regimes that brainwash and impose particular forms of opinions or beliefs on individuals would easily pass muster. Instead, Gould's theory and the contemporary expression of autonomy found in the Article 8 ECHR and human dignity literature is about fundamentally *engaging* the individual in defining and exercising her positive agency within a socially informed existence. The active role of the individual in doing so can be contrasted with the passive reception of sociality this objection forwards. Seen this way, the instrumental atomisation point still stands. The individual is instrumentalised through datafication, taking the individual away from the situated socialisation that allows for her positive engagement into a receptacle of imposed sociality, of those of ostensible likeness to her.[56]

Thus, AI/ML systems can operate in ways that divorce the individual not only from their social situatedness and its enabling conditions for positive agency but also from modulations that individuals should, as a matter of a human

[49] Henrik Skaug Sætra and Stuart Mills, 'Psychological interference, liberty and technology' (2022) 69 Technology in Society 101973, 6.

[50] Amos Tversky and Daniel Kahneman, 'The framing of decisions and the psychology of choice' (1981) 211(4481) Science 453.

[51] Wachter (n 40).

[52] Daniel Susser, Beate Roessler, and Helen Nissenbaum, 'Online manipulation: Hidden influences in a digital world' (2019) 4 Georgia Law Technology Review 1.

[53] Marion Fourcade and Daniel N Kluttz, 'A Maussian bargain: Accumulation by gift in the digital economy' (2020) 7 Big Data & Society 1.

[54] Sætra and Mills (n 49).

[55] I thank the anonymous reviewer who raised this astute point.

[56] See generally Ian Hacking, 'Making Up People' in TL Heller, M Sosna, and DE Wellbery (eds), *Reconstructing Individualism* (Stanford University Press 1985); Geoffrey C Bowker and Susan Leigh Star, *Sorting things out: Classification and its consequences* (MIT press 2000); Wachter (n 40).

right, be able to co-determine. The contention is not that the structural atomisation is wrong because it gets things wrong in the process. It is wrong because it constrains the individual's ability to reinvent and change themselves in ways that comport with a socially situated exercise of autonomy. Doing so in ways that make it difficult for these modulations to be challenged ex post, due to lack of socially salient comparator experiences, further compounds the problem. As we shall see in the next section, instrumental atomisation of the individual also occurs through the practice of computational law.[57]

Instrumentalisation through Optimisation

The second pressure point identified from the structural atomisation of AI/ML versus the social ontology of human rights is that the individual is instrumentalised through optimisation. Whether one takes as a starting point that AI/ML systems are systems displaying intelligent behaviour[58] or that they are taking rational (as opposed to random) actions,[59] the underlying purpose of the design and deployment of AI/ML systems orient around the optimisation of a given goal or performance metric.[60] The optimisation of goals or performance metrics brings about immense benefits in relation to efficiency, cost reduction and in streamlining and promoting consistency in diverse public and private sector endeavours. These alleged benefits in turn straightforwardly enhance the access to and exercise of human rights. This process of optimisation involves a certain level of abstraction, abstracting away unnecessary noise in relation to the task at hand, and utilising data that is both considered relevant and computationally tractable. Selbst notes that:

> Desirable properties of a system can then be described in terms of inputs and outputs alone: the internals of the system and the provenance of the inputs and outputs have been abstracted away.[61]

This decoupling from context similarly applies to the data which is extracted to train the AI/ML models. It is said that every single data set consists of some 'intrinsic, hidden, not yet unearthed value, and companies are engaged in a race to discover how to capture and rate this value.'[62]

In relation to optimisations, private sector use of AI/ML such as within social media platforms that are reliant on engagement and growth metrics wield algorithmic power over interactions and individuals on the platform.[63] AI/ML systems deployed in the public sector may, on the other hand, be driven by societally desirable goals such as the attainment of fairer outcomes, more equal distribution of opportunities and resources as well as cost and time savings associated with efficiency. Many public sector domains demand swift and accurate responses that cater to the overall public good. Such benefits are also a key feature of the rule of law and access to justice – the old adage of justice delayed is justice denied – and they can similarly be applied to justify the attraction of computational law.

However, the optimisation logics applied to public administration are problematic on several fronts. They can potentially infringe on the moral reasoning, reflected in administrative law and through rule of law principles on the need to state reasons – thereby requiring transparency and accountability over public-facing decision making.[64] In rendering the individual computationally legible for purposes of optimisation, the optimisation necessarily instru-

[57] I relate to computational law in this paper to encompass the prospects or practices of data-driven law. See COHUBICOL, 'Data-driven law' ⟨https://www.cohubicol.com/about/data-driven-law/⟩.

[58] Nilsson (n 10) xiii.

[59] Stuart J Russell and Peter Norvig, *Artificial intelligence: a modern approach* (Third edition, Prentice Hall 2010).

[60] Andrew D Selbst and others, 'Fairness and abstraction in sociotechnical systems' in *Proceedings of the conference on fairness, accountability, and transparency* (2019).

[61] ibid.

[62] José Van Dijck, 'Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology' (2014) 12(2) Surveillance & society 197 (quoting Mayer-Schoenberger and Cukier).

[63] Bucher (n 48); Rasmus Kleis Nielsen and Sarah Anne Ganter, *The power of platforms: Shaping media and society* (Oxford University Press 2022).

[64] Karen Yeung, 'Algorithmic regulation: A critical interrogation' (2018) 12(4) Regulation & governance 505; Jeremy Waldron, 'Theoretical Foundations of Liberalism' (1987) 37 The Philosophical Quarterly 127.

mentalises them in ways that treat individuals systemically as a means to an end.

The example of an algorithm used in Spain to assess the risk of the violence that women could face when they report cases of abuse revealed curiosities such as getting lower scores on account of not having children in the household. This, however, does not reveal whether or not the woman in question is at an (imminent) threat of violence.[65] An algorithmic review undertaken showed that only a small minority of women received scores that are 'medium' or higher which would qualify them for police protection.[66] Even though the system used is that of a classical statistical model, the point about computational tractability and instrumentalisation remains. Here, the social phenomena of violence is reduced and flattened to scores traced to arbitrary proxies such as children in the household. This indicates that what was intended for the system (i.e. assessing potential risk of violence) versus what is measured and optimised (ostensibly expending state resources on abuse prevention) were out of step as violence is not a straightforward computationally tractable phenomena. This way, the individual in question is deprived of the capacity to articulate, make sense of and understand the circumstances of her own life situation.

A second example in relation to optimisation through computational tractability relates to the use of SyRI, a social-welfare fraud detection algorithm used in the Netherlands. The use of SyRI has been found to fall foul of the respect for privacy[67] but also as lacking transparency due to the black box nature of the algorithm.[68] The former UN Special Rapporteur on Extreme Poverty, Philip Alston, cautioned against 'stumbling, zombie-like, into a digital welfare dystopia' whereby small technical irregularities through, for example, data-mismatching, can bring about serious negative impacts.[69] The deployment of the algorithm saw 'fraud suspicion' being measured through supposedly innocuous acts such as missing medical appointments and water usage levels. The report highlighted that the rigidity engendered by strict algorithmic accounting fails to take into account 'extenuating circumstances such as being late for an appointment because of urgent care obligations, or being unable to understand a written communication due to a disability or a personal crisis.'[70] Similar examples are now abundant.[71] Thus, key public administration services such as social welfare provisions reduce measurability of trust and entitlement through computational legibility instead of putting the individuals affected at the front and center. This is an ironic outcome considering that few social provisions matter more for the individual than the safety nets provided by social welfare.

An objection may however be offered. Is the computational logic that enables optimisation, including that of public services, really the culprit here? Examples abound on the biases, prejudices, inconsistencies and downright discriminatory effect of human decision-making within public administration.[72] Huq argues that with proper safeguards in place, the benefits of machine decision making can far outweigh human decision making[73] and notes that a well-calibrated machine decision' marries both worlds – bringing fairness and accountability. The purpose of deploying AI/ML systems then pertains not merely to the optimisation of time and cost but also due to their ostensibly fairer means of recommendation and decision making.[74]

[65] Eticas, The External Audit of the VioGén System, 'Association Eticas Research and Innovation' (2022).

[66] A lack of human oversight and an over-reliance on the risk scores generated were other reasons that were cited in the Eticas evaluation report on the VioGen System.

[67] NJCM et al. v The Dutch State (2020) The Hague District Court ECLI: NL: RBDHA:2020:1878 (*SyRI*).

[68] AlgorithmWatch, 'How Dutch Activists Got an Invasive Fraud Detection Algorithm Banned' (2020) ⟨https://algorithmwatch.org/en/syri-netherlands-algorithm/⟩.

[69] Philip Alston, 'Report of the Special Rapporteur on Extreme Poverty and Human Rights on Digital Welfare States' (2019) ⟨https://documents-dds-ny.un.org/doc/UNDOC/GEN/N19/312/13/PDF/N1931213.pdf?OpenElement⟩ A/74/493 para 77.

[70] Hacking (n 56).

[71] Eva Constantaras and others, 'Inside the Suspicion Machine' (*Wired*, 2023) ⟨https://www.wired.com/story/welfare-state-algorithms/⟩.

[72] Daniel Kahneman, Olivier Sibony, and Cass R Sunstein, *Noise: A flaw in human judgment* (Brown Spark 2021); Cass R Sunstein, 'Algorithms, correcting biases' (2019) 86(2) Social Research: An International Quarterly 499.

[73] Aziz Z Huq, 'A right to a human decision' (2020) 106 Virginia Law Review 611.

[74] Kahneman, Sibony, and Sunstein (n 72).

An algorithm has no personal or socialised prejudice after all, hidden or otherwise. Thus, Langford's argument that '[d]iscussions of automation and digitalization should be guided by a logic of minimizing danger, regardless of whether its origin is machine or human' is both relevant and on point.[75]

To this, a two-pronged response is offered. First, I agree with such critique that calls for abandoning a techno-determinist stance that focuses upon the technology in question without situating its development and deployment within a sociotechnical setting. Science and Technology scholars have cautioned against the myopia of fixating upon artefacts to the exclusion of their social and political impact and influences. Thus, the 'digital welfare state' that Alston fears did not emerge out of nowhere. The AI technologies that facilitate it are but one of the manifestations of a neo-liberalist governmentality that has steadily shifted the role of the state from being guardians and providers of economic, social and cultural rights protection into one that individuates risks and reverses the burden of proof in relation to claims.[76] However, instead of critiquing the technology in question, the human rights lens offers a set of conditions wherein human flourishing is enabled rather than constrained. From this vantage point, the criticism of human rights as necessarily being dependent upon the technological artefact is a misdirected one. It matters not that smart technologies enter into the fray. Instead, when technologies enable the granular surveillance and the reading of individuals in such a way that they have no corresponding means to know, let alone contest, the way AI/ML systems frame them, the core of the social ontology of human rights is engaged. In other words, the danger here pertains not per se to harms that occasion, which individual discrete rights (e.g. non-discrimination, freedom of expression etc.) can largely adequately address, but towards the unravelling of conditions that allow the individual to make sense of their social environment and therein exercise positive agency.

Second, while a technological determinist approach that is focused on the artefact should be avoided, this does not negate the fact that AI/ML systems enable particular affordances that did not exist before (or not at the same scale) on account of purported values of neutrality and objectivity.[77] Hong critiques the way in which the ostensible neutrality and objectivity of data conveys power upon the party deploying algorithmic predictions while at the same time depriving targeted individuals of the power that comes from understanding their socially situated condition.

> When predictive models are baked into decision-making processes, they do not simply shift the entire apparatus towards inhuman objectivity, but rather empower new norms on who gets to impose their discretion upon whom. Often, the promise of data as a universal illuminator conceals the reality in which it is data for me, and not for thee.[78]

He continues: '[t]his deprivation of someone's ability to anticipate, plan, and adjust their own conditions of working and living is but one episode in which making labour more predictable for some requires making it less predictable for others.'[79]

Going back to examples of computational optimisation, where the computational logics of optimisation hold sway, the converse is also evident in how issues of unfairness within AI/ML tools are addressed. The complexity of so-

[75] Malcolm Langford, 'Taming the digital leviathan: Automated decision-making and international human rights' [2020] , 114 AJIL Unbound 141, 145; The focus on actual harms can also be motivated by the field of computational phenomenology, which demonstrates that the human brain (and the subjective experiences therein) can be computationally modeled. See for example Pierre Beckmann, Guillaume Köstner, and Inês Hipólito, 'An alternative to cognitivism: computational phenomenology for deep learning' (2023) 33(3) Minds and Machines 397.

[76] Lina Dencik, 'The datafied welfare state: A perspective from the UK' in Andreas Hepp, Juliane Jarke, and Leif Kramp (eds), *New perspectives in critical data studies: The ambivalences of data power* (Springer International Publishing Cham 2022); Luke Henriques-Gomes, 'Robodebt Class Action: Coalition Agrees to Pay $1.2bn to Settle Lawsuit' (*The Guardian*, 2020) ⟨https://www.theguardian.com/australia-news/2020/nov/16/robodebt-class-action-coalition-agrees-to-pay-12bn-to-settle-lawsuit⟩.

[77] Ben Green and Salomé Viljoen, 'Algorithmic realism: expanding the boundaries of algorithmic thought' in *Proceedings of the 2020 conference on fairness, accountability, and transparency* (2020); See generally Theodore M Porter, 'Trust in numbers: The pursuit of objectivity in science and public life' [2020] .

[78] Sun-ha Hong, 'Prediction as extraction of discretion' (2023) 10(1) Big Data & Society 20539517231171053.

[79] ibid 7.

cial phenomena, imbued with the messiness of social and moral contestation, is replaced with questions of fair data representation and fairness of AI/ML. This new form of knowledge representation replaces the necessarily contestable spaces of justice and fairness into questions of machine fairness, compacting incommensurable values into computational optimisations. In fact, research has shown that even within technical domains of fairness, contesting notions of what counts as fairness in specific instances might yield mutually incompatible solutions.[80] Furthermore, questions of fairness are not readily amenable to surgically precise adjustments as it arises not only within outcomes but also as embedded within processes (design choices, optimisation parameters etc) in the AI/ML pipeline.[81] The measurements of efficiency and demands for accuracy of AI/ML assume that the data representing a phenomenon counts as ground truth, and that the phenomenon is a desirable state of affairs worth replicating and measuring against.[82] Privileging accuracy (measured through ground truth) over questions of social justice risks cementing implicit relationships of power – including that of racial inequality and gender representation. Christianini agrees: the 'accuracy of predictions has been privileged over other aspects of knowledge, in this way subtly redefining what knowledge and models are for.'[83] Elsewhere, a narrow focus on metrics of fairness and accuracy obscures larger concerns of whether or not AI/ML systems should be deployed in the first place.[84]

Can human rights then come to the rescue? When juxtaposed against the scale and speed of AI/ML and the optimisation based upon efficiency logic, human rights may fail on several fronts. In order to rely on protections afforded under international human rights law, the individual claiming human rights accountability on account of a detriment suffered, needs to, obviously, know that a detriment has been experienced and that this detriment was non-trivial.[85] It necessarily locates the source of harm as one that is discrete, a deviation from an otherwise acceptable state of affairs. However, for certain types of detriment, such as privacy harms, the implicit state of affairs assumed might no longer hold. On account of the ubiquitous use of AI/ML tools and the mediation role it plays in human perceptions and experiences, it is not only impracticable to expect individuals to navigate this terrain – for example through notice and consent of privacy provisions, but outright impossible precisely due to its widespread yet paradoxical invisibility.[86] Even if one is able to overcome these systemic disaffordances, it can be extremely hard to gain insight in a given system's interdependence and complexity.[87] Can human rights then play a role upstream and be part of the optimisation process? That is to say, can human rights to be factored into the design of such AI/ML systems?

To this, the first response is that, unlike AI/ML, human rights do not scale. As an example, human rights cannot scale to properly address issues of hate speech on social media platforms.[88] For the latter, the prohibition of ad-

[80] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan, 'Inherent trade-offs in the fair determination of risk scores' [2016] arXiv preprint arXiv:1609.05807; see however Reuben Binns, 'On the apparent conflict between individual and group fairness' in *Proceedings of the 2020 conference on fairness, accountability, and transparency* (2020).

[81] David Lehr and Paul Ohm, 'Playing with the data: what legal scholars should learn about machine learning' (2017) 51 UCDL Review 653.

[82] Sandra Wachter, Brent Mittelstadt, and Chris Russell, 'Bias preservation in machine learning: the legality of fairness metrics under EU non-discrimination law' (2020) 123 West Virginia Law Review 735.

[83] Nello Cristianini, 'Shortcuts to Artificial Intelligence' in Marcello Pelillo and Teresa Scantamburlo (eds), *Machines We Trust: Perspectives on Dependable AI* (MIT Press 2021).

[84] See for example NBC News, 'Microsoft Is Removing Emotion Recognition Features from Its Facial Recognition Tech' ⟨https://www.nbcnews.com/tech/tech-news/microsoft-removing-emotion-recognition-features-facial-recognition-tec-rcna35087⟩.

[85] Wachter (n 40); Janneke Gerards and Frederik Zuiderveen Borgesius, 'Protected grounds and the system of non-discrimination law in the context of algorithmic decision-making and artificial intelligence' (2022) 20 Colorado Technology Law Journal 1. See also the admissibility requirements under the ECHR, Art 35.

[86] Yoni Van Den Eede, 'In between us: On the transparency and opacity of technological mediation' (2011) 16 Foundations of Science 139; Cohen, 'What privacy is for' (n 43).

[87] Van Dijck (n 62); Woodrow Hartzog and Evan Selinger, 'Big data in small hands' (2013) 66 Stanford Law Review 81; Astrid Mager, 'Algorithmic ideology: How capitalist society shapes search engines' (2012) 15(5) Information, Communication & Society 769.

[88] Douek (n 5).

vocacy of national, racial or religious hatred (the Rabat Plan of Action) requires an examination of six factors to determine whether or not hate speech has taken place – these include the need for contextualisation and looking at the intent of the speaker. Achieving this at scale is extremely difficult with AI mediated content moderation and curation.[89] Contextualisation necessarily demands individual assessment, taking into account embedded circumstances. The scaled application of AI/ML in the mediation of online speech is ontologically different and brings forth a different set of concerns. These concerns focus, for example, on acceptable error rates, types of errors (false positives/negatives) to minimise and calls for transparency, instead of the accountability framework of human rights.[90] Douek notes that 'IHRL does not speak in these terms of error choice, but rather in the language of individual cases. Practically, however, these considerations pervade platform decision-making.'[91] On this note, even as we acknowledge that measures to design for rights at the outset, such as 'privacy by design,'[92] are laudable and gaining traction, such measures function in a less optimal manner when contestations occur. As the hate speech example has shown, contestations lie at the heart of human rights operationalisation.

Secondly, human rights have also been said to be about optimising for outcomes. The 'necessity and proportionality' balancing processes involved in gauging whether or not certain rights are infringed when weighed against other interests – national security, public health, public morals etc., engage optimisation parameters. The optimal outcome is one that balances the rights of one person, or a class of persons against that of a *larger* interest. Verdirame called this a 'category error'[93] as it necessarily reduces human rights to a utilitarian (and hence optimisation) concern, displacing the primacy of the individual as the primary subject of concern in international human rights law. Yet, one could argue that human rights is not straightforwardly about optimisations. The balancing processes hinge upon the (democratic) legitimacy of the state – which in turn imbues this balancing process with legitimacy.[94] The state has both the economic resources and political legitimacy for the proper exercise of taking the interests of public morality, public health and public security, amongst others, into consideration. The same cannot be said for private interests that continue to monopolise the development and deployment of AI/ML systems. Mixed into this concoction is the private sector's fiduciary shareholder interests focused on growth and profits. Although most evidently traced to the commercial motivations of 'Big Tech', similar concerns permeate through private sector technologies deployed within the public sector.

This section has exposed twin weaknesses. On the one hand, it has shown that the instrumental optimisation of the individual through the efficiency framing, in ways that are computationally tractable, structurally atomises the individual from the exercise of positive agency, undergirded by the social ontology of human rights. On the other hand, incongruence still exists between the protection afforded by international human rights law oriented around the language of individual harms and the sociality that orients human rights. An individual who has been instrumentalised through optimisation afforded by AI/ML within sociotechnical settings does not engage the vernacular of human rights violation, because such optimisation does not *appear* as a typical human rights harm in the first place. Furthermore, the vernacular of human rights viola-

[89] Douek (n 5) 67–68.

[90] The EU's Digital Services Act, adopted in 2022 to regulate digital intermediaries, implicitly acknowledges the difficulties raised here as the focus of the Regulation is upon transparency, risk management and reporting duties rather than the substance of content moderation.

[91] Douek (n 5) 70.

[92] Ann Cavoukian, 'Operationalizing Privacy by Design: A Guide to Implementing Strong Privacy Practices' (*Information and Privacy Commissioner*, 2012) ⟨https://gpsbydesigncentre.com/wp-content/uploads/2021/08/Doc-5-Operationalizing-pbd-guide.pdf⟩.

[93] Guglielmo Verdirame, 'Rescuing human rights from proportionality' in Rowan Cruft, S Matthew Liao, and Massimo Renzo (eds), *Philosophical Foundations of Human Rights* (Oxford University Press 2015).

[94] Fabienne Peter, 'Political legitimacy' (2010) ⟨https://plato.stanford.edu/archives/spr2014/entries/legitimacy/⟩.

[95] Douek (n 5); Brenda Dvoskin, 'International Human Rights Law Is Not Enough to Fix Content Moderation's Legitimacy Crisis' (*Berkman Klein Center Collection*, 2020) ⟨https://medium.com/berkman-klein-center/international-human-rights-law-is-not-enough-to-fix-content-moderations-legitimacy-crisis-a80e3ed9abbd⟩; Rachel Griffin, 'Rethinking Rights in Social Media Governance' (*Verfassungsblog*, 2022) ⟨https://verfassungsblog.de/rethinking-rights⟩.

tion seems structurally unable to address the paradoxical problem of the dynamism yet pervasiveness and the siloing at scale engendered by AI/ML systems,[95] revealing that perhaps that the ontologically different underlying affordances of law and computational systems such as AI/ML do not match.[96] The design of human rights responds to the language of violations towards the individual, less well to optimisations and conditions of subterfuge and obfuscation afforded by AI/ML.

Contextual Atomisation

The final point of tension between the sociality acknowledged within international human rights law and AI/ML pertain to contextual atomisation, namely through the epistemic and contextual precarity afforded by AI/ML. The contextual atomisation through AI/ML mediated shaping of epistemic and enabling conditions can threaten the condition antecedent of a socially situated exercise of moral agency and with it, human rights. In the example of deepfakes, Rini writes that the advent of such technologies threaten the epistemic backstop which has been accepted as a barometer of truth and falsity. This privilege was previously held by ironically, video and audio technologies, both of which were accepted as marks of the veracity of testimony.[97] The displacement of these technological stamps of veracity muddles the epistemic conditions into one of contextual precarity and uncertainty. It is not truth or falsity that is at issue here, as epistemic precarity goes beyond deepfakes convincing anyone that the deepfaked audio or video is real. Instead, deepfakes are said to cause a 'sense of displaced epistemic reality'.[98] This challenge is arguably amplified in the context of generative AI models – encompassing multimodal audio, video and text generation.[99] Citron and Chesney have offered numerous examples of how deepfakes can threaten social and political discourse, such as a deepfake video showing a gun control advocate allegedly tearing up a copy of the US Constitution which was in fact an image of a bullseye. In relation to politically charged issues such as gun control in the US, the contextual precarity brought forth by deepfakes can have severe political consequences.[100] The same goes for deepfakes used during politically charged periods such as elections where timing is of the essence.[101] The amount of time it takes for deepfakes to be refuted would mean that political fraught lines are drawn even before verification can take place. Further, Rini acknowledges the problem of social epistemology in that fake news tend to have a wider and faster reach compared to subsequent corrections.[102]

As mentioned, epistemic and contextual precarity is not limited to deepfakes. The conceptual point can similarly be made by the example of content modulation and amplification on social media platforms. During the initial stages of the Covid-19 pandemic, misinformation surrounding the vaccine and the virus was rife, and amplification of such content introduced widespread epistemic uncertainty (which was no doubt amplified by the lack of settled scientific evidence at that time). Once again, there is no need prove that amplification of misinformation led to reduced rates of vaccination but rather the danger lies in the mere fact of algorithmic amplification that introduces precarity into epistemic conditions. Content modulation has also generally been said to contribute towards echo chambers and epistemic bubbles, influencing the types of information one encounters online.[103] Although research point in opposite directions on the extent of the dangers

[96] Mireille Hildebrandt, *Law for computer scientists and other folk* (Oxford University Press 2020) ch 11.

[97] Regina Rini, 'Deepfakes and the epistemic backstop' (2020) 20 Philosophers' Imprint 1.

[98] ibid 8.

[99] See for example Dani Di Placido, 'Why Did "Balenciaga Pope" Go Viral?' (*Forbes*, 2023) ⟨https://www.forbes.com/sites/danidiplacido/2023/03/27/why-did-balenciaga-pope-go-viral/⟩; Abené Clayton, 'Fake AI-Generated Image of Explosion near Pentagon Spreads on Social Media' (*The Guardian*, 2023) ⟨https://www.theguardian.com/technology/2023/may/22/pentagon-ai-generated-image-explosion⟩.

[100] Bobby Chesney and Danielle Citron, 'Deep fakes: A looming challenge for privacy, democracy, and national security' (2019) 107 California Law Review 1753.

[101] Maja Brkan, 'Artificial intelligence and democracy: The impact of disinformation, social bots and political targeting' (2019) 2 Delphi – Interdisciplinary Review of Emerging Technologies 66.

[102] Rini (n 97) 7.

[103] C Thi Nguyen, 'Echo chambers and epistemic bubbles' (2020) 17(2) Episteme 141.

of echo chambers and filter bubbles,[104] it is important to stress that it is not per se its effects in engendering negative outcomes that this section aims to highlight. Instead, the wrong lies in the precarious background conditions through a splintering of a common context for the exercise of moral agency within social settings. Under such conditions, the very nature of what counts as true and what is considered as a shared social space is up for grabs. It is this element of contextual precarity that atomises the individual from contextual sense-making.

Contextual precarity can also be affected through the practice of computational law. Building on section 2.2 and Huq's contention that human decision making has no antecedent superiority to its machine counterpart, Kahneman et. al. also demonstrated that human decision making is fraught with noise, leading to inconsistent decision making within consequential areas such as asylum determinations.[105] Consequently, relying on AI/ML systems in turn promises consistency and reliability. Leaving aside the question of whether the claims of the potential of AI/ML systems have been matched in practice,[106] this section looks instead at the second order problem, that is at the *value substitution*. The value of trust (towards citizens as moral agents with the capacity to understand and self-apply the law)[107] is replaced by the value of the *perceived* reliability of AI systems.[108] Winsberg speaks of this as 'reliability without truth.'[109] The values of trust and self-reliance respects the moral agency of individuals that underpins the rule of law. Raz observed that:

> A legal system which does in general observe the rule of law treats people as persons at least in the sense that it attempts to guide their behaviour through affecting the circumstances of their action. It thus presupposes that they are rational autonomous creatures and attempts to affect their actions and habits by affecting their deliberations.[110]

Thus, where the perceived reliability of AI/ML systems displaces the need to rely on the (unpredictability) of trusting persons as moral agents who are addressees of law, this can at the same token deprive the individual of knowledge pertaining to the application of the law (to them) and obscures the ability of the individual to express their own afflictions. This can complicate navigating the demands of the law or complicate matters for those seeking to contest its increasingly computational form.

These points demonstrate that epistemic and contextual precarity cuts into the conceptual and normative foundations of human rights. Human rights are conceptually deficient because accountability for human rights wrongs is through the occurrence of a breach of discrete rights. This exposes the inadequacy of human rights responses that focus upon harms through its exogenous (perceivable and observable) typology instead of focusing on structural conditions as *enablers* of harm.[111] Harm transforms itself from a source-based concern (be it technological source or a moral agent as the source) amenable to targeted redress to one in which the very contextual conditions become precarious. Additionally, this precarity also challenges the

[104] On studies indicating that social media does not lead to polarization, see Maria Nordbrandt, 'Affective polarization in the digital age: Testing the direction of the relationship between social media and users' feelings for out-group parties' (2023) 25(12) New media & society 3392; Levi Boxell, Matthew Gentzkow, and Jesse M Shapiro, 'Greater Internet use is not associated with faster growth in political polarization among US demographic groups' (2017) 114(40) Proceedings of the National Academy of Sciences 10612; For opposing views, see Jonathan Haidt, 'More Social Media Regulation' (*Politico*, 2019) ⟨https://politico.com/interactives/2019/how-to-fix-politics-in-america/polarization/more-social-media-regulation/⟩; Steve Rathje, Jay J Van Bavel, and Sander Van Der Linden, 'Out-group animosity drives engagement on social media' (2021) 118 Proceedings of the national academy of sciences 26; Antoine Banks and others, '# polarizedfeeds: Three experiments on polarization, framing, and social media' (2021) 26(3) The International Journal of Press/Politics 609.

[105] Kahneman, Sibony, and Sunstein (n 72).

[106] Alexander Campolo and Kate Crawford, 'Enchanted determinism: Power without responsibility in artificial intelligence' (2020) 6 Engaging Science, Technology, and Society; Frederike Kaltheuner, *Fake AI* (2021).

[107] Jeremy Waldron, 'How law protects dignity' (2012) 71(1) The Cambridge Law Journal 200.

[108] John Danaher and Henrik Skaug Saetra, 'Technology and moral change: the transformation of truth and trust' (2022) 24(3) Ethics and Information Technology 35.

[109] Eric Winsberg, 'Reliability without Truth' in Eric Winsberg (ed), *Science in the Age of Computer Simulation* (University of Chicago Press 2010).

[110] Joseph Raz, *The authority of law: essays on law and morality* (Oxford University Press 2009) 222.

[111] Teo, 'How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework' (n 5).

bottom-line conditions for a moral community[112] – potentially impacting human dignity premised upon individuals as social autonomous beings.[113]

However, even this account is inadequate. The process of mediation between emerging technologies such as AI is a process of co-creation – individuals shape and mediate the contours of these interactions. Individuals can be intertwined as co-authors of their own resulting harm. Thus, this is not about Zuboff's portrayal of Big Tech corporations as appropriating human experiences, but about individuals playing a mediating role as well.[114] Consent is one element of this complicity. Hurd denotes consent as 'moral magic' and traces this capacity as autonomy – namely the capacity of a person to act as a self-legislator.[115] The idea of consent is also dominant within data protection discourses centered around self-management of privacy practices.[116] However, as discussed in Section 'The Instrumental Atomisation of the Individual', the mode of engagement by many technology companies relies upon psychological tools – often in sub-conscious ways that seek to undermine the very premise of autonomous decision-making. Susser has noted that some of these measures amount to manipulation – which he defines as a form of hidden influence, 'the covert subversion of another person's decision-making power.'[117] In fact, this covertness itself is a factor in the precarity of the structural factors, acting as enablers of harm. Such 'hidden influences' are by their nature hidden and seamless – within the architectures of AI/ML.[118] Barber considers the potential for monopoly over information architectures as a form of tyranny, arguing that: '[t]here is no tyranny more dangerous than an invisible and benign tyranny, one in which subjects are complicit in their victimisation, and in which enslavement is a product of circumstance rather than intention.'[119] The idea of meaningful individual consent, for example within the GDPR, is laudable as a theoretical premise but fails to account for the structural atomisation of individuals from their embedded social settings through forms of epistemic and contextual precarity, as afforded by AI. Individual human rights remain ironically intact even as the rug is pulled from under their feet.

Conclusion

The three forms of structural atomisation show that the intuition of AI/ML, reflecting the values of the social ontology of human rights, may be misplaced. At the same time, the international human rights framework, oriented around protection of the individual and through individual rights, is left ironically intact even as the individual herself is *othered* through AI/ML systems. This paper has identified three ways in which this might occur. First, the instrumental atomisation of the individual occurs through the modulation and mediation of AI/ML systems that construct the individual. This takes place through group based correlations applied to the individual in such a way that the individual has no say nor direct knowledge of how these systems shape their construction. The second form of structural atomisation occurs where the individual is instrumentally optimised – in ways that go against their own interest. Optimisation parameters necessitate computational tractability of social phenomena and in effect use the individual as a means to an end. Third, structural atomisation occurs through the contextual and epistemic precarity of AI/ML modulated contexts. Such contexts can act as structural enablers of harm, by *othering* the individual and yet do not typically engage the vernacular of

[112] Roger Brownsword, 'From Erewhon to AlphaGo: for the sake of human dignity, should we destroy the machines?' (2017) 9(1) Law, Innovation and Technology 117.

[113] *Pretty v. the United Kingdom*, no. 2346/02, ECHR 2002-III

[114] Nick Clegg, 'You and the Algorithm: It Takes Two to Tango' (2021) ⟨https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2⟩.

[115] Heidi M Hurd, 'The moral magic of consent' (1996) 2(2) Legal theory 121.

[116] Daniel J Solove, 'Introduction: Privacy self-management and the consent dilemma' (2012) 126 Harvard Law Review 1880.

[117] Susser, Roessler, and Nissenbaum (n 52) 3.

[118] Van Den Eede (n 86).

[119] Benjamin R Barber, 'Three scenarios for the future of technology and strong democracy' (1998) 113(4) Political science quarterly 573; See also Sætra who quotes Barber. Henrik Skaug Sætra, 'The tyranny of perceived opinion: Freedom and information in the era of big data' (2019) 59 Technology in Society

harms under human rights law.[120] This paper has thus demonstrated how the computational turn through computational thinking (datafication), computational law and computational environments *others* the individual in ways that threaten to foreclose the normative expectations associated with the social ontology of human rights.[121]

This paper sets out to challenge and provoke. For the former, by deploying the lens of the three forms of structural atomisation, I hope to challenge the AI/ML community to think about how AI/ML affordances can divorce the individual from their situated and embedded sociality. On the other hand, the paper also serves as a provocation to human rights theorists and practitioners to think critically about the transposition of the rights framework to concerns raised by AI/ML systems, presuming the sufficiency of the 'normative equivalency' paradigm. The mapping this article provides can help to better account for the social ontology of human rights in our computational environments. The three pressure points highlighted in this paper exhort that room be left for the individual to define and engage with the positive agency that is underpinned by their socially situated existence.

References

AlgorithmWatch, 'How Dutch Activists Got an Invasive Fraud Detection Algorithm Banned' (2020) ⟨https://algorithmwatch.org/en/syri-netherlands-algorithm/⟩.

Alston P, 'Report of the Special Rapporteur on Extreme Poverty and Human Rights on Digital Welfare States' (2019) ⟨https://documents-dds-ny.un.org/doc/UNDOC/GEN/N19/312/13/PDF/N1931213.pdf?OpenElement⟩.

Banks A and others, '# polarizedfeeds: Three experiments on polarization, framing, and social media' (2021) 26(3) The International Journal of Press/Politics 609.

Barber BR, 'Three scenarios for the future of technology and strong democracy' (1998) 113(4) Political science quarterly 573.

Beckmann P, Köstner G, and Hipólito I, 'An alternative to cognitivism: computational phenomenology for deep learning' (2023) 33(3) Minds and Machines 397.

Beitz C, 'What human rights mean' (2003) 132(1) Daedalus 36.

Berlin I, 'Two Concepts of Liberty' in *Reading Political Philosophy* (Routledge 2014).

Besson S, 'The law in human rights theory' (2013) 7 Zeitschrift für Menschenrechte-Journal for Human Rights 120.

Binns R, 'On the apparent conflict between individual and group fairness' in *Proceedings of the 2020 conference on fairness, accountability, and transparency* (2020).

Bowker GC and Star SL, *Sorting things out: Classification and its consequences* (MIT press 2000).

Boxell L, Gentzkow M, and Shapiro JM, 'Greater Internet use is not associated with faster growth in political polarization among US demographic groups' (2017) 114(40) Proceedings of the National Academy of Sciences 10612.

Brandeis L and Warren S, 'The Right to Privacy' (1890) IV Harvard Law Review 193.

Brincker M, 'Privacy in public and the contextual conditions of agency' in *Privacy in public space* (Edward Elgar Publishing 2017).

Brkan M, 'Artificial intelligence and democracy: The impact of disinformation, social bots and political targeting' (2019) 2 Delphi – Interdisciplinary Review of Emerging Technologies 66.

Brownsword R, 'From Erewhon to AlphaGo: for the sake of human dignity, should we destroy the machines?' (2017) 9(1) Law, Innovation and Technology 117.

Buchanan A, 'Why International Legal Human Rights?' in R Cruft, SM Liao, and M Renzo (eds), *Philosophical Foundations of Human Rights* (Oxford University Press 2015).

[120] The EU's Artificial Intelligence Act which was adopted in 2024 and regarded as the world's first attempt to comprehensively regulate AI, indirectly addresses some concerns raised in this article. For example, the risk-based approach requires providers of AI systems used in high risk use cases to live up to heightened obligations, including on data governance, record keeping, having in place a risk management system, transparency obligations and enabling 'human-in-the-loop.' These help in tempering the risk to individuals, such as those seeking social welfare, as highlighted in this article. However, the three pressure points on the social ontology of human rights is not directly addressed by the Act.

[121] I thank the anonymous reviewer who helped to nuance this point.

Bucher T, 'Want to be on the top? Algorithmic power and the threat of invisibility on Facebook' (2012) 14(7) New media & society 1164.

Campolo A and Crawford K, 'Enchanted determinism: Power without responsibility in artificial intelligence' (2020) 6 Engaging Science, Technology, and Society.

Cavoukian A, 'Operationalizing Privacy by Design: A Guide to Implementing Strong Privacy Practices' (*Information and Privacy Commissioner*, 2012) ⟨https://gpsbydesigncentre.com/wp-content/uploads/2021/08/Doc-5-Operationalizing-pbd-guide.pdf⟩.

Chesney B and Citron D, 'Deep fakes: A looming challenge for privacy, democracy, and national security' (2019) 107 California Law Review 1753.

Clayton A, 'Fake AI-Generated Image of Explosion near Pentagon Spreads on Social Media' (*The Guardian*, 2023) ⟨https://www.theguardian.com/technology/2023/may/22/pentagon-ai-generated-image-explosion⟩.

Clegg N, 'You and the Algorithm: It Takes Two to Tango' (2021) ⟨https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2⟩.

Cohen JE, *Configuring the Networked Self: Law, Code, and the Play of Everyday Practice* (Yale University Press 2012).

— 'What privacy is for' (2012) 126 Harvard Law Review 1904.

— 'Turning privacy inside out' (2019) 20(1) Theoretical inquiries in law 1.

COHUBICOL, 'Data-driven law' ⟨https://www.cohubicol.com/about/data-driven-law/⟩.

Constantaras E and others, 'Inside the Suspicion Machine' (*Wired*, 2023) ⟨https://www.wired.com/story/welfare-state-algorithms/⟩.

Cristianini N, 'Shortcuts to Artificial Intelligence' in M Pelillo and T Scantamburlo (eds), *Machines We Trust: Perspectives on Dependable AI* (MIT Press 2021).

Danaher J and Saetra HS, 'Technology and moral change: the transformation of truth and trust' (2022) 24(3) Ethics and Information Technology 35.

Dencik L, 'The datafied welfare state: A perspective from the UK' in A Hepp, J Jarke, and L Kramp (eds), *New perspectives in critical data studies: The ambivalences of data power* (Springer International Publishing Cham 2022).

Donnelly J, *The concept of human rights* (Cornell University Press 2013).

Douek E, 'The limits of international law in content moderation' (2021) 6 UC Irvine Journal of International, Transnational & Comparative Law 37.

Douzinas C, *The end of human rights: Critical thought at the turn of the century* (Bloomsbury Publishing 2000).

Dror-Shpoliansky D and Shany Y, 'It's the end of the (Offline) World as we know it: From Human rights to Digital Human rights – a proposed typology' (2021) 32(4) European Journal of International Law 1249.

Dvoskin B, 'International Human Rights Law Is Not Enough to Fix Content Moderation's Legitimacy Crisis' (*Berkman Klein Center Collection*, 2020) ⟨https://medium.com/berkman-klein-center/international-human-rights-law-is-not-enough-to-fix-content-moderations-legitimacy-crisis-a80e3ed9abbd⟩.

Eticas, The External Audit of the VioGén System, 'Association Eticas Research and Innovation' (2022).

Fourcade M and Kluttz DN, 'A Maussian bargain: Accumulation by gift in the digital economy' (2020) 7 Big Data & Society 1.

Freeman M, *Human rights* (Third edition, Polity Press, Cambridge 2017).

Gandy OH, 'Statistical surveillance: Remote sensing in the digital age' in D Lyon, KD Haggerty, and K Ball (eds), *Routledge handbook of surveillance studies* (Routledge 2012).

Gerards J and Borgesius FZ, 'Protected grounds and the system of non-discrimination law in the context of algorithmic decision-making and artificial intelligence' (2022) 20 Colorado Technology Law Journal 1.

Glendon MA, *A World Made New: Eleanor Roosevelt and the Universal Declaration of Human Rights* (First edition, Random House Trade Paperbacks 2002).

Goffman E, *The Presentation of Self in Everyday Life* (Anchor Books 1990).

Gould CC, *Marx's Social Ontology Individuality and Community in Marx's Theory of Social Reality* (MIT Press 1980).

— *Rethinking democracy: Freedom and social cooperation in politics, economy, and society* (Cambridge University Press 1990).

— 'A Social Ontology of Human Rights', in R Cruft, SM Liao, and M Renzo (eds), *Philosophical Foundations of Human Rights* (Oxford University Press 2015).

Green B and Viljoen S, 'Algorithmic realism: expanding the boundaries of algorithmic thought' in *Proceedings of the 2020 conference on fairness, accountability, and transparency* (2020).

Griffin R, 'Rethinking Rights in Social Media Governance' (*Verfassungsblog*, 2022) ⟨https://verfassungsblog.de/rethinking-rights⟩.

Hacking I, 'Making Up People' in T Heller, M Sosna, and D Wellbery (eds), *Reconstructing Individualism* (Stanford University Press 1985).

Haidt J, 'More Social Media Regulation' (*Politico*, 2019) ⟨https://politico.com/interactives/2019/how-to-fix-politics-in-america/polarization/more-social-media-regulation/⟩.

Hartzog W and Selinger E, 'Big data in small hands' (2013) 66 Stanford Law Review 81.

Heaven D, 'Why faces don't always tell the truth about feelings' (2020) 578(7796) Nature 502.

Henriques-Gomes L, 'Robodebt Class Action: Coalition Agrees to Pay $1.2bn to Settle Lawsuit' (*The Guardian*, 2020) ⟨https://www.theguardian.com/australia-news/2020/nov/16/robodebt-class-action-coalition-agrees-to-pay-12bn-to-settle-lawsuit⟩.

Hildebrandt M, 'Privacy and Identity' in E Claes, A Duff, and S Gutwirth (eds), *Privacy and the Criminal Law* (Intersentia 2006).

— *Smart technologies and the end (s) of law: novel entanglements of law and technology* (Edward Elgar Publishing 2015).

— *Law for computer scientists and other folk* (Oxford University Press 2020).

Hong S.-h, 'Prediction as extraction of discretion' (2023) 10(1) Big Data & Society 20539517231171053.

Huq AZ, 'A right to a human decision' (2020) 106 Virginia Law Review 611.

Hurd HM, 'The moral magic of consent' (1996) 2(2) Legal theory 121.

Jasanoff S, *Science at the bar: Law, science, and technology in America* (vol 9, Harvard University Press 1997).

Kahneman D, Sibony O, and Sunstein CR, *Noise: A flaw in human judgment* (Brown Spark 2021).

Kaltheuner F, *Fake AI* (2021).

Kapur R, *Gender, Alterity and Human Rights : Liberal Freedom in a Fishbowl* (Edward Elgar Publishing 2018).

Kleinberg J, Mullainathan S, and Raghavan M, 'Inherent trade-offs in the fair determination of risk scores' [2016] arXiv preprint arXiv:1609.05807.

Kosinski M, 'Facial recognition technology can expose political orientation from naturalistic facial images' (2021) 11(1) Scientific reports 100.

Langford M, 'Taming the digital leviathan: Automated decision-making and international human rights' [2020].

Lehr D and Ohm P, 'Playing with the data: what legal scholars should learn about machine learning' (2017) 51 UCDL Review 653.

Liu H.-Y, 'The digital disruption of human rights foundations' in *Human Rights, Digital Society and the Law* (Routledge 2019).

— 'AI challenges and the inadequacy of human rights protections' (2021) 40(1) Criminal Justice Ethics 2.

Mager A, 'Algorithmic ideology: How capitalist society shapes search engines' (2012) 15(5) Information, Communication & Society 769.

Morsink J, 'World war two and the universal declaration' (1993) 15 Human Rights Quarterly 357.

Moyn S, *The last utopia: human rights in history* (Harvard University Press 2012).

Mutua M, 'The Ideology of Human Rights' (1996) 36 Virginia Journal of International Law 589.

NBC News, 'Microsoft Is Removing Emotion Recognition Features from Its Facial Recognition Tech' ⟨https://www.nbcnews.com/tech/tech-news/microsoft-removing-emotion-recognition-features-facial-recognition-tec-rcna35087⟩.

Nguyen CT, 'Echo chambers and epistemic bubbles' (2020) 17(2) Episteme 141.

Nielsen RK and Ganter SA, *The power of platforms: Shaping media and society* (Oxford University Press 2022).

Nilsson NJ, *The quest for artificial intelligence: A History of Ideas and Achievements* (Cambridge University Press 2010).

NIMHD, 'Structural Racism and Discrimination' (2022) ⟨https://web.archive.org/web/20220628032157/https://www.nimhd.nih.gov/resources/understanding-health-disparities/srd.html⟩.

Nissenbaum H, 'Privacy as contextual integrity' (2004) 79 Washington Law Review 119.

Nordbrandt M, 'Affective polarization in the digital age: Testing the direction of the relationship between social media and users' feelings for out-group parties' (2023) 25(12) New media & society 3392.

Nussbaum MC, *Creating capabilities: The human development approach* (Harvard University Press 2011).

OECD, 'The OECD Artificial Intelligence (AI) Principles' ⟨https://oecd.ai/en/ai-principles⟩.

Peter F, 'Political legitimacy' (2010) ⟨https://plato.stanford.edu/archives/spr2014/entries/legitimacy/⟩.

Phan T and Wark S (2021) ⟨https://culturemachine.net/vol-20-machine-intelligences/what-personalisation-can-do-for-you-or-how-to-do-racial-discrimination-without-race-thao-phan-scott-wark/⟩.

Placido DD, 'Why Did "Balenciaga Pope" Go Viral?' (*Forbes*, 2023) ⟨https://www.forbes.com/sites/danidiplacido/2023/03/27/why-did-balenciaga-pope-go-viral/⟩.

Porter TM, 'Trust in numbers: The pursuit of objectivity in science and public life' [2020].

Ranjan R and others, 'Deep learning for understanding faces: Machines may be just as good, or better, than humans' (2018) 35(1) IEEE Signal Processing Magazine 66.

Rathje S, Van Bavel JJ, and Van Der Linden S, 'Out-group animosity drives engagement on social media' (2021) 118 Proceedings of the national academy of sciences 26.

Raz J, *The authority of law: essays on law and morality* (Oxford University Press 2009).

Rini R, 'Deepfakes and the epistemic backstop' (2020) 20 Philosophers' Imprint 1.

Russell SJ and Norvig P, *Artificial intelligence: a modern approach* (Third edition, Prentice Hall 2010).

Sætra HS, 'The tyranny of perceived opinion: Freedom and information in the era of big data' (2019) 59 Technology in Society.

Sætra HS and Mills S, 'Psychological interference, liberty and technology' (2022) 69 Technology in Society 101973.

Selbst AD and others, 'Fairness and abstraction in sociotechnical systems' in *Proceedings of the conference on fairness, accountability, and transparency* (2019).

Sen A, 'Human rights and capabilities' (2005) 6(2) Journal of human development 151.

Sloot B van der, 'Do groups have a right to protect their group interest in privacy and should they? Peeling the onion of rights and interests protected under article 8 ECHR' in L Taylor, L Floridi, and B Van der Sloot (eds), *Group privacy: new challenges of data technologies* (Springer 2017).

Solove DJ, 'Introduction: Privacy self-management and the consent dilemma' (2012) 126 Harvard Law Review 1880.

Sunstein CR, 'Algorithms, correcting biases' (2019) 86(2) Social Research: An International Quarterly 499.

Susser D, Roessler B, and Nissenbaum H, 'Online manipulation: Hidden influences in a digital world' (2019) 4 Georgia Law Technology Review 1.

Tasioulas J, 'On the foundations of human rights' in R Cruft, SM Liao, and M Renzo (eds), *Philosophical Foundations of Human Rights* (Oxford University Press 2015).

Teo SA, 'How Artificial Intelligence Systems Challenge the Conceptual Foundations of the Human Rights Legal Framework' (2022) 40 Nordic Journal of Human Rights 216.

— 'Human dignity and AI: Mapping the contours and utility of human dignity in addressing challenges presented by AI' (2023) 15(1) Law, Innovation and Technology 241.

Tversky A and Kahneman D, 'The framing of decisions and the psychology of choice' (1981) 211(4481) Science 453.

UNESCO, 'Recommendation on the Ethics of Artificial Intelligence' (*UNESCO General Conference*, 2021) ⟨https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14⟩.

Valcke P, Clifford D, and Dessers VK, 'Constitutional Challenges in the Emotional AI Era' in A Reichman and others (eds), *Constitutional Challenges in the Algorithmic Society* (Cambridge University Press 2021).

Van Den Eede Y, 'In between us: On the transparency and opacity of technological mediation' (2011) 16 Foundations of Science 139.

Van der Sloot B, 'Is the Human Rights Framework Still Fit for the Big Data Era? A Discussion of the ECtHR's Case Law on Privacy Violations Arising from Surveillance Activities' in S Gutwirth, R Leenes, and PD Hert (eds), *Data Protection on the Move: Current Developments in ICT and rivacy/Data Protection* (Springer 2016).

Van Dijck J, 'Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology' (2014) 12(2) Surveillance & society 197.

Verdirame G, 'Rescuing human rights from proportionality' in R Cruft, SM Liao, and M Renzo (eds), *Philosophical Foundations of Human Rights* (Oxford University Press 2015).

Wachter S, 'The theory of artificial immutability: Protecting algorithmic groups under anti-discrimination law' (2022) 97 Tulane Law Revuew 149.

Wachter S, Mittelstadt B, and Russell C, 'Bias preservation in machine learning: the legality of fairness metrics under EU non-discrimination law' (2020) 123 West Virginia Law Review 735.

Waldron J, 'Theoretical Foundations of Liberalism' (1987) 37 The Philosophical Quarterly 127.

— 'How law protects dignity' (2012) 71(1) The Cambridge Law Journal 200.

Winner L, *The whale and the reactor: A search for limits in an age of high technology* (University of Chicago Press 2020).

Winsberg E, 'Reliability without Truth' in E Winsberg (ed), *Science in the Age of Computer Simulation* (University of Chicago Press 2010).

Yeung K, 'Algorithmic regulation: A critical interrogation' (2018) 12(4) Regulation & governance 505.

Yu R and others, 'Deep learning: A generic approach for extreme condition traffic forecasting' [2017] Proceedings of the 2017 SIAM international Conference on Data Mining 777.

Zuckerberg M, 'Bringing the World Closer Together' (2021) ⟨https://www.facebook.com/notes/393134628500376/⟩.

A reply: Some counterarguments

Wayne Wobcke • Associate Professor, UNSW Sydney, rw.wobcke@unsw.edu.au

The crux of the paper is the claim that there are three ways that AI/ML systems undermine the 'social ontology' of human rights. First, note that while the author refers to 'AI/ML systems', a wider definition is adopted that covers *both* other types of AI system such as Recommender Systems, Decision-Support Systems, etc., *and* other technical approaches, including Statistics, etc. Thus the broader issues are not new, and not specific to AI or ML.

For clarity, the main argument is outlined below; (1)–(4) are accepted, while the main claims are in three parts 5(a)–(c).

1. Human rights are construed socially (Gould).
2. ECHR Article 8 requires the ability for individuals to meaningfully exercise autonomy in establishing and developing social relations.
3. The exercise of this autonomy is situational and contextual (Cohen).
4. Meaningfully exercising autonomy involves an individual having a 'say' in how one is 'read and modulated' (by an ML system) in a digital environment (Hildebrandt).
5. (a) ML and other systems represent individuals as separate from the social contexts necessary for them to meaningfully exercise autonomy, limiting their capacity to 'self-present' differently in different circumstances; (b) ML systems are typically designed to optimize some loss function that privileges predictive accuracy, limiting an individual's capacity to understand (therefore, have a 'say' in) decisions made using an ML system; and (c) Recommender Systems mediate the disruption of enabling conditions for exercising autonomy through facilitating the spread of misinformation, etc.

I will focus on the three main arguments 5(a)–(c), which do not contradict (1)–(4); the claim is that these 'create tension' with (1)–(4). Each is of the form 'B because A', and the author argues for A, B, the causal connection between A and B, and the tensions with (1)–(4), so there are plenty of lines of counterargument, though I will mention just some of them.

Atomization of Individuals in ML Representations.

The primary examples are from social media and e-commerce. But as noted in the paper, social media *does* 'afford' the ability for people to form new 'communities', develop new interests, and in general develop new social relations (facilitated by Recommender Systems), but also to 'self-present' multiple identities (e.g. family persona on Facebook, professional persona on LinkedIn), so it is odd to be arguing *against* ML supporting a social conception of personal development. As to the 'co-creation' of one's profile generally, surely everyone nowadays knows that web sites are tracking their every mouse click, so users *are* able to 'co-determine how [they] will be read' through anticipating the behaviour of the ML system. The author accepts this very point in the discussion of 5(c).

Another example where this supposedly bites is mass surveillance via facial recognition. This is problematic for more reasons than 'displacement of the subject', but it would be good to know how interpretation of the ECHR balances the right to privacy with the right to safety from, say, a terrorist attack in a public arena. Here I believe we as a society, though not we as individuals, need a 'say' in what data is collected in the name of national security.

ML uses Optimization.

Two inappropriate applications are described, one using a statistical model for estimating the likelihood of violence

against women, the other a risk assessment tool for fraud detection. A major issue is that the underlying models are not very good, and this is likely because the training data is flawed, leading to incorrect predictions. But I don't think the problem is with ML *per se*, or the reliance on particular metrics. In terms of the overall structure of the paper, the argument about the causal role of ML should apply even when the ML model works well, such as our work addressing SDG 1 (No Poverty) through developing improved ML methods for poverty targeting. The representation of individuals in such models is inscrutable, however evaluation against metrics is essential to demonstrate the method's effectiveness. Does the author's account imply that we should *always* prioritize the individual right to contest decisions made using an ML system, and thus automatically reject the use of a superior ML system? If so, the right to contest decisions made using ML systems is overly strict; if not, focus shifts to the usual practical question of how competing ethical principles are weighed against one another. Does the 'social ontology' conception of human rights have anything to say about this?

Also, nothing requires fairness for ML applications to be addressed in a technical manner. I think that ethical frameworks are much better suited to defining and achieving fair processes and outcomes with deployed ML systems.

Big Tech and Social Media.

We are again on familiar territory with the observation that deepfakes/fake news and misinformation/disinformation, via Recommender Systems, create echo chambers/filter bubbles that disturb belief systems and promote hate speech and political polarization. This is a problem on anyone's account of human rights, one that can be addressed with regulation. I suggest that social media companies (which are much more than mere 'platforms') have a 'duty of care' to their users, and that AI could be better utilized to help fulfil this duty. Is such a duty enshrined in the ECHR or a foundational principle of the Digital Services Act?

Conclusion

An important aspect of the paper is the concern with ML *systems*, and no system is deployed in a vacuum. This foregrounds the sociotechnical nature of ML systems that includes the users, decision/policy making organizations using the systems, those involved in the creation and distribution of data used by the ML models, and those responsible for validating the models and maintaining the systems over time. Rather than focusing on the underlying techniques *vis à vis* social ontology of human rights, an analysis of if and how the ECHR confers rights on groups and organizations, and of whether there is any guidance on how those rights are to be balanced against individual rights, would be more in keeping with this point of view.

Author's reponse

Sue Anne Teo

I would like to thank Prof. Wobcke for his thoughtful and critical response my paper. This Response gives me a chance to clarify pressure points – addressing the *what*, the *why* and the *how*.

Prof. Wobcke observes the concerns in the article are not new, nor are they specific to AI/ML. At the same time, never in the history of humanity has the thoughts and expression of billions been centrally controlled nor accessible by corporations, let alone one corporation. This form of power, alongside the lack of transparency, means that we do not participate in this increasingly public space in equal terms. Data accumulation is in this way 'a core component of political economy in the 21st century.'[122] Determining or even co-determining algorithmically mediated sociality is onerous when algorithms can be tinkered with without transparency nor accountability.[123] At the same time, the lack of knowledge and visibility of algorithmic curation of content raises the spectre of possible manipulation.[124] This not only does not engage sociality, but actively subverts it.

Addressing the 'why', the existence of social media platforms ostensibly engages sociality in name through its ability to connect many at once. An incongruence is however present. The business model is driven by engagement, primarily through recommender systems, and targeted advertising, not upon connection. The 'pay-or-consent' debate, where the European Data Protection Board called for real user choice in either consenting to or paying for platform access,[125] highlights the tension between the right to privacy (including connecting with others) versus the underlying attention economy. Might human rights potentially come with a price?[126]

Finally, in addressing the 'how' question, the article invited reflection on how the design of our computational environments can better respect and enable the social ontology of human rights. I agree with Prof. Wobcke that technology is not the focus. Rather, it should be about its design and application within a contextualised sociality. The human rights-based approach, shortened to PANEL, which foregrounds participation, accountability, non-discrimination, empowerment and legality,[127] complements human rights law and existing regulatory efforts underway, including through stakeholder participation required under the EU AI Act.[128] It can be a first step in engaging with and enabling sociality within our computational environments.

[122] Jathan Sadowski, 'When data is capital: Datafication, accumulation, and extraction' (2019) 6 Big data & society 1.

[123] Kai Kupferschmidt, 'A Study Found Facebook's Algorithm Didn't Promote Political Polarization. Critics Have Doubts' (*Science*, 2024) ⟨https://www.science.org/content/article/study-found-facebook-algorithm-didnt-promote-political-polarization-critics-doubt⟩.

[124] Sue Anne Teo, 'How to think about freedom of thought (and opinion) in the age of AI' (2024) 53 Computer Law & Security Review.

[125] European Data Protection Board, 'Opinion 08/2024 on Valid Consent in the Context of Consent or Pay Models Implemented by Large Online Platforms'.

[126] Reuters, 'EU Set to Charge Meta over "Pay or Consent", FT Reports' (2024) ⟨https://www.reuters.com/technology/eu-charges-meta-over-pay-or-consent-ft-reports-2024-07-01⟩.

[127] Australian Human Rights Commission, 'Human Rights Based Approaches' ⟨https://humanrights.gov.au/our-work/rights-and-freedoms/human-rights-based-approaches⟩.

[128] See Article 27 of the EU AI Act on fundamental rights impact assessments and Recital 96 on stakeholder participation.