Gavin Duffy. A Virtuous Ethics of AI: Conviviality as a Regulatory Framework.
Goldsmiths, University of London.

g.duffy@gold.ac.uk

Abstract

In recent years, we have seen the AI industry grow astronomically, becoming a technology that will seemingly impact all elements of our daily lives in the near future. AI omnipresence is now treated by many as almost inevitable, leaving only the question of who should control this technology. This has, understandably, drawn much concern from regulators (both at the state and international level), as well as from many within the AI industry. Unfortunately, there has been less agreement on how we should regulate AI and what the ethical framework for such regulation should be. This article presents two contrasting ethical frameworks of justice in relation to AI: John Rawls (1999) theory of justice as fairness and Ivan Illich's (1973) notion of conviviality. This article critiques the Rawlsian approach as being too concerned with an abstract notion of a 'fair' playing field when establishing notions of justice (through its concepts of the original position and difference principle) and ignoring, or even embracing, injustice of outcomes. In contrast, this article argues in favour of the conviviality approach, presenting it as an ethical framework based in virtue and concerned primarily with outcomes and material reality, rather than hypothetical and semantic notions of fairness. This includes showing how conviviality can be applied practically, applying a comprehensive (or 'thick') notion of sustainability to AI. This thick sustainability considers the entire lifecycle of AI development in considering regulation, including the impacts on ecology as well as the impacts on people. Thus, the conviviality approach de-centres technology and re-centres both humans and our natural environment, providing a holistic ethical framework which must underpin any serious regulation of AI.

Keywords: AI, conviviality, Illich, Rawls, justice, degrowth

Introduction

In recent years, the rise of artificial intelligence (AI) has been astronomical. The public release of ChatGPT is already seen as a watershed moment in re-organising society around AI (Baker, 2024). Nvidia's (a GPU company now specialising in AI chips (Oi, 2024)) growth reflects the changes already being brought about by AI, both economically (briefly becoming the most valuable company in the world (Labiak, 2024)) and geopolitically (becoming a proxy for US-

China tensions (McMorrow and Olcott, 2024)). The presence of AI is felt in domestic governance too, with increasing numbers of countries developing national AI policies, generally with the outlook that AI will inevitably become a central element in our everyday lives (DSIT, 2025a). Combined with well-known and established 'Digital Lords' embracing the technology (Brevini, 2023), AI has seemingly become an unavoidable prospect for even the most technologically hesitant.

Naturally, this raises questions about the governance of AI. If this technology is to be such a terrific force across society, how should it be governed? This article addresses this question through advocating for an ethics of conviviality: a socially oriented form of AI governance, rooted in the notions of human flourishing and equity. In short, a conception of human-centred AI ethics. The following section will detail why such a conception of AI ethics is needed, examining what makes AI distinct from previous digital technologies for regulatory purposes. This article will then discuss two central approaches to AI ethics. The first is a liberal approach, inspired by John Rawls (1999) and dominant in the field of AI ethics (Franke, 2021; Barsotti and Koçer, 2024). This approach suggests a minimisation of harm caused by AI, stemming from a deontological judgement of what constitutes a 'fair' playing field. The subsequent section presents a contrasting, more critical and expansive view on AI ethics, based on Illich's (1973) notion of conviviality. This convivial approach takes a distinctly more outcome-driven approach than the Rawlsian viewpoint, exemplified through examining AI in the media sector. As such, this article argues that the conviviality approach to AI ethics is more practical, more comprehensive, and desirable of the two frameworks, even if (or possibly because) it is also more demanding.

Why do we need an ethics of AI?

One pressing issue in the governance of AI is defining what we mean by "artificial intelligence". Burkhardt and Rieder (2024) note that one of the difficulties in assessing AI is that it is not a single technology but that current AI models represent something 'new' due to their generative, pre-trained, transformer (GPT) capabilities. These models are intended to be domain and task agnostic, based on large-scale foundation models which are much less specialised than generative adversarial network (GAN) models (which use large but narrow datasets to generate convincing outputs of a specific concept). This move towards a generalisable model with generative capabilities has political and ethical implications for societies: if current AI models are believed to be applicable to all tasks, they in turn can influence how we understand the world and what is capable within it. Amoore et al (2024, 2) describe this as AI "instantiating a model of the world,

and with it a set of political logics and governing rationalities that have profound and enduring effects on how we live today". With AI claiming the ability to see underlying (or latent) trends in large datasets, these technologies are becoming powerful actors in shaping our world, often to the (economic) benefit of already wealthy companies, such as Palantir, and at the expense of those already in precarious positions. We are already seeing various AI systems being used in ways which reproduce and reify existing inequalities for refugees (Madianou, 2021), intensify biases in fraud detection with Sweden's welfare system (Amnesty International, 2024), and to identify potential Palestinian targets (including civilians) in Israel's attacks on Gaza (Birch, 2024). This demands questions around the acceptable uses of AI in our world.

Alongside questions about the political characteristics of AI, there are questions of infrastructure. In defining AI, Crawford (2021) includes the creation, maintenance, and disposal processes of AI, rather than focusing solely on the AI product or marketing as experienced by the end user. Thus, AI is also the process of mining rare earth minerals and metals. Crawford reflects on how these practices are not entirely new. Instead, they echo the colonial and extractive history of other technological developments, such as the use of gutta-percha (a natural white latex) for insulating transatlantic telegraph cables. Again, we see the current regime of AI as maintaining long-standing systems of oppression, at both the national and international level, often in pursuit of new media and communication tools favouring the Global North.

These issues point to the need for comprehensive regulation around AI, one which considers the human and environmental impacts of these products first and foremost. Simultaneously, the breadth of these issues points to the difficulty in creating such regulation: if AI is so generalisable and to be used across all elements of society, identifying a specific AI 'regulatory target' seems almost impossible. Indeed, this already appears to be the case for the media sector in many areas, with AI being enabled to undermine intellectual property (IP) laws and the labour of human workers. In the EU, for example, AI developers use an exemption for data scraping in the 2019 Copyright Directive to justify training their models through practices which would normally be considered copyright infringement (Rankin, 2025). Similarly, the British government intends to relax copyright restrictions for developers in training AI models, effectively permitting what would otherwise be considered IP theft (Milmo, 2025). Miltner (2024) points to several more news articles discussing the theft of content and predatory data practices by AI (as well as AI models creating discriminatory or biased outputs) from a range of countries, primarily the US, Mexico, the UK, and India. For media and communication sectors, AI (and the broad scope of GPT-based AI in particular) evidently threatens the ability for people to create new art or

content, both through AI dominating the generation of content and continuing to devour any new human-produced media. In this context, regulation is often framed as being either impossible to create effectively (i.e. AI companies will find a way around it) or is simply in favour of AI companies.

As Miltner (2024, 27) highlights, even in media which laments the predatory and biased data regimes of AI tends to frame this as "just the way things are". This does not simply have to be the case, however. This is a discursive technique which naturalises the power of AI through its supposed inevitability and the powerlessness of citizens to resist (Markham, 2021). This is exemplified in the UK-based NGO Tony Blair Institute for Global Change suggesting that workers should move "beyond narratives of unemployment and Terminator" through a "greater emphasis… on how human workers can be empowered by robots" (Macon-Cooney et al, 2024, 45). In discussing how AI could be *better* regulated in this article, it is therefore imperative to first examine these underlying ways in which AI is conceived of and understood. Subsequently, this article will employ the framework of conviviality to outline a more holistic ethics of AI, informing more effective AI regulation. However, it is first important to outline the current ethical framework used to understand AI, namely a liberal one.

Approaches to AI ethics and regulation

The liberal perspective

Despite numerous critiques of neoliberalism within academic literature, a good deal of research continues to promote a liberal view of AI regulation. In particular, John Rawls' (1999) *A Theory of Justice* (TJ) continues to be influential for political philosophy in general (Laden, 2003) and egalitarianism more specifically (Stone, 2022). Further, Rawls' text transcends academic spheres, finding commercial success in its own time and maintaining prominence in (neo)liberal movements since (Coman, 2020). TJ therefore makes enduring contributions to normative understandings of social issues, justice, ethics, and rationality, extending beyond political theorists and social scientists, influencing economists, lawyers, and even theologians (Richardson and Weithman, 1999). It is this width of influence, across domains and time, that makes Rawls and TJ relevant to AI, a technology that promises to be so generalisable that it will be central to all elements of society. Regulating such a comprehensive technology requires an equally comprehensive ethical framework, given the difficulty described above in regulating the technology in a more piecemeal fashion. This article will therefore examine how Rawls' theory of justice has been applied to AI, highlighting the theory's shortcomings generating comprehensive, effect regulation. First, however, it is important to set out TJ's central concepts.

In TJ, Rawls (1999, 10) articulates the idea of 'justice as fairness', or "the principles that free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association [with society]" and which "regulate all further agreements". This does not need to lead to 'fair' outcomes, only that the principles of justice are initially agreed upon in a fair situation. Rawls outlines two primary principles for achieving this 'fair' justice: distributive justice and the difference principle. Pogge (1982) describes the first principle as guaranteeing the basic liberties of all people (emphasising that this should be understood as global in scope), with these basic liberties only constrained if it promotes greater liberty overall (e.g. the basic liberties of the intolerant may be restricted if it ensures the liberty of those they target and, by proxy, all others). Secondary to this is the difference principle, which states a society should seek to maximise the state of the least advantaged citizens, without violating the first principle (Estlund, 1996).

These are laudable ideas that few would disagree with. Less generously, they may be seen as so vague that few *could* disagree with them. It is therefore worth returning to Rawls (1999) for more detail on these principles. Regarding the difference principle, Rawls states that society should arrive at a conception of fairness (represented through equal liberties) through the original position. Rawls (1999, 11) compares the original position as "[corresponding] to the state of nature in the traditional theory of the social contract… [i.e.] a purely hypothetical situation characterized so as to lead to a certain conception of justice". This hypothetical situation occurs as a contractual negotiation with the intended outcome that "the principles that would be chosen, whatever they turn out to be, are acceptable from a moral point of view" for all (Rawls, 1999, 104). To achieve such results, however, requires that all actors reason from the *original position* behind a *veil of ignorance*. The veil of ignorance is again a hypothetical situation in which one does not know their place in society, his conception of the good, or even the circumstances of their own society overall. Instead, "the only particular facts which the parties know is that their society is subject to the circumstances of justice and whatever this implies" (Rawls, 1999, 119). The intention of the veil of ignorance is therefore to ensure that no one will "design principles to favor his particular condition", meaning that the principles of justice established "are the result of a fair agreement or bargain" and so will be rational (Rawls, 1999, 11). As such, the original position is "a status quo in which any agreements reached are fair" (Rawls, 1999, 104).

Secondly, when discussing the difference principle, Rawls measures what constitutes working to advantage the most disadvantaged not through changed outcomes but through altered

expectations. Specifically, Rawls (1999, 69) recommends that "we simply maximize the expectations of the least favored position subject to the required constraints... [as] the estimated gains from the situation of hypothetical equality are irrelevant, if not largely impossible to ascertain anyway". The difference principle is fulfilled through a positive change in *expectations* of the most disadvantaged in a society, justifying actual material inequalities and "initial inequality in life prospects" (Rawls, 1999, 68). Further, Rawls (1999, 68) positions the greater expectations of the already advantaged as fair and even positive for society as "the greater expectations allowed to entrepreneurs encourages them to do things which raise the prospects of laboring class". Thus, when examining the Rawlsian framework more closely, we can see the ways in which 'justice as fairness' acts to permit and justify inequalities, allowing only for very restricted redress to these issues.

It is, at this point, worth asking: how have Rawls' concepts been applied to AI? Westerstrand (2024) uses the Rawlsian framework to promotes ethical design and use of AI. Regarding Rawls' first principle (on basic liberties), Westerstrand (2024, 5) states that "Rawls offers a preliminary list of basic liberties… to be equally distributed". This includes "liberty and integrity of the person (including freedom from psychological oppression and physical assault and dismemberment)" (Rawls, 1999, 53). Expanding on this, Westerstrand (2024, 8) posits that "AI systems should not harm but support the liberty and integrity of the person, including freedom from psychological oppression and physical assault and dismemberment". This is a pressing matter, according to Westerstrand (2024, 8) as "AI has already being [sic] used in military to automate warfare" which risks causing physical oppression and assault. Regarding Rawls' second principle (the difference principle), Westerstrand (2024, 10) raises concerns that AI "could also lead [to] discrimination of people working certain professions", such as freelance designers or writers, concluding that AI should not be used is it could "negatively impact people's opportunities to seek income and wealth". Again, these are hardly objectionable concerns; they are legitimate insofar as they are both real and material, with NATO investing in Palantir's Maven Smart System (an AI-powered tool that sifts through battlefield data to "scan for targets and speed up attacks") (Foy and Bradshaw, 2025) and AI already being slated to cause massive job losses (Robinson, 2025).

It is, however, unclear how useful Rawls' principles of justice are in either example. Westerstrand (2024) does caution against the use of AI systems which impinge on liberty through physical assault. However, citing Johansson (2018), Westerstrand (2024) also claims that the AI-driven weapons could reduce causalities and so may adhere to Rawls' (1999) notion of liberty (this,

however, appears to ignore Johansson's (2018) warning that this applies only to the possessor of such weaponry and may actually lower the threshold for instigating a war as a result). It is initially somewhat clearer how the difference principle relates to those made unemployed by the use of AI systems, particularly within the arts. Indeed, Westerstrand (2024, 13) states that "following Rawls' theory, AI systems should always thus encourage societal improvement when used in processes that lead to inequalities". As always, it is less clear what this would look like in practice, with Westerstrand (2024) simply suggesting private corporations include the difference principle in their ethical frameworks. Further, Rawls (1999, 68) states that entrepreneurs may be granted unequal benefit under the difference principle should they "do things which raise the prospects of" the least advantaged, including making economic processes more efficient and innovation more rapid. This is exactly the claim made by AI boosters, e.g. the UK government's AI Opportunities Action Plan (DSIT, 2025b), which views AI as a part of the creative industries. Applying TJ and its principles at the case-by-case level can therefore become little more than semantic negotiation around what constitutes an acceptable amount of inequality, rather than eliminating this inequality.

This does not mean that Rawls can have no salience for AI regulation. It may merely mean that it is more important (and productive) to apply the principles of TJ to underlying principles of AI, rather than specific use cases. Indeed, Bay (2023), in critiquing Ashrafian's (2023) notion of a Rawlsian AI agent, suggests that the veil of ignorance, the original position, and difference principle are decidedly macro-principles, rendering them of limited utility for assessing specific AI. Gabriel (2022, 218) utilises a macro-principle approach, stating that AI is now a part of the background justice of our societies, playing an important role in many major institutions and social practices. However, this amounts to little more than recommendations for a public rationale being provided when governments use AI, including "nontechnical explanations of their performance", greater research on antidiscrimination practices and outcomes, and consideration of privacy as a basic right (Gabriel, 2022, 223). These recommendations come with some broad and limiting stipulations: rationale requirements for AI merely apply to "certain public contexts", and solely objects to "purely private goals"; antidiscrimination remains exclusively a matter of discussion; and privacy is only a basic right unless there is an "adequate justification" to the contrary (Gabriel, 2022, 223, 224). This, ultimately, provides only vague suggestions that AI should be reasonably transparent and interfering in certain contexts, to some degree, provided there is not a justification to act otherwise.

Gabriel's limited recommendations point to a central issue with applying a Rawlsian framework to AI and, simultaneously, why Rawls' notion of fairness remains a common one amongst AI-related ethicists (e.g. Larson, 2017; Hashimoto et al, 2018; Heidari et al, 2019; Li et al, 2021; Franke, 2024). As Jørgensen and Søgaard (2023) draw out, the continued use of Rawlsian fairness is due to the permissive nature of TJ, providing a range of exceptions and loopholes to its two central measures of equality. For example, Jørgensen and Søgaard (2023, 1186) state that through "Subgroup Test Ballooning" (tailoring a technology specifically to early adopters, with the argument that it will eventually be adapted for all end users) and "Snapshot-Representative Evaluation" (taking a sample population from the current userbase, rather than an fully representative or even weighted population sample), AI developers can give their products the appearance of 'fairness' (and so 'justness') through ignoring inconvenient (and generally the most precarious) population groups. As such, Rawlsian fairness "is too permissive to prevent common AI/NLP practices that actively contribute to global and social inequality gaps", while purporting to do the opposite (Jørgensen and Søgaard, 2023, 1190).

As noted above, Rawls (1999) discusses such exceptions in TJ, justifying income inequality as fair, for example, provided expectations of workers are managed appropriately. Rawls' notion of justice as fairness is intended to legitimate (at least some of) the inequalities experienced in liberal democracies when examined as a whole system. Applying the Rawlsian approach to AI serves primarily to justify inequalities encoded within and executed by these technologies as *one piece* of the whole system, framing these inequalities simply because of this system alone, rather than as being reified by AI and its developers. This produces distinct negative outcomes e.g. the further centralisation of English as the *lingua franca* at the expense of all other languages (Jørgensen and Søgaard, 2023) and a specific form of standardised English at the expense of other less nondominant Englishes (de Roock, 2024). Such a focus on a specific type of English shapes the ways in which AI models can 'think', perpetuating (dominant) Anglophone understandings of the world, including that of fairness and justice (Tacheva and Ramasubramanian, 2023). When considering the generalisable promises of AI and the universal standards demanded by TJ (Pogge, 1982), it is difficult to see how these exceptions should be justified as fair. In reality, through the permissive broadness of TJ, the Rawlsian framework enables a rhetorically robust but practically loose regulation of AI. This threatens inclusivity in media in ways much broader than the freelancers described by Westerstrand (2024), legitimating an extremely narrow and already dominant understanding of the world through the apparent vastness and consequent omnipotence of AI, leaving room for little else.

As a result, this article suggests that an alternative understanding of justice and fairness is needed for understanding and regulating AI in a manner that is more human-centred. Due to the tension between the deontological Rawls and de-deontological AI, this alternative approach must be more considerate of AI's consequences. This approach is Illich's (1973) conviviality.

Conviviality as an alternative approach

Before making an argument for a convivial approach to AI ethics, it is essential to outline what is meant by "convivial" here, understood through Illich's (1973) definition and application of the term. Instead, Illich uses convivial as a technical term to describe a society in which there is a responsibly limited usage of tools, with modern technologies serving politically interrelated citizens, rather than solely serving managers. Illich (1973, 11) explains that conviviality is an "intrinsic ethical value", that of "individual freedom realized in personal interdependence". A convivial society is therefore one in which people act in creative and autonomous relations with one another and their natural environment. This is contrasted with industrial society in which the power of machines consistently increases at the expense of the individual person, who is degraded to being a mere consumer and subject to the demands of others within a man-made environment.

This is not a binary distinction. Instead, it is only when a society falls below a certain level of conviviality (and industrial productivity rises above a certain level) that the populace becomes plagued by a sense of amorphousness and meaninglessness. Thus, conviviality does not equate to a complete rejection of technology nor that there is an inherently negative quality to technology. Rather, Illich notes that societies and their technologies can either be variously convivial or industrial depending on how they are owned, controlled, and used. Convivial societies are those which ensure a just distribution of unprecedented power (manifest through new technologies), ensuring that the autonomy of one person does not necessitate the subjugation of another. As such, a convivial approach to ethics is one which is interested in full participatory justice. This is in resistance to the ongoing amassing of power by professional elites "who promise to build up the machinery to deliver" futures which are dependent upon high production levels via increasing inequality and energy slaves (Illich, 1973, 12).

It is in this sense that convivial regulation should be understood: rooted in the notion of human flourishing and as a shared virtue. This again stands in contrast to regulation created around a Rawlsian framework of "justice as fairness", in which outcomes are rendered secondary to the

imagined conditions in which they were created. Conviviality as a shared virtue can also be seen in the origins of the term, underpinning the suggested notion of convivial regulation in this article. Illich's definition of conviviality draws upon Aquinas' (1947) argument that austerity is a virtue but must exist in conjunction with pleasure, that neither should be inordinate, instead balancing one another. Such a balance is essential, Aquinas claims, to prevent one from becoming burdensome upon others (should they excessively lack mirth) or to becoming boorish and rude (should lack austerity). It is this balance of mirth and austerity that we see in Illich's (1973) definition of conviviality as personal freedom through mutual interdependence. It is therefore important to note that conviviality is neither negative nor admonishing, even if it does make arguments against the current regulatory regimes. Instead, conviviality is a normative approach rooted in virtue, around the question of the good life at both the individual and collective level.

The convivial approach to ethics thus shares a similarity with the Rawlsian view. Both seek to maximise societal fairness through justice and see individual-level justice as contingent upon the societal-level organisation of fairness. However, the conviviality and Rawlsian approaches differ significantly in what this fairness means and how it is reached. As outlined above, Rawls (1999) puts forward the original position as a means of judging fairness. Once again, this necessitates that, due to the veil of ignorance, no one will "design principles to favor his particular condition" meaning that the principles of justice established "are the result of a fair agreement or bargain" and so will be rational (Rawls, 1999, 11). Such a suggestion appears to be, in itself, irrational. Our understandings of the present and imaginaries of the future are influenced by structural powers, including shaping our perceptions of what a just society is at all (Lukes, 2005). Within an industrial society, industrial forms of justice are to be an expected outcome of the original position, not because of an unwillingness of participants to engage with the *idea* of the original position but because ideas of what is rational (e.g. what values should be prioritised over others and to what extent, to achieve fairness) are inherently shaped by ontological viewpoints. We no longer sacrifice animals to god(s) as a means of repenting for our sins (van Dijk, 2008) but this does not make such activities irrational *in toto*; they simply exist within older forms of rationality. Unless it is believed that the entirety of history was irrational and that the present will always be viewed as rational, any outcomes of the original position must be assumed to be influenced by the context of their place, time, and culture. Illich implicitly recognises this through making an argument for a different form of rationality (conviviality over industrial). TJ does not.

This is a vital point of contention in the context of AI. It is not difficult to see how the current discourse around AI parallels Illich's (1973) warning of professional elites shaping how we imagine the future and political institutions promoting the goal of increased output through conflating the idea of "the good" with what is good for powerful institutions. This logic of industrial society is clearly seen through both national and supranational governments competing to most successfully curry favour with the digital lords of AI, e.g. the British government's AI Opportunities Action Plan (DSIT, 2025b) or the US government's immediate courting of SoftBank and OpenAI for greater AI investment (Hammond, 2025). Further, this approach is rationalised as promoting a common good through notions of increased employment, economic productivity, and environmental regeneration. This is despite many of these claims being visibly untrue and, further, incompatible with one another (Latouche, 2009), particularly given AI's resource intensiveness (Li et al, 2023). As such, any justice derived from an original position under this logical framework could not rationally arrive at a convivial perspective on AI, regardless of how "rational" such an outcome may be. Instead, the outcome from this original position would rationally be one which promotes increased use of AI in all spaces and an increasing allocation of resources and priority to AI. This is, in fact, what many AI boosters suggest and what many governments are seeking to do (DSIT, 2025b; Hammond, 2025). Whether or not such decisions are correct is immaterial to whether or not they are rational; they are rational within the given framework of thinking. Rawls (1999, 11) states that justice as fairness "does not mean that the concepts of justice and fairness are the same, any more than the phrase "poetry as metaphor" means that the concept of poetry and metaphor are the same". Similarly, rationality and correctness are not the same, even if something can be correct under a certain rationality.

Conversely, Illich's (1973) conviviality framework has been influential for many degrowth-oriented approaches to contemporary digital technologies, including AI. In particular, Illich's conviviality framework has inspired means of testing for 'fairness' in ways which are decidedly more robust and less permissive than Rawls' (1999) TJ. In considering specific products, for example, Vetter (2018) establishes a matrix of convivial technology which can act as a guide for what human-centred AI regulation may privilege. This involves promoting technologies which: recognise that humans exist in a series of relations to one another and so seek to promote positive relations between people; consider both material (hardware) and immaterial (software, knowledge) accessibility, as well as accessibility across different groups (e.g. addressing the traditionally male biases in technological development); have clear utility in their ecologies,

including ethical plans for the product's end-of-life, rather than simply being 'less harmful'; and consider the appropriateness of the product, with serious consideration of where it may *not* be useful, including where technologies may be desirable but not necessary. This, evidently, goes beyond a Rawlsian notion of fairness through a strict, clearly articulated criteria by which technologies should be measured across their lifecycle and its chain of production, resulting in a substantially less permissive framework for justice.

Considering sustainability at a more macro level, Heilinger et al (2024) develop a framework for assessing and regulating for the "thick" sustainability of AI. Thick sustainability is an approach to sustainable AI which looks not just at how the technology is used *for* sustainability purposes but also sustainable *as* a technology. This includes not only the environmental sustainability of AI but its social sustainability as well, discussed in the context of media in the following section. Heilinger et al contrast this with thin sustainability, which only examines the direct impacts of AI's immediate ecological actions, e.g. identifying more efficient strategies to deal with climate change, while prioritising economic sustainability over social sustainability. It is this 'thin' sustainability which AI ethicists and developers appeal to through the Rawlsian framework to make claims toward thin sustainability, relying on 'fair' exceptions carved out in ambiguous regulations (Gabriel, 2022), statistically and rhetorically concealing their supply chains (Crawford, 2021) and those othered by AI (Jørgensen and Søgaard, 2023).

In contrast, conviviality-based approaches such as that of Heilinger et al (2024) avoid the permissiveness of TJ through making companies responsible for the whole lifecycle of their product, and particularly its impacts. Through focusing on the life of a product, rather than theoretical assessments of fairness enabled through the Rawlsian approach, frameworks inspired by Illich (1973) pro-actively and continuously seek a society in which people are able to exist with greater agency, living in conjunction with technology rather than subject to it, i.e. a more convivial society. Rather than being permissive of an unjust outcome due to the supposedly fair nature of the contractual bargaining process which created the injustice, a framework of conviviality demands an outcome-oriented approach to fairness and justice. In practice, this is likely to come at the expense of the economic 'sustainability" (i.e. perpetual growth) prized by thin sustainability, recognising that this economic growth is inequitable and undesirable for a majority of the world's population, yoking them to an unjust economy of AI to enable the flourishing of a few.

The conviviality framework therefore operates as a more human-centred approach to regulation through this systematic approach to AI, in contrast to the narrower frame often used to assess

what an AI "does" or "is". Conviviality resists technosolutionist or technologically deterministic regulation through maintaining a critical (but not cynical) disposition to new digital technologies, seeing AI as yet another tool to be regulated and managed rather than as a digital Leviathan. This distinction is important, as we already see how AI is often framed as being almost mythical (Leaver and Srdarov, 2025), as opposed to a new watershed in the timeline of digital technology. This demystification of AI de-centres the technology, and the economic sustainability associated with it, in favour of greater human (and environmental) sustainability.

It should be noted that this article primarily argues for the adoption of such a convivial framework, rather than suggesting that this framework is already entirely constructed. The approaches to convivial AI discussed here represent practical steps towards ethical regulation of AI. In particular, the focus on developers' responsibility for their products throughout their production, use, and end-of-life states ensure a less permissive, more demanding idea of just regulation for AI than is seen through the use of Rawls (1999) and TJ. However, there remains more to be done in establishing comprehensive regulation. The following section raises some of these concerns, focusing on the interaction of AI and the media, discussing already emerging issues and the inability for the current, Rawlsian view of ethics to properly address these problems. These are issues which must be dealt with by future research, with a convivial approach presenting the best framework for achieving a practical, humane, and ultimately fair outcome.

What does this mean for media and communication?

As has been noted throughout, a great deal of the issues around regulating AI impact media and communications. Perhaps the most well-known issue (mentioned above) is that of AI models scraping data from across news sources, often being made exempt from copyright laws or simply infringing upon them (Grynbaum and Mac, 2023). Large AI companies are not only interested in existing media, however, but in producing media as well. De-Lima-Santos and Ceron (2022) find that the use of AI in news media largely relies on news organisations purchasing AI models from third-party companies, particularly large technology companies such as Alphabet. De-Lima-Santos and Ceron do note that AI produced text is seen less frequently in non-English languages, due to the English-centric nature of these models. While this could be taken to mean that non-English news media is not under threat by AI, it is more likely that this means non-English media will see an indirect harm by AI by being made more peripheral (de Roock, 2024).

Local news is particularly vulnerable to this kind of economic interference of AI. In the UK, for example, Reach PLC (the nation's largest local news company and owners of national papers such as the *Mirror* and *Express*, cumulatively reaching 69% of the country's population online) have been using AI since 2023, focusing on replicating articles across sites in a manner favoured by AI's ranking system (Gupta, 2024; Tribune, 2025) and ensuring content is considered 'appropriate' for advertisers (IBM, 2019). Similarly, Google's Digital News Initiative Innovation Fund awarded a grant to PA Media (then the Press Association) to develop their RADAR-AI (Gregory, 2017). RADAR-AI uses national level data to generate local news, including on children in custody, welfare payments, and council spending on temporary accommodation for homeless households (Care, 2025). AI companies are increasingly embedding themselves within the production and dissemination of news media, shaping what is considered 'valuable' in a story (i.e. how well it appeals to search algorithms and digital advertisers), and increasingly financialising an already precarious sector. This is worsened by the inaccuracies repeatedly found within such tools (Rahman-Jones, 2025), a limiting of journalists' editorial freedom (Thäsler-Kordonouri, 2025) and simply a lack of real knowledge about local areas (Tribune, 2025). This is felt by news readerships as well, with AI journalism undermining the trust readers have in the news, even when the content itself is still seen as being accurate and fair (Toff and Simon, 2023).

The risks posed by AI in news media therefore go well beyond making freelance journalism more difficult (Westerstrand, 2024), instead posing issues for the sector at every point of production and reception. Without a strong regulatory framework, one which considers the ways in which people either can or must interact with technology, it is difficult to imagine how this phenomenon will not worsen. This poses an issue for the deontological Rawlsian framework. Unless the decreasing number of jobs in journalism is considered a fundamental impingement upon the basic liberty of all citizens (although it seems unlikely that an equal job-to-demand ratio is a fundamental freedom and, if so, Rawlsians should take issue with *all* technologies since the industrial revolution), the rise of AI does not appear to threaten TJ's primary principles of justice. Further, provided that a government provides a reasonable justification for allowing AI use in such a manner, the issue of publicity as set out by Gabriel (2022) is averted. A convivial approach, conversely, prioritises the relationship that citizens have with technology (and with the societal institutions which own and deploy these technologies).

This approach to news media is not an aberration but rather is indicative of the wider perspective taken toward communicative and creative media by the AI sector. This is perhaps best exemplified by OpenAI CEO Sam Altman's recent interview at TED2025 (Cadwalladr,

2025). During this interview, Altman was asked if ChatGPT was committing IP theft, to which the present audience applauded. Altman simply responded, "you can clap about that all you want, enjoy... I think that people have been building on the creativity of others for a long time... I think there are incredible new business models that me and others are excited to explore" (TED staff, 2025). There is a clear desire from AI developers to further the economic precarity established through the platformised economics of media creation and dissemination (Drott, 2024), with AI developers becoming central to the political economy of creative expression in media.

Further, during this interview, Altman made a statement that exemplifies the underlying perspective on AI developers around creativity: "if you can't tell the difference, how much do you care?". This is in reference to being unable to know if AI is 'thinking' or just repeating data from its training set, but speaks to the wider implications of AI produced content in general (Altman himself prefaces this statement by describing it as an "incredible meta-answer") (TED, 2025). This statement articulates a direct response to concerns over the consequences of AI for human-centred creative outputs and the displacement of professional media careers: who cares? Altman's statement belies the perspective of AI developers around creation, i.e. all that matters is the end product, devoid of its context for creation or reason for being. This, in a sense, is a coherent viewpoint. If AI is a machine built upon and generative of consequences, it logically follows that those who create AI would be consequence focused as well. AI's perspective does not originate from the void; it is reflective of the viewpoint of its creators (which are in turn influenced by the products they create and so on).

This again returns us to the need for a consequence focused idea of justice to act as a regulatory counterweight to the ongoing AI-ification of the world. Donahue (2025) argues that there is value to maintaining a burden of collective moral achievement amongst a populous, i.e. the opportunity for individuals to come to and make their own moral decisions over time, as well as being a part of a larger society that makes moral judgements over time. Without the opportunity to make poor moral judgements, making good moral judgements is rendered less meaningful. Similarly, for media and communications, this article argues there is a collective creative achievement which would be undermined by loose non-human-centred AI regulation. This includes the individual level of being able to create art poorly, which gives greater meaning to art which is created well; and the collective level in which there must be opportunity to create art with potentially limited mass appeal but substantial value to those whom it does appeal (in the

context of AI, this may include non-English language content, something which has substantially wide appeal but is not necessarily captured by AI).

In TJ, Rawls (1999) constructs a fluid framework for society, which makes few normative claims about what justice looks like beyond provided it adheres to an ex-ante agreement on the fairness of society (and so the fairness and justness of its inequalities). Conversely, Illich's (1973) conviviality offers a framework for society based on normative ideas of how justice should be experienced and what just relations should look like in society, primarily based on our relations with one another and with technology. This framework therefore continues to make human-centred demands of justice ad tempus, in which justice is less concerned about previous agreements of what, in an abstract sense, is a fair and contractual agreement but instead sees justice as something to be constantly renegotiated in the face of new sociotechnical and material conditions. In order to preserve a thick sustainability of creative media output (and, indeed, improve current conditions), such an approach is necessary to counter the entirely outcome-driven ideology of AI. Without this, we risk an even greater enclosure of media creation, one which does not see an intrinsic value in the creation process (and the processes preceding creation, such as learning), instead seeing value only in quantifiable metrics such as data created and economic value. Seemingly, all that the Rawlsian approach can offer here is a demand for 'publicity', that we be made aware that AI is used and given justifications for this use, managing the expectations of citizens and so meeting TJ's criteria for fairness but evidently failing any measure of collective creative achievement.

Conclusion

With AI currently occupying such a large space in public discourse, particularly around how ubiquitous it should be in everyday life, it is vital to consider how this emerging technology should be regulated. It is for this reason that this article presents two opposing views when considering what constitutes a human-centred, ethical approach to AI regulation. The first is the liberal, Rawlsian view of justice as fairness. This position begins with the idea that justice should be distributive, established through the original position and difference principle. This is not to say that all should be equal. Rather, there *is* acceptance of an "appropriate division of advantages" by Rawls (1999, 15), provided that this distribution is generally acceptable to all when considered from the original position. Thus, the Rawlsian view is a deontological ethical framework and has been popular with many AI ethicists.

The alternative approach suggested in this article is based in Illich's (1973) notion of conviviality. Conviviality, as it is used here, is distinct from the Rawlsian view in that it is concerned with outcomes, rather than a more abstracted ethical position. Fundamentally, a convivial approach to regulation is based in the notion that technologies should exist to serve people, rather than people existing to serve technologies (for the benefit of a small number of people). In viewing AI as a technology, existing in a genealogy of other digital technologies, the conviviality approach emphasises that AI is malleable to human agency, rather than seeing AI as somehow inevitable. Conviviality therefore operates as a distinctly human-centred position, seeing the 'technology' itself as secondary to the social relations which surround it. This is considered in the context of AI through the matrix of conviviality and thick sustainability, which consider the importance of social and cultural sustainability alongside environmental sustainability. These two perspectives on justice are finally applied to news media, discussing the need for a comprehensive means of regulating AI in the news media and media more generally. Thus, through providing a more outcome-oriented framework that is interested in promoting the greatest level of virtue within society, the conviviality approach provides a practical and impactful starting point for regulating AI. This stands in contrast to the Rawlsian approach of seeking out the 'least bad' outcome and a hoped-for minimisation of disadvantage: in any human-centred ethics, we must demand more than this.

References

Amnesty International. (2024). Sweden: Authorities must discontinue discriminatory AI systems used by welfare agency. Retrieved 22 January, 2025 from https://www.amnesty.org/en/latest/news/2024/11/sweden-authorities-must-discontinue-discriminatory-ai-systems-used-by-welfare-agency/

Amoore, L., Campolo, A., Jacobsen, B., & Rella, L. (2024). A world model: On the political logics of generative AI. Political Geography, 113, https://doi.org/10.1016/j.polgeo.2024.103134

Aquinas, T. (1947). Summa Theologica. Einsiedeln, Switzerland: Benziger Bros.

Ashrafian, H. (2023). Engineering a social contract: Rawlsian distributive justice through algorithmic game theory and artificial intelligence. AI and Ethics, 3(4), 1447-1454. https://doi.org/10.1007/s43681-022-00253-6

Baker, S. (2024). Rise of ChatGPT and other tools raises major questions for research. *Nature*, *633*(8030), 5-5. https://doi.org/10.1038/d41586-024-02984-4

Barsotti, F., & Koçer, R. G. (2024). MinMax fairness: from Rawlsian Theory of Justice to solution for algorithmic bias. AI & Society, 39(3), 961-974. https://doi.org/10.1007/s00146-022-01577-x

Bay, M. (2024). Participation, prediction, and publicity: avoiding the pitfalls of applying Rawlsian ethics to AI. AI and Ethics, 4(4), 1545-1554. https://doi.org/10.1007/s43681-023-00341-1

Birch, M. (2024). Who did that? AI assisted targeting and the lowering of thresholds in Gaza. Medicine, Conflict and Survival, 40(2), 97-100. https://doi.org/10.1080/13623699.2024.2364937

Brevini, B. (2023). Global Digital Lords and Privatisation of Media Policy: The Australian Media Bargaining Code. Javnost-The Public, 30(2), 268-283. https://doi.org/10.1080/13183222.2023.2207427

Burkhardt, S., & Rieder, B. (2024). Foundation models are platform models: Prompting and the political economy of AI. Big Data & Society, 11(2), 1-15. https://doi.org/10.1177/20539517241247839

Cadwalladr, C. (2025). It's not too late to stop Trump and the tech broligarchy from controlling our lives, but we must act now. The Guardian. Retrieved 29 April, 2025 from https://www.theguardian.com/global/2025/apr/20/carole-cadwalladr-ted-talk-this-is-what-a-digital-coup-looks-like-its-not-too-late-to-stop-trump-and-the-silicon-valley-broligarchy-from-controlling-our-lives-but-we-must-act-now

Care. A. (2025). RADAR round-up: A look at the stories delivered by the RADAR-AI editorial team in the month of March. PA Media. Retrieved 29 April, 2025 from https://pa.media/blogs/editorial-data/radar-round-up-a-look-at-the-stories-delivered-by-the-radar-ai-editorial-team-in-the-month-of-march-2/

Coman, J. (2020). John Rawls: can liberalism's greatest philosopher come to the west's rescue again? The Guardian. Retrieved 30 April 2025 from https://www.theguardian.com/inequality/2020/dec/20/john-rawls-can-liberalisms-great-philosopher-come-to-the-wests-rescue-again

Crawford, K. (2021). Atlas of AI. New Haven, CT: Yale University Press.

de Roock, R. S. (2024). To become an object among objects: Generative artificial "intelligence," writing, and linguistic white supremacy. Reading Research Quarterly, 59(4), 590-608. https://doi.org/10.1002/rrq.569

de-Lima-Santos, M. F., & Ceron, W. (2021). Artificial intelligence in news media: current perceptions and future outlook. Journalism and media, 3(1), 13-26. https://doi.org/10.3390/journalmedia3010002

Department for Science, Innovation and Technology (DSIT). (2025a). Prime Minister sets out blueprint to turbocharge AI. Gov.uk. Retrieved 30 April 2025 from https://www.gov.uk/government/news/prime-minister-sets-out-blueprint-to-turbocharge-ai

Department for Science, Innovation and Technology (DSIT). (2025b). AI Opportunities Action Plan. Gov.uk. Retrieved 22 January, 2025 from https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan

Donahue, S. (2025). AI rule and a fundamental objection to epistocracy. AI & Society, 1-13. https://doi.org/10.1215/9781478027874

Drott, E. (2024). Streaming music, streaming capital. Durham, NC: Duke University Press. https://doi.org/10.1215/9781478027874

Estlund, D. (1996). The survival of egalitarian justice in John Rawls's political liberalism. The Journal of Political Philosophy, 4(1), 68-78. https://doi.org/10.1111/j.1467-9760.1996.tb00042.x

Foy, H. & Bradshaw, T. (2025). Nato acquires AI military system from Palantir. Financial Times. Retrieved 30 April, 2025 from https://www.ft.com/content/7f80b1bc-114c-4a00-ad06-6863fb435822

Franke, U. (2021). Rawls's original position and algorithmic fairness. Philosophy & Technology, 34(4), 1803-1817. https://doi.org/10.1007/s13347-021-00488-x

Franke, U. (2024). Rawlsian Algorithmic Fairness and a Missing Aggregation Property of the Difference Principle. *Philosophy & Technology*, *37*(3), 1-19. https://doi.org/10.1007/s13347-024-00779-z

Gabriel, I. (2022). Toward a theory of justice for artificial intelligence. Daedalus, 151(2), 218-231. https://doi.org/10.1162/daed_a_01911

Gregory, J. (2017). Press Association wins Google grant to run new service written by computers. The Guardian. Retrieved 29 April, 2025 from https://www.theguardian.com/technology/2017/jul/06/press-association-wins-google-grant-to-run-news-service-written-by-computers

Grynbaum, M. M., & Mac, R. (2023). The Times sues OpenAI and Microsoft over AI use of copyrighted work. The New York Times. Retrieved 29 April, 2025 from https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html

Gupta, N. (2024). How UK's Reach is using AI to help produce more content faster. WAN-IFRA. Retrieved 29 April, 2025 from https://wan-ifra.org/2024/10/how-uks-reach-is-using-ai-to-help-produce-more-content-faster/

Hammond, G. (2025). SoftBank and OpenAI back sweeping AI infrastructure project in US. Financial Times. Retrieved 22 January, 2025 from https://www.ft.com/content/48eb53a1-67ca-4509-8c62-401f0cf8b099

Hashimoto, T., Srivastava, M., Namkoong, H., & Liang, P. (2018, July). Fairness without demographics in repeated loss minimization, International Conference on Machine Learning, 1929-1938. https://doi.org/10.48550/arXiv.1806.08010

Heidari, H., Ferrari, C., Gummadi, K., & Krause, A. (2019). Fairness behind a veil of ignorance: A welfare analysis for automated decision making. Advances in neural information processing systems, 31, 1-17. https://doi.org/10.48550/arXiv.1806.04959

Heilinger, J. C., Kempt, H., & Nagel, S. (2024). Beware of sustainable AI! Uses and abuses of a worthy goal. AI and Ethics, 4(2), 201-212. https://doi.org/10.1007/s43681-023-00259-8

IBM. (2019). Reach and IBM launch Mantis, using IBM Watson to make brand safety smarter. IBM. Retrieved 29 April, 2025 from https://uk.newsroom.ibm.com/2019-10-17-Reach-and-IBM-launch-Mantis-using-IBM-Watson-to-make-brand-safety-smarter

Illich, I. (1973). Tools for Conviviality. London, UK: Calder & Boyars.

Jørgensen, A. K., & Søgaard, A. (2023). Rawlsian AI fairness loopholes. AI and Ethics, 3(4), 1185-1192. https://doi.org/10.1007/s43681-022-00226-9

Labiak, M. (2024). AI frenzy makes Nvidia the world's most valuable company. BBC. Retrieved 22 January, 2025 from https://www.bbc.co.uk/news/articles/cyrr40x0z2mo

Laden, A. S. (2003). The house that Jack built: Thirty years of reading Rawls. Ethics, 113(2), 367-390. https://doi.org/10.1086/342855

Larson, B. (2017). Gender as a Variable in Natural-Language Processing: Ethical Considerations. 3, 1-11. https://doi.org/10.18653/v1/W17-1601

Latouche, S. (2009). Farewell to Growth. Cambridge, UK: Polity Press.

Leaver, T., & Srdarov, S. (2024). Generative AI and children's digital futures: New research challenges. Journal of Children and Media, 1-6. https://doi.org/10.1080/17482798.2024.2438679

Li, M., Namkoong, H., & Xia, S. (2021). Evaluating model performance under worst-case subpopulations. Advances in Neural Information Processing Systems, 34, 17325-17334. https://doi.org/10.48550/arXiv.2407.01316

Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making ai less" thirsty": Uncovering and addressing the secret water footprint of ai models. Communications of the ACM, 1-10. https://doi.org/10.48550/arXiv.2304.03271

Lukes, S. (2005). Power: A Radical View (Second Edition). Basingstoke, UK: Palgrave Macmillan.

Macon-Cooney, B., Mökander, J., Stanley, L., & Decorte, R. (2024). A New National Purpose: The UK's Opportunity to Lead in Next-Wave Robotics. Tony Blair Institute for Global Change. Retrieved 22 January, 2025 from https://assets.ctfassets.net/75ila1cntaeh/55qX0nXXRWUSe7q4x4xntd/e93f86c94f9d7b1910005a1ca6d45bd5/2Xu2ib2C6zxbYJzxch9L1R--152708102024

Madianou, M. (2021). Nonhuman humanitarianism: when 'AI for good' can be harmful. Information, Communication & Society, 24(6), 850-868. https://doi.org/10.1080/1369118X.2021.1909100

Markham, A. (2021). The limits of the imaginary: Challenges to intervening in future speculations of memory, data, and algorithms. New media & society, 23(2), 382-405. https://doi.org/10.1177/1461444820929322

McMorrow, R., & Olcott, E. (2024). Nvidia's AI chips are cheaper to rent in China than US. Financial Times. Retrieved 22 January, 2025 from https://www.ft.com/content/10aacfa3-e966-4b50-bbee-66e13560deb4

Milmo, D. (2025). UK copyright law consultation 'fixed' in favour of AI firms, peer says. The Guardian. Retrieved 29 April, 2025 from https://www.theguardian.com/technology/2025/feb/11/uk-copyright-law-consultation-fixed-favour-ai-firms-peer-says

Miltner, K. M. (2024). "AI is holding a mirror to our society": Lensa and the discourse of visual generative AI. Journal of Digital Social Research, 6(4), 13-33. https://doi.org/10.33621/jdsr.v6i440456

Oi, M. (2024). Nvidia: US tech giant unveils latest artificial intelligence chip. BBC. Retrieved 22 January, 2025 from https://www.bbc.co.uk/news/business-68603198

Pogge, T. (1982). The Interpretation of Rawls' First Principle of Justice. Grazer Philosophische Studien, 15, 119, 119-147.

Rahman-Jones, I. (2025). AI chatbots unable to accurately summarise news, BBC finds. BBC. Retrieved 29 April 2025, from https://www.bbc.co.uk/news/articles/c0m17d8827ko

Rankin, J. (2025). EU accused of leaving 'devastating' copyright loophole in AI Act. The Guardian. Retrieved 29 April, 2025 from https://www.theguardian.com/technology/2025/feb/19/eu-accused-of-leaving-devastating-copyright-loophole-in-ai-act

Rawls, J. (1999). A Theory of Justice (Revised Edition). Oxford, UK: Oxford University Press.

Richardson, H. & Weithman, P. (1999). The Philosophy of Rawls: A Collection of Essays. New York, NY: Garland Publishing.

Robinson, B. (2025). Fears about AI job loss: New study answers if they're justified. Forbes. Retrieved 30 April, 2025 from https://www.forbes.com/sites/bryanrobinson/2025/02/09/fears-about-ai-job-loss-new-study-answers-if-theyre-justified/

Stone, P. (2022). In the Shadow of Rawls: Egalitarianism Today. Ethical Theory and Moral Practice, 25(1), 157-168. https://doi.org/10.1007/s10677-022-10272-1

Tacheva, J., & Ramasubramanian, S. (2023). AI Empire: Unraveling the interlocking systems of oppression in generative AI's global order. Big Data & Society, 10(2), 1-13. https://doi.org/10.1177/20539517231219241

TED Staff. (2025). What's next for AI? A conversation with OpenAI's Sam Altman – Live at TED2025. TEDBlog. Retrieved 29 April, 2025 from https://blog.ted.com/whats-next-for-ai-a-conversation-with-openais-sam-altman-live-at-ted2025/

Thäsler-Kordonouri, S. (2025). News automation in UK newsrooms, in N. Thurman, I. Henkel, S. Thäsler-Kordonouri, & R. Fletcher (Eds.), UK Journalists in the 2020s: Who they are, how they work, and what they think, 27-31.

The Tribune. (2025). AI is taking over local journalism. Does it matter? Sheffield Tribune. Retrieved 29 April, 2025 from https://www.sheffieldtribune.co.uk/ai-is-taking-over-local-journalism-does-it-matter/

Toff, B. & Simon, F. (2023). "Or they could just not use it?": The Paradox of AI Disclosure for Audience Trust in News. https://doi.org/10.31235/osf.io/mdvak

Van Dijk, J. (2008). In the shadow of Christ? On the use of the word "victim" for those affected by crime. Criminal Justice Ethics, 27(1), 13-24. https://doi.org/10.1080/0731129X.2008.9992224

Vetter, A. (2018). The matrix of convivial technology–assessing technologies for degrowth. Journal of cleaner production, 197, 1778-1786. https://doi.org/10.1016/j.jclepro.2017.02.195

Westerstrand, S. (2024). Reconstructing AI Ethics Principles: Rawlsian Ethics of Artificial Intelligence. Science and Engineering Ethics, 30(5), 1-21. https://doi.org/10.1007/s11948-024-00507-y