



# Towards just democracies in the age of pervasive digital systems—a Rawlsian approach

Salla Westerstrand<sup>1</sup>

Received: 14 November 2024 / Accepted: 17 July 2025  
© The Author(s) 2025

## Abstract

Digital systems are increasingly pervasive, influencing people's lives and societal structures. This paper uses John Rawls's concept of a basic structure of society to analyse the implications of pervasive digital systems on democracy and how to steer them to a more societally sustainable direction. The analysis shows that the profound impact of pervasive digital systems, the power of their providers, as well as the inability of mere regulation to tackle the challenges imply a need for us to attribute responsibilities of basic structure also to private organisations, such as providers of impactful technologies. However, organisations such as technology companies that are new to the basic structure have not aligned their operations with Rawlsian conception of justice, which contributes to democratic decline. As a response, the paper proposes a framework for the basic structure organisations towards developing societally sustainable pervasive digital systems.

**Keywords** Artificial intelligence · Algorithmic justice · Pervasive digital systems · Democracy · Digital democracy · Digital power

## 1 Introduction

Digital systems have an increasingly prevalent role in the range of opportunities we are offered in life. Artificial neural networks and other techniques of Artificial Intelligence (AI) can be used in deciding whether you will get a loan, an apartment, or your next job based on the success of previous applicants (O'Neil 2016). Generative AI (GenAI) systems are used to produce convincingly real-looking but fake content online that can affect political opinion-formation and freedom of elections (Coeckelbergh 2024; Mainz et al. 2024). After the release of popular interfaces for systems based on Large Language Models (LLMs), deepfakes have reached a new level of manipulative potential, which has been harnessed to stir already volatile democratic settings (e.g., deepfakes of related to Donald Trump's campaign event in July 2024 have fomented political chaos around the upcoming US presidential election: see, e.g., Harwell et al. 2024). These systems are underpinned by complexity, ubiquity, scalability, socio-technical dimensions and

general-purpose nature, which Grover and Lyytinen (2023) call *the pervasive digital*.

Pervasive digital systems come with convoluted questions about their impacts on the moral and political fundamentals of democracy. Although there are many different ideals of democratic societies and a myriad of different designs of democratic governments, in the heart of all these perspectives lie the idea of *rule by the people*. Democratic legitimacy of the use of power by public governments is thus partly based on the idea that people (the *demos*) hold the ultimate power to rule (*kratos*), whether through electing representatives or participating directly in collective decision-making. On the one hand, pervasive digital systems can help include more people in collective decision-making (Bernholz et al. 2021), facilitate online deliberations (Bonechi 2024; Alnemr 2020) and potentially decrease cognitive biases occurring in human decision-making (König & Wenzelburger 2020), which can all strengthen the democratic legitimacy of collective decisions. They could be used in community-building and thus strengthening the dialogue between actors holding varying political opinions (Coeckelbergh 2024, pp. 114–115). On the other hand, studies indicate that the use of algorithmic systems can lead to biased decisions (Kordzadeh and Ghasemaghaei 2022) and decrease the level of human autonomy in a way that

✉ Salla Westerstrand  
salla.k.westerstrand@utu.fi

<sup>1</sup> Turku School of Economics, Information Systems Science, University of Turku, Turku, Finland

threatens fundamental rights (Coeckelbergh 2022; Jungherr 2023; Kilovaty 2019). Our use of AI on social media could lead to filter bubbles that erode rather than build democratic deliberation (e.g., Karimov 2024; Rodilosso 2024). AI-generated deepfakes are used to manipulate voters, endangering freedom of elections (Łabuz and Nehring 2024). The current directions can be seen to indicate a trend towards eroding rather than supporting democratic governance (Coeckelbergh 2024, p. 118; Susskind 2022).

Whereas some scholars have contributed to the accumulation of understanding of the impacts these systems can have on democracy (Coeckelbergh 2024; Jungherr 2023; Manheim and Kaplan 2019; Nemitz, 2018; Simons 2023; Susskind 2022), we have only scratched the surface and revealed more questions than answers. What happens when the above-discussed impacts take place? If pervasive digital systems erode democracy, what will our future forms of governance look like and what are their moral foundations? Who will hold the power, and what is the role of the people in relation to public and private institutions? And finally, to highlight an aspect that has gained in prevalence after the popularisation of AI systems: will we find such societies *just*?

In this paper, I argue that the proliferation of pervasive digital systems influences fundamental structures of democratic societies, which challenges our ability to sustain contemporary democracies and thus requires our attention and action. I adopt a critical perspective and approach the topic through the lens of John Rawls's theory of justice. I discuss how the changes these systems provoke influence what Rawls calls the *basic structure of society* that holds the primary responsibility over fair distribution of justice in democratic societies. Rawls intended his theory to form the most appropriate moral basis for democratic societies (Rawls 1971, p. viii) and proposed a set of principles for institutions that play a key role in the distribution of justice in society. His theory has been shown relevant in the context of ethics of emerging technology (Douglas 2015; Gabriel 2022; Keeling 2018; Leben 2017, 2018), as well as organisational and business studies (Cohen 2010; Fia and Sacconi 2019; Wood 1994), not to mention the body of literature that has drawn from Rawls in democratic theory and theories of justice (e.g., Chandler 2023; Nozick 2013; Sen 2010, only to mention a few), which makes it a theoretically interesting basis for studying the impacts of pervasive digital systems on democratic societies.

This paper is thus a conceptual exploration into the impacts of pervasive digital systems on democratic societies and how to address them. In this exploration, Rawls's basic structure serves as a tool that helps understand the nature of societal changes provoked by pervasive digital systems and actions needed to change the course towards more just digital democracies that are resilient and thus can absorb the

shocks of digital disruption. This is thus not an attempt to develop a model for *global governance*, as Rawls's theory still faces several limitations when it comes to its applicability to global (and especially non-democratic) settings, but a step in understanding societal sustainability of pervasive digital systems in democratic contexts.

Although much of the discussion in this paper will be dedicated to the impacts of AI systems and digital platforms with shown impacts and a considerable amount of available literature, it needs to be recognised that many other technologies come with societal implications, as well. Therefore, I have adopted a relatively inclusive technological scope. Rather than defining precise technical details of systems discussed, I address modern digital technologies underpinned by complexity, ubiquity, scalability, socio-technical dimensions and general-purpose nature, which Grover and Lyytinen (2023) call pervasive digital systems. This includes technologies such as AI systems, digital platforms, algorithmic decision-making assistants and chatbots currently in use, but also systems underpinned by similar characteristics that have not yet seen the light of day. In both contemporary academic and popular discussion, these systems are often grouped under the umbrella of AI. For the sake of coherence, I follow this terminological grouping and use the term AI in this paper in a broad sense, including advanced rule-based systems, algorithmic systems deploying machine learning, as well as neural networks and LLMs alike. I use the term societal sustainability to refer to the ability to maintain—or sustain—democratic structures that are beneficial to citizens. Therefore, societally sustainable digital systems are those that do not erode beneficial societal structures but enable their maintenance or improvement over time.

I also do not suggest that the future of democracy depends exclusively on technological developments. Scholars have extensively studied phenomena such as democratic backsliding (e.g., Bermeo 2016) and the crisis of democracy (e.g., Przeworski 2019), which serve as a backbone for us to understand the broader democratic context where digital systems are being introduced. In the light of the existing literature, however, there is reason to suspect that the role of pervasive digital systems in democratic change merits more attention, as will be shown below.

In what follows, in Sect. 2, I discuss Rawls's concept of the basic structure of society and argue for the inclusion of private organisations in the basic structure. In Sect. 3, I analyse the implications of digital technologies on democracy through the lens of the basic structure. This is followed by a discussion on the potential directions where the changes provoked by digital systems are steering democracies (Sect. 4.1) and a framework for organisational governance of pervasive digital systems towards a societally sustainable direction (Sect. 4.2). In Sect. 5, I discuss the results of this conceptual exploration, and in Sect. 6, I draw conclusions.

## 2 Basic structure for just digital democracies

Pervasive digital systems increasingly influence people's opportunities in life. But who should take accountability over securing fair distribution of those opportunities? Rawls offers a lens through which we can explore the role of organisations in securing social justice amid digitalisation, which helps us reveal whether the current institutional structure around digitalisation is societally sustainable.

For Rawls, the basic structure of society holds the responsibility over securing social justice. This basic structure comprises '[the] political constitution and the principal economic and social arrangements' because of their profound impact on people's chances in life (Rawls 1999, p. 6). Since we cannot guarantee equal conditions in life for everyone, Rawls argues that his principles of justice are primarily targeted to the basic structure responsible for the distribution of necessary inequalities. According to Rawls, the purpose of these institutions is:

*'to secure just background conditions against which the actions of individuals and associations take place. Unless this structure is appropriately regulated and adjusted, an initially just social process will eventually cease to be just, however free and fair particular transactions may look when viewed by themselves'* (Rawls 2005, p. 266).

The basic structure thus comprises influential organisations and societal structures, forming an 'important complex of institutions, given the deep and pervasive nature of its social and psychological effects' (Rawls 2005, p. 260). These institutions should follow Rawls's principles of justice, which are the ones he argues would be agreed upon in the fairest possible setting. These principles are:

- a) *'Each person has an equal right to a fully adequate scheme of equal basic liberties which is compatible with a similar scheme of liberties for all.'*
- b) *'Social and economic inequalities are to satisfy two conditions. First, they must be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they must be to the greatest benefit of the least advantaged members of society.'* (Rawls 2005, p. 291.)

However, defining which institutions belong to the basic structure is not obvious. The basic structure is intended as a conceptual tool and is here used as such. However, to apply it effectively in the context of digitalisation, some clarifications are needed. First, the role of businesses and other private organisations and whether they thus should

be subjected to the principles of justice merits discussion. For example, in *Political Liberalism*, Rawls states that it is the 'political constitution, the legally recognised forms of property, and the organization of the economy, and the nature of the family' that belong to the basic structure (Rawls 2005, p. 258). However, soon after, he takes libertarianism as a contrastive example to his conception of the basic structure. He argues that libertarians consider state like any private association, and that the relationship between citizens and states are 'just like their relation with any private corporation with which they have made an agreement' (Rawls 2005, p. 264). He continues: 'By viewing the state as a private association the libertarian doctrine rejects the fundamental ideas of the contract theory, and so quite naturally it has no place for a special theory of justice for the basic structure' (Rawls 2005, p. 265). This implies that private corporations do not necessarily fulfil his requirements of institutions belonging to the basic structure.

According to the *coercive account*, only institutions with legally coercive power should be included (Berkey 2021). Singer (2015), for example, argues that corporations should be excluded from the basic structure because that would be in line with Rawls's arguments for freedom of association and individual transactions. Instead, to secure a just society, businesses should be regulated from above by the basic structure institutions (Singer 2015, p. 17). On the other hand, some consider that the main criterion for determining the institutions belonging to the basic structure should be the *profoundness of the impacts* those institutions have on people's lives (Berkey 2021). For example, Blanc and Al-Amoudi (2013) argue that the current context of weakening welfare states justifies the revision of which institutions belong to the basic structure. They conclude that private companies should be part of the basic structure because they 'bear an effect on the expectations of primary goods associated with relevant (that is, non-voluntary) social position' (Blanc and Al-Amoudi 2013, p. 519). They thus base their argument on the purpose of the basic structure given by Rawls: distribution of social justice.

A third perspective to the debate is offered by Berkey (2021), who argues that the applicability of Rawls's theory to private corporations is not a matter of whether they belong to the basic structure. Rather, the evaluation should be based on whether their action affects people's ability to act as free and equal beings, benefiting from the primary goods (Berkey 2021, pp. 197–198; 205), which is one of the main goals of Rawls's theory. Berkey's perspective, however, risks bending Rawls's theory beyond Rawls's appreciation, as it contradicts with the original assumption that only the basic structure should be subjected to the principles of justice. Yet, it brings to discussion the need to revisit our understanding of the basic structure in a context where power

relations are shifting, such as amid digital disruption we are currently experiencing.

When using Rawls's theory in analysing real-life phenomena, attempting to define an *empirical manifestation* of a basic structure—i.e., a list of institutions that in current democracies belong to the basic structure—would be unrealistic and unhelpful at best. Still, in the present analysis, the question of private organisations is essential as it defines the very subject of justice and the attributed moral duties: should AI businesses also be held accountable over maintaining unfair distribution of inequalities, or is that the sole responsibility of public governments? Who needs to act to steer the development of digital technologies towards societally sustainable direction? Is coercive power through binding regulation the only way forwards, and does it really suffice for a fair society intended by Rawls to take place?

One of the most recent contributions to the discussion comes from the context of digital technologies. Gabriel (2022) argues that in modern societies, the basic structure should be 'best understood as a composite of sociotechnical systems: that is, systems that are constituted through the interaction of human and technological elements' (p. 220). For Gabriel, AI influences this structure, which subjects its design, development and deployment to the realm of principles of justice as fairness. He gives two justifications: firstly, the societal functions that are relevant in guaranteeing the just background conditions for people are increasingly algorithmically mediated. The second justification revolves around the profound impact these systems have on people's lives. As a result, Gabriel argues that demanding justice from AI systems providers, as well as the use of AI in hands of the basic structure organisations, is justified.

Similar position has also been adopted by Westerstrand (2024), who proposes that in the age of AI, also private organisations should belong to the basic structure of society and align their activities with a set of Rawlsian principles of AI ethics because of 'a) their profound impact on people's lives, b) the way in which they limit people's ability to act as free and equal beings, and c) their key role in securing just background conditions for people and associations to function in the society' (p. 4). Thus, there is reason to assume that also private organisations, including providers of pervasive digital systems, should participate in securing social justice.

For Rawls, justice and democratic resilience go hand in hand, as fulfilling the principles of justice requires democratic governance, and democracy requires fulfilment of the principles of justice. As Cohen (2003) explains, the contents of Rawls's principles—political liberties of the first principle in the forefront ('the right to vote and to hold public office', as in Rawls 1999, p. 53)—presumes a democratic society. Moreover, Cohen demonstrates how the foundations Rawls's just society overlap with

democratic fundamentals: Rawls sees people as equals in society based on their equal moral capacity to 'understand requirements of justice that provide the fundamental standards of public life' (Cohen 2003, p. 96). He continues: 'And a democratic society is a society of equals whose members are regarded in the political culture as having that capacity' (Cohen 2003, p. 96). Finally, Cohen notes how Rawls's idea of fairness in legislation is connected to the deliberative conception of democracy: for Rawls, laws are not just without a sufficient level of citizen participation in the process (Cohen 2003, pp. 101–103).

Rawls was not alone in highlighting the importance of justice in democratic societies. Rather, the idea of interdependence between justice and democracy has been shared by several scholars (Buchanan 2002; Pettit 2012, pp. 133–134). Considering the increasing body of literature that studies algorithmic fairness as a response to recent technological developments (e.g., Franke 2021; Hacker 2018; Mitchell et al. 2021) with numerous ethics guidelines given to AI systems that call for justice and fairness in their development and deployment (e.g., Franzke 2022; Jobin et al. 2019), taking Rawls's basic structure as the key concept offers an arguably relevant lens for the present paper. It is hence this role of justice as the moral foundation of democratic societies that allows us to see how pervasive digital systems impact democracy: if they change the basic structure in a way that it cannot (or does not) secure just background conditions for people and associations, then, in Rawlsian terms, they erode democracy. On the other hand, if they erode democracy, they also erode structures needed for justice to take place—a vicious cycle that takes us further away from just digital societies.

It needs to be noted that in his oeuvre, Rawls did not offer a fully clear vision to what an ideal democracy would look like. He also left the relationship between justice and democracy for others to interpret, and thus the above reflects our interpretation based on the reading of his theory and secondary literature. Rather than an empirical description of institutional structures, Rawls's theory is essentially a conceptual tool for analysing societal structures, their justice and moral duties different institutions that hold power over people's lives should have. Hence, along similar lines, I use the basic structure as a tool to make sense of the societal changes provoked by pervasive digital systems, what they mean for democratic societies, and how organisations should react to contribute to the sustainability of just democratic societies.

Building on this theoretical background, in the next Section, I discuss the role of pervasive digital systems in this change and what it implies for today's democratic societies, which demonstrates the potential of this interpretation to serve scholars studying the impacts of technology—whether in information systems, computer science, science and technology studies or beyond—on democratic societies.

### 3 How pervasive digital systems change the basic structure

In the context of emerging technologies, we have witnessed a shift of power over people's lives from democratic institutions to corporations (e.g., Coeckelbergh 2024; Muldoon 2022; Susskind 2022). Those deciding on the decision rules and weights of algorithms in digital systems seem to have slowly gained more influence over life-changing decisions, which thus are no longer made under democratic scrutiny but behind closed doors, or, in the case of complex algorithms, in black boxes.

In this context, the idea that *all* private organisations should be excluded from the basic structure of society does not seem to hold. To see why, we need to look at the societal role of companies providing pervasive digital systems and their ethical impacts, i.e., the very background conditions against which the actions of individuals and associations take place. Drawing from literature on digital ethics, I do so by analysing the profoundness of impacts these systems have on people's lives (Sect. 3.1), the power of providers of pervasive digital systems to impose such impacts (Sect. 3.2), and the role of regulation by the basic structure as a proposed remedy by the coercive account to secure a just society (Sect. 3.3).

#### 3.1 Profound impacts of pervasive digital systems

Today, pervasive digital systems have been deployed in public governance to define what kinds of services people receive as support for employment in Estonia (Brioscú et al. 2024, p. 40). They have been used to predict the length of people's sentences in courts in the US (Simons 2023, pp. 38–50) and to determine who gets a loan or a job all around the world (O'Neil 2016). People's irises are being scanned by a private company as an exchange for cryptocurrency, with little details provided about the purpose and intention of the use of such database, targeting primarily people in already vulnerable locations (Gent 2023). Algorithms developed in the public sector have been used in predicting childcare benefit frauds in the Netherlands (Heikkilä 2022). The results given by all these algorithms are being used as a support for decisions and other activities that arguably come with profound impacts on people's lives both by private and public sector institutions.

Similarly, the conditions for free speech are increasingly moderated by social media companies when people communicate on their digital platforms (York 2021). Those who design the algorithms for these platforms determine which topics we see on our feeds, whose voice we hear,

whether we get banned or gratified on the public forums these platforms maintain, and the conditions for people to earn their living in gig economy ecosystems (Muldoon 2022; Muldoon and Raekstad 2023; Susskind 2022). If we define coercion as an act of persuading or forcing someone to do something using threat of punishment, it seems that the power of social media companies has reached a coercive dimension: If we fail to comply with the rules dictated by, e.g., Meta on its digital platforms, our content will be deleted, or we can be entirely blocked from their platforms. As a significant volume of public deliberation happens on these proprietary platforms and established communities within, it seems unreasonable to suggest that the possibility to switch to another product makes this use of power less coercive.

All the above-discussed systems pervade our everyday lives. From the perspective of Rawls's basic structure of society, it seems that the use of pervasive digital systems has a profound impact on key life events, and this impact often emerges in situations where people do not have effective means to choose otherwise. The unemployed, those in need for childcare benefits, and people being judged in courts or even job seekers—all of which are examples discussed above—can rarely choose whether they are subjected to these algorithms. Hence, the grounds on which these decisions are made shape the background conditions against which people operate in society and the fairness thereof.

To exemplify, in the case of the infamous recidivism algorithm, COMPAS, a group of investigative journalists of ProPublica studied the algorithm and found out it was more likely to predict an overly high recidivism rate when the defendant in court was black, and more likely to underestimate the recidivism rate when the defendant was white (Larson et al. 2016; for discussion, see Simons 2023, p. 38–46). Incidents have been reported with big social media providers' moderating policies that have discriminated against certain groups of people—e.g., when Facebook banned the use of word 'Palestinian' when creating a new site on its social media platform, while 'Israeli' and 'Afghan' were still allowed (York 2021, pp. 52–54). The Dutch taxation office used a harmfully biased algorithm to detect child care benefit fraud, falsely accusing families in precarious positions with drastic consequences (e.g., Heikkilä 2022).

It thus seems that the design of these algorithms influences the justice of the background conditions we have in democratic societies. But which institutions hold the power over deciding on the design of the systems? Whose values form the normative basis of these decisions?

### 3.2 Shift of power over people's lives to tech providers

The ones holding power over the design of impactful algorithms are often institutions outside democratic scrutiny. For example, in the case of COMPAS, the company that developed the algorithm, Northpointe (now Equivant), responded to ProPublica with saying they had used different fairness metrics to define fairness for their recidivism algorithm than ProPublica, and thus argued their algorithm was fair and worked as it should have worked (Simons 2023, pp. 42–43). The company thus held the power over defining the concept of fairness used by courts when COMPAS was used, rather than an institution enjoying democratic legitimacy. Although the sentences were not given by a machine but by a human judge, neither the courts nor people being judged had visibility to the grounds of the algorithm's recommendations, making it impossible to critically assess its potential biases. Indeed, the debate about which model of fairness would be preferable took place between ProPublica and Northpointe rather than the court and the defendant, as discussed by Simons (2023, pp. 38–46).

As another example, Finnish media reported how Meta's platforms have increased their censure on political contents, taking as an example a Finnish clothing brand whose print shirts with texts 'human rights' and 'diversity' Instagram's algorithms caught as too political, hiding them from the followers' feeds and banning paid promotion (Nieminen 2024). Doing so, the platform dictates which topics can be discussed and which not. Considering that social media is for some an important source of news and political content (e.g., Bode 2016), algorithmic content moderation has a profound impact on these platforms as forums for deliberation and political speech. Whereas the Finnish example was only a setback for a clothing brand, media have discussed the implications of such censure on political deliberation under the many elections held in 2024 (e.g., De Guzman 2024), or the plans of Trump's administration in the US to 'strip foreign nationals of the visas that allow them to remain in the country' based on 'AI analysis of their social media accounts'—depending on the acceptability of ideologies their social media represents (e.g., Stokel-Walker 2025). This demonstrates the broader impacts of pervasive digital systems on people's ability to participate in political discussion and forming political opinions free from manipulation by external interests.

In the context of the digital economy, algorithms are used to exercise mass surveillance (Zuboff 2015, 2019), which affects how we behave even when surveillance does not lead to concrete actions (Coeckelbergh 2022, p. 14). Whilst the main concern around surveillance was once almost exclusively reserved for state-operated surveillance mechanisms, an increasing interest is focused on companies providing

surveillance tools for governments, but also exercising surveillance themselves. It seems that it has become harder even for technology companies to justify why they should be able to exercise all this power without democratic scrutiny, which is why some have taken on initiatives to 'democratise' their AI systems through, e.g., citizen consultations (e.g., Anthropic 2023). These attempts, however, still misunderstand the fundamentals that would be required for their use of power to enjoy democratic legitimacy, as the companies eventually still hold the power over deciding whether citizens' voice will be heard, who will be involved in participation, and whether citizens will ever be heard again in the next round of policy changes and model updates.

This shift of power has been noted by several scholars. For example, Coeckelbergh (2024) argues that AI undermines democracy due to its impact on the distribution of power: 'the future of AI is decided by corporate managers and their shareholders. [...] If big tech is more powerful than many countries, in what sense does the *demos*, the people, still rule?' (p. 49). Although the issue of the legitimacy of corporate power is much older than AI, it seems to have been amplified because of the concentration of power to a few powerful actors in the tech field. As Susskind (2022) points out, such course is rarely fully dictated by one individual actor but many, leading to a situation where 'a thousand technologies might combine to form an atmosphere of oppression, even if, individually, each technology seems perfectly reasonable' (p. 80). AI companies also seem to have gained control over the discourse around the benefits and harms of their own technologies in journalistic media (Westerstrand et al. 2024), which has traditionally had a role of a watchdog to increase transparency and critically observe the use of power in democracies.

Therefore, from a Rawlsian perspective, private organisations now play a key role in the distribution of justice in society and thus should alleviate rather than upkeep or amplify social injustices. Instead of inspecting individual decisions made by specific institutions and evaluating whether they are fair, the basic structure should be seen as a holistic entity, the institutions of which and their interrelations together shape the just background conditions in a society. Indeed, in *Political Liberalism*, Rawls writes:

'[F]air background conditions may exist at one time and be gradually undermined even though no one acts unfairly when their conduct is judged by the rules that apply to transactions within the appropriately circumscribed local situation. The fact that everyone with reason believes that they are acting fairly and scrupulously honoring the norms governing agreements is not sufficient to preserve background justice. [...] We might say: in this case the invisible hand guides things in the wrong direction and favors oligopolistic

configuration of accumulations that succeeds in maintaining unjustified inequalities and restrictions on fair opportunity.’ (Rawls 2005, p. 267.)

In this light, in societies influenced by pervasive digital systems, the argument according to which private organisations should be fully excluded from the basic structure (Singer 2015, p. 17) seems difficult to maintain. Given the profound impact of the systems they develop (Sect. 3.1), it is difficult to argue that the power they hold is so inherently different that they should not bear responsibilities similar to those of democratic institutions.

It is worth keeping in mind that negative impacts are not reserved only for proprietary systems. As the childcare benefit scandal in the Netherlands demonstrated, public sector organisations also develop algorithms themselves with potentially devastating consequences to people’s lives (Heikkilä, 2022). Moreover, not all impacts of proprietary or publicly developed algorithms are negative—there are many examples of ways in which AI systems can be beneficial for humans and societies (e.g., Du et al. 2023; Javaid et al. 2023; König and Wenzelburger 2020). Defining which technology providers offer solutions that come with profound enough impacts to secure them a spot in the basic structure is also a task with obvious difficulties, as there are no metrics available, nor clear definitions of what makes an impact *profound*. Amid the current lack of discussion around such definitions and uncertainty around the impacts of pervasive systems overall, it can seem tempting to steer attention towards institutions that have traditionally set the rules and responsibilities of different actors in societies: the regulators.

### 3.3 Cannot we just regulate? A response to coercive account

Finally, I return to the role of regulation in the basic structure. As discussed above (Sect. 2), the proponents of the coercive account suggest that private organisations do not belong to the basic structure but should rather be regulated by the basic structure to reach social justice (e.g., Singer 2015). Indeed, in a complex society, it sounds tempting to allow private organisations the freedom to innovate our future in the frameworks set in place by regulators, without responsibility over social justice. Could we respond to the profound impacts of pervasive digital systems (Sect. 3.1) and the shift in power relations (Sect. 3.2) with regulation rather than attributing responsibility to private organisations?

As noted by Rawls, one of the tasks of the basic structure itself is to regulate, as it ‘enforces through the legal system another set of rules that govern the transactions and agreements between individuals and associations’ (Rawls 2005, p. 268). However, he also notes that the basic structure itself needs to be ‘appropriately regulated and adjusted’

(Rawls 2005, p. 266) for it to fulfil its role of securing the just background conditions for democratic societies. For Rawls, the ability to regulate is not a prerequisite for basic structure institutions, as he includes family and religious institutions in the basic structure despite them having no legislative power (e.g., Rawls 2005, p. 256). Hence, all institutions—whether public or private, whether inside or outside the scope of the basic structure—are subjected to regulation, making legislative power an unhelpful qualifier for defining the contours of the basic structure. Instead, it forms a complex entity of institutions with a unifying characteristic of securing the just background conditions in a society—indifferent to whether they hold regulatory power.

Moreover, looking at the current regulatory frameworks around pervasive digital systems, it seems that regulatory institutions alone cannot guarantee social justice. Despite regulations stemming from international human rights conventions to national, sectorial regulations, AI systems have already been deployed in ways that violate Rawls’s basic liberties through, e.g., mass surveillance (Almeida et al. 2022; Zuboff 2015, 2019) and behavioural steering (Coeckelbergh 2022, pp. 16–20), biases leading to unfair discrimination (e.g., Kordzadeh and Ghasemaghahi 2022; Tolan 2019) and influence on freedom of elections (Jones 2023; Łabuz and Nehring 2024). In the EU, such impacts have been addressed with regulatory initiatives around data and AI, the recently adopted EU AI Act being perhaps the best-known example. As shown by Westerstrand (2025), however, despite a goal-setting that supports Rawls’s basic liberties (i.e., protection of health, safety and fundamental rights), the effectiveness of the EU AI Act in securing Rawls’s principles of justice appears limited. It seems that in today’s democracies, regulators are challenged by pervasive system characteristics and their complex supply chains that involve a variety of actors and components, the combined effects of which often become fully understood only after deployment (Scherer 2015). Introduction of generative AI applications has shown how seemingly innocuous general-purpose systems can easily be used for malicious purposes in contexts including cybersecurity (Schmitt and Flechais 2024; see also uses for cyber security Kaur et al. 2023) and manipulation (e.g., Coeckelbergh 2022; Kilovaty 2019; Klenk 2024). AI-powered chatbots and digital duplicates come with ethical implications (Danaher and Nyholm 2024; Stahl and Eke 2024), yet they are available for individuals to purchase or to use for free without effective regulatory intervention.

What is more, companies that hold power over designing the most impactful technologies have an active role in regulatory initiatives. Whereas corporate involvement in regulation is a welcomed way to ensure inclusion of a variety of interest (see, e.g., Christiansen 2024 on corporatism), excess influence can lead to erosion of democratic legitimacy of the regulation and over-emphasise corporate

interests over public interest. In the process of drafting of the EU AI Act, Corporate Europe Observatory revealed a significant increase in Big Tech lobbying directed to European legislators towards the end of the EU AI Act negotiations (Corporate Europe Observatory 2023a), potentially leading to weaker protection of people's rights in favour of business interest (Corporate Europe Observatory 2023b). Similar tendencies towards regulatory capture were already observed in the context of the Digital Markets Act and the Digital Services Act (Laux et al. 2021). As discussed by, e.g., Chomanski (2021), even calls for heavier technology regulation can increase corporate effort in regulatory capture, putting to question the effectiveness of introducing *more* technology regulation as a response to lobbying.

One could still argue that today's regulators have simply failed at their task of properly regulating technology providers, which is a legitimate concern. In a capitalist economy where businesses are incentivised to maximise shareholder value over other purposes, demanding that companies contribute to democratic sustainability can seem like an empty hope. I argue, however, that the idea of inherent conflict between societal sustainability and profitable business is a misconception. For example, Bednar and Spiekermann (2024) show how ethics-based innovation can generate more value ideas with more variation and depth compared to traditional product roadmapping. Including ethical considerations into digital business has been shown to support business in several instances through, e.g., increased customer satisfaction and trust (Shah et al. 2024). Industry studies have also pointed towards the importance of ethical and governance structures in organisations when adopting AI—Deloitte's 2024 survey showed that 89% of all surveyed C-suite leaders agreed that their organization's 'existing ethical frameworks and governance structures encourage and support technological innovation' in their organisations.<sup>1</sup> The insufficiency of regulation is also highlighted in academic literature on AI governance, where laws and governmental regulation are only some of the measures in efforts towards effective organisational AI governance (Birkstedt et al. 2023; Mäntymäki et al. 2023). Governance measures are thus ideally designed to advance the organisation's strategic goals and vision, which supports their duty towards shareholders to generate profit. Dedicating all the responsibility over securing just democracies to regulators in times of pervasive digital should thus not be justified by alleged incompatibility between capitalism and social justice, as taking responsibility over the broader impacts of

business practices is likely to enhance rather than hurt the competitiveness of these companies.

Moreover, while more regulation can solve some of the persisting challenges (see, e.g., discussion in Susskind 2022, p. 119 onwards), the complexity of pervasive digital systems challenges regulatory mechanisms (Moses 2013; Scherer 2015), and the current established position of power of technology providers make relying only on regulation ineffective in striving towards Rawlsian ideal of a fair democracy. Hence, viewing today's providers of pervasive digital systems as basic structure institutions with responsibility over social justice seems like a more effective way of reaching social justice in digital democracies compared to only regulating them. Next, I turn the attention towards how these impacts on the basic structure influence democratic resilience, and how we can steer the basic structure forwards more societally sustainable direction.

## 4 Implications for democracy and a way forward

Up to this point, I have discussed how the popularisation of pervasive digital systems has challenged the ability of our current institutional structures to secure social justice due to the profound impacts these systems have on people's lives (Sect. 3.1), and how these impacts are often controlled by private organisations rather than democratic institutions (Sect. 3.2). I have discussed the inability of regulation to address these issues in full and called for moral responsibility to the providers of pervasive digital systems as institutions belonging to the basic structure of society (Sect. 3.3). Considering that for Rawls, even his basic liberties can and sometimes should be modified to fit the 'social, economic and technological' context of the society under scrutiny (Rawls 1999, p. 54), this analysis has provided us with understanding of technological change that can provoke a need for such consideration. In this section, I discuss the implications of this analysis on democratic societies and how we could steer the development of pervasive digital systems towards a more sustainable direction.

### 4.1 Democracy in the age of the pervasive digital

'So in the end there are some cases in which the collapse of democracy is manifest, marked by some discrete event, but there are some in which democracy slides down a continuous slope, so not only do we not have discrete markers but we can reasonably disagree about whether a particular regime is still democratic or already past the point of no return' (Przeworski 2019, 26).

<sup>1</sup> Deloitte (August 2024), Technology, Trust, Ethics. Leadership, governance and workforce decision-making about ethical AI: C-suite perspectives. Available at: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/about-deloitte/us-ent-pulse-survey.pdf>. Last accessed 4 April 2025.

Above, I have posited that as a response to recent technological developments, the responsibility over social justice should be shared between governmental institutions and other organisations belonging to the basic structure, including providers of pervasive digital systems. Yet, it seems that some of these organisations have not recognised such responsibility. This has come with ethical implications, such as erosion of Rawlsian basic liberties like freedom and human autonomy (Coeckelbergh 2022; Prunkl 2024; Sio 2024) and political liberties (Jones 2023; Łabuz and Nehring 2024), as well as equality of opportunity and the difference principle (Westerstrand 2024). As Gabriel (2022) notes, AI systems challenge Rawls's requirement of publicity, according to which the functioning, or the 'grounds' of the basic structure institutions should be open for public scrutiny.

If the basic structure is failing in its moral duty to ensure justice as fairness, what does that mean for the future of democracy as a form of governance? For Rawls (Rawls 1999, p. 6), the basic structure is the very subject of justice. If it fails its moral duty, we can only state the absence of social justice. If the basic structure does not live up to its task and fails to protect basic liberties, equality of opportunity and the worst-off members of society, we lack foundations for *people to rule*, namely, the very fundamental normative justification of democratic legitimacy.

Democratic theorists have for long argued that democracy is in crisis (Przeworski 2019), or that we are witnessing democratic backsliding (Bermeo 2016). Adam Przeworski (who is a proponent of minimalist democratic ideal) brings forth erosion of party systems, the rise of right-wing populism and a general decline in support for democracy demonstrated in surveys as ongoing phenomena that feed democratic crises in election-based democratic systems (Przeworski 2019). Furthermore, although traditional and sudden forms of democratic backsliding, such as coups d'état seem to be in decline, Bermeo (2016) noted how phenomena such as election harassment and manipulation, as well as executive aggrandizement were challenging democratic regimes. Eight years later, the introduction of generative AI systems has amplified the issues related to election manipulation (e.g., Łabuz and Nehring 2024), and the opacity of algorithms used in public governance are making it increasingly difficult to challenge executive politics, thus undermining democratic legitimacy (Beckman et al. 2024; Erman and Furendal 2024).

Therefore, democratic backsliding or crises of democracy are multifaceted phenomena. As noted by Przeworski, we cannot always pinpoint one single event that would lead to erosion of democracy (Przeworski 2019, p. 26). AI—or any digital technology for that matter—alone cannot provoke such change. Still, the current ways in which we develop, deploy and (perhaps under)regulate pervasive digital systems can contribute to the decline of democratic regimes

and thus accelerate the slide of democracies down the slope. Using Rawls's basic structure as a concept foundational for democratic societies, we can see that the changes in the basic structure and our inability as a society to adequately respond to such changes can contribute to democratic backsliding through erosion of social justice. But what democracies are sliding towards still seems to be a mystery: is it a corporatocracy, i.e., the autocracy by private corporations? Is it an authoritarian regime? A guardianship of not philosophers, as Plato hoped for, but technologists? A new form of governance where corporations eventually enjoy democratic legitimacy and carry their part as guardians of just background conditions? Or something yet to be imagined?

Whereas the impacts of pervasive digital systems can be felt in democracies across the globe, it is not to say that all democracies would be sliding in the same direction. As shown by, e.g., Wunsch and Blanchard (2023), trajectories of democratic backsliding vary depending on variables such as the nature of existing electoral processes, citizen's capacity to opposition and constraints against executive power in place. It is thus likely that the trajectories of contemporary democracies are just as unique as their institutional and social contexts.

To avoid the trap of unnecessarily sombre, pessimistic future scenarios, it is worth imagining the opportunities the current situation offers. Let us imagine that providers of pervasive digital systems, such as Big Tech companies, became to adopt a role as basic structure institutions and made efforts to align their activities with the principles of justice. Doing so would improve the situation of the least advantaged members of society, strengthen equality of opportunity and the basic liberties. This could have direct and relatively quickly realised positive impacts on social justice and thus the foundations of just democracies. To stretch one's imagination even further, one could argue that the decision to enforce justice on corporations relies just as much on people's willingness to upkeep such social order as does the decision of some to strengthen the concentration of wealth at the expense of social justice.

In the spirit of Rawls's theory (and despite the fact that it has received critique on its idealistic nature, e.g., Harsanyi 1975; Nozick 2013; Sen 2010), we should not entirely close off the possibility of reaching a just society, despite the challenges apparent in today's societies and the direction they point. I thus next envision a way forwards for organisations to strive towards technology development informed by Rawlsian theory of justice.

#### 4.2 A way forwards: toward Rawlsian governance of pervasive digital systems

Using Rawls's theory to envision fair and democratically sustainable governance of pervasive digital systems offers an

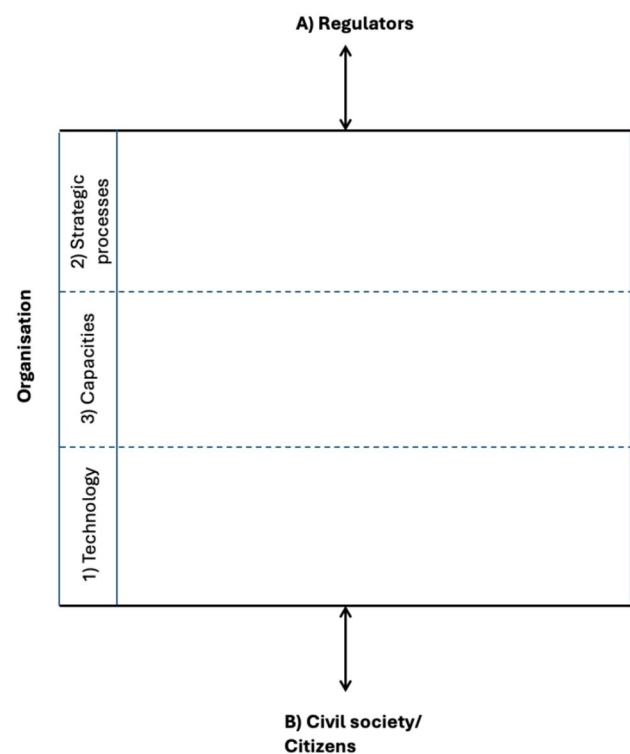
approach that enjoys robust moral- and socio-philosophical justifications, which is a welcome addition to the scattered landscape of digital ethics that has recently suffered from a lack of robust philosophical justification (Franzke 2022; Westerstrand 2024). In addition, it contributes to AI governance literature by giving well-reasoned justifications for why certain governance measures—whether organizational, governmental, or supranational—would be morally justified. This remedies the lack of attention to the societal and political context of AI governance and the lack of social awareness in existing ethics principles recognised in existing literature (Birkstedt et al. 2023).

As one of the main pain points in the current trajectory identified above is the lack of accountability and moral agency by private organisations developing and deploying pervasive digital systems, I focus on organisational governance and build on existing AI governance literature. This choice is motivated by the relatively established body of literature that considers the characteristics of pervasive digital systems in a more comprehensive manner compared to, e.g., IT or data governance. Although the present argument addresses systems beyond AI, doing so enables us to address gaps in related literature and approach governance mechanisms from a perspective familiar to organisations.

I build on the dimensions identified by Birkstedt et al. (2023) in their literature review on themes and knowledge gaps in organisational governance of AI systems, as well as the metarequirements identified by Mäntymäki et al. (2023) for AI governance. Accordingly, organisational AI governance revolves around four themes: technology (data, algorithms, AI systems), stakeholders and context (internal stakeholders, external stakeholders and sociopolitical context), regulation (hard and soft regulation), and processes (oversight, auditing, impact assessment) (Birkstedt et al. 2023, p. 145).

From this starting point, I propose an approach that distinguishes three internal tracks for organisational governance and two main external stakeholders that organisations should continuously keep in the loop in their governance. The three internal tracks are: 1) technology, 2) strategic processes, and 3) capacities. The two main external stakeholders are A) regulators, and B) civil society/citizens. This structure is illustrated in Fig. 1, and the contents of each component are discussed in the following sub-sections.

Compared to some existing governance frameworks (e.g., Mäntymäki et al. 2022), this framework does not consider principles of justice as external guidance. Rather, justice is seen as an inherent goal of the organisation itself that is attached to its role as a basic structural institution. As noted above, responsibility over ethical and social impacts should be seen as a driver rather than a barrier of innovation and competitiveness (see, e.g., Shah et al. 2024; Bednar and Spiekermann 2024), which from a Rawlsian perspective



**Fig. 1** Basic building blocks of Rawlsian governance of pervasive digital systems in basic structure organisations

requires using the principles of justice in all layers of organisational governance – from strategic processes to technical operations. This mitigates issues pointed out by, e.g., Bleher and Braun (2023), according to whom only relying on principles that are poorly justified can lead to operationalisations that are empty, or practices that are mere end in themselves—leading to ineffectiveness in securing just digital democracies but also benefits to the business itself.

Next, I discuss the tracks on the organisational level, how they address challenges revealed by the analysis in Sect. 3 and 4.1, and how they interact with regulators and civil society, forming a framework for Rawlsian governance of pervasive digital systems.

#### 4.2.1 Technology

On the technology track, Rawlsian principles need to be operationalised into practices that eventually guarantee that processes such as development and monitoring lead to systems aligned with the principles of justice. In this effort, data and algorithmic bias mitigation and explainability techniques take centre stage, as biases and opacity of algorithms have been identified as key challenges for social justice in existing systems (Sect. 3.1). Also, they play a key role in securing equality of opportunity and some of the basic liberties (e.g., Westerstrand 2024, pp. 10–11), as well as key

measures for algorithmic fairness (e.g., Leben 2017; 2018). Successful governance requires taking into consideration the characteristics of pervasive digital systems (Grover and Lyytinen 2023; Mäntymäki et al. 2023), which is why the choice of technical governance measures should be based on a reflection of those characteristics against Rawlsian principles of justice. For example, Rawlsian governance should address biases in algorithms that are used in decision-making with profound impact on people’s lives, such as employment (Brioscú et al. 2024, p. 40), sentencing (Simons 2023, pp. 38–50) or recruitment (O’Neil 2016) (for more discussion, see Sect. 3.1). As these challenges persist in existing systems, governance efforts should start from eliminating infringements of basic liberties or equality of opportunity, securing thus a minimum level of protection for the first two of Rawls’s principles in order of priority. I call this *passive protection* of social justice.

However, Rawls’s theory does not contend only with the prevention of harm. Instead, it argues that the basic structure is responsible over *improving* the status quo through the difference principle, i.e., ensuring that all necessary inequalities are to the greatest benefit of the least advantaged members of society. Therefore, I propose a gradual approach for operationalisation, which advances from passive protection towards *active protection* of social justice, which means implementing features that proactively advance basic liberties and equality of opportunity, as well as enable the greatest benefit to the least advantaged members of society. To give an example of what an active protection could mean in the development of pervasive digital systems, Simons (2023) introduces *positive equality duties*, which would require institutions to ‘demonstrate in designing decision-making systems that they took reasonable measures to explore how best to advance equality among protected and non-protected groups’ (p. 95). For machine learning systems, this would mean including consideration of how the design and development choices can rectify existing injustices and improve the situation of the least advantaged when, e.g., choosing bias mitigation, embeddings and decoding techniques, or the attributes that the algorithms should explicitly consider in their outputs (Simons 2023, 99–101).

Both passive and active protection can be implemented into an operating framework (e.g., MLOps/LLMOps) that integrates the measures into the whole lifecycle of the pervasive digital system, which has been identified as one of

the key requirements by, e.g., Mäntymäki et al. (2023) in organisational AI governance. The technical track with these elements is illustrated in Fig. 2.

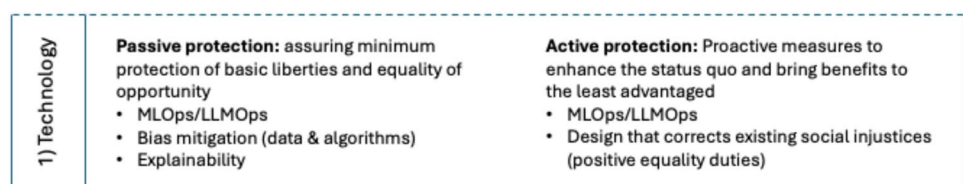
Nevertheless, technical measures come with limitations. For example, Leben has attempted to integrate Rawlsian considerations into algorithms (Leben 2017; 2018), but so far with contested results (e.g., Keeling 2018). As Sahlgren (2024) notes, fully mitigating biases in algorithms would require us to tackle structural biases in societies that the data represents, as technical mitigations always rely on rigid patterns that do not recognise the complexity of fairness in real life. Moreover, as effective consideration of ethical questions such as fairness is an active process of deliberation rather than a fully technical exercise (e.g., Ressayguier and Rodrigues 2020; Bleher and Braun 2023), only relying on technical measures is hardly enough to ensure that pervasive digital systems contribute to societal sustainability. Therefore, the limits of the technical measures need to be recognised and complemented with other governance measures: strategic processes and capacity building.

#### 4.2.2 Strategic processes

Strategic processes play a key role in ensuring that the governance mechanisms are not a hindrance but an integral part of organisation’s strategic roadmap (e.g., Birkstedt et al. 2023), bringing together social considerations and the capitalist logic of private organisations. To encourage strategic commitment to building just digital democracies, I argue that Rawlsian governance requires commitment from the leadership to advancing social justice. This commitment is essential in encouraging active protection of social justice also in situations where regulation fails to provide and enforce sufficient normative frameworks.

In addition, strategic processes are needed to implement legal requirements. In democracies, implementing legal requirements can serve as a baseline for Rawlsian governance of pervasive digital systems, as they often rely on principles similar to basic liberties and equality of opportunity (see, e.g., Westerstrand 2025). Such measures include, for example, adopting regular auditing and impact assessment processes called for in AI governance literature (Birkstedt et al. 2023), which in the case of Rawlsian governance should highlight the extent to which organisational processes and digital systems advance social justice. This can

**Fig. 2** Technical track of organisational governance of pervasive digital systems



help reveal responsibility gaps, opening a path for dialogue amongst basic structure institutions on the best ways to close such gaps. For those operating in the European market, EU AI Act gives partial coverage for Rawls's basic liberties that are of highest priority (Westerstrand 2024; Westerstrand 2025), which makes prioritising compliance a justified starting point for strategic efforts.

To work towards both compliance and strategic commitment, an oversight unit should be established to ensure implementation into practice (Birkstedt et al. 2023). This facilitates clarity in the division of responsibilities and labour, as well as support in conducting the measures needed for effective governance, such as risk management.

Finally, organisations should revisit their business model to reflect on its potential conflicts with advancing social justice. At first, aligning business with social justice might seem like a challenging if not unrealistic requirement for businesses operating in capitalist economy. This discontent, however, arises only if we assume that social justice and capitalism cannot coexist. Although Rawls was sceptical towards capitalism, research indicates that fairness and profit generation can be mutually supporting rather than conflicting goals. Ethically and societally sustainable solutions can be more competitive than their less sustainable competitors through, e.g., stronger innovation (e.g., Bednar and Spiekermann 2024) and increased user satisfaction and trust (Shah et al. 2024) (for more discussion, see Sect. 3.3). In addition, less sustainable business models that seems more profitable in short-term take us further away from reaching fair democracies. When such business models contribute to the erosion of democracies rather than societal stability, it is detrimental for value-creation—conflicting also with capitalist interest (see, e.g., Wang et al. 2024). Hence, adopting strategic processes that support societal sustainability contribute to both shareholder value creation expected of organisations belonging to the capitalist system, as well as advancing social justice expected of basic structure institutions, making them a justified part of effective governance of pervasive digital systems for public and private organisations alike.

As Rawls notes in the very beginning of *A Theory of Justice*, '[a] theory, however, elegant and economical must be rejected or revised if it is untrue; likewise laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust' (Rawls 1999, p. 3). Hence, from a Rawlsian perspective, a highly profitable

business model should be abandoned and replaced with another if it is in conflict with social justice, as such a method of accumulation of wealth is not morally justified. In times when wealth is increasingly concentrated, such challenges are arguably welcome, if not inevitable, if we want to sustain democracies in the age of pervasive digital systems.

Yet, for organisations that operate in the current system, challenging the entire market economy is an unrealistic request. Therefore, the present argument does not call for abandoning the capitalist logic of the market economy where businesses operate, as that would much broader societal discussion that is beyond the scope of organisational governance. Rather, it calls for organisations to opt for business models (i.e., models for profit generation) that improve the situation of the least advantaged members of society, which is a realistic endeavour. In this perspective, any other model should be deemed unfair and thus undesirable. In practice, consideration over the business model can be included in the innovation pipeline, periodic reviews and reporting to stakeholders, as well as regular strategy rounds or similar forums that guide the operations of the organisation, defining organisation's strategic direction and vision.

The strategic process track is illustrated in Fig. 3.

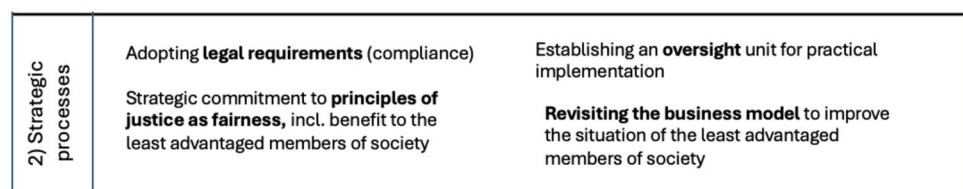
Technical measures and strategic processes discussed herein are likely to be unfamiliar to some organisations, in particular those that are new to the duties of basic structure. Therefore, capacity building measures are needed to equip organisations with the necessary skills and competences that enable the fulfilment of the principles of justice when developing and deploying pervasive digital systems, which I next discuss in more detail.

#### 4.2.3 Capacities

As demonstrated by AI ethics literature, establishing principles alone cannot ensure the sustainable development of pervasive digital systems. Rather, organisations need to exercise continuous deliberation and justify trade-offs and other choices of moral relevance (e.g., Bleher and Braun 2023; Rességuier and Rodrigues 2020). Capacity building is a way to train experts and leaders in social considerations on how their actions either advance or hinder the sustainability of democratic societies.

Whereas some of the responsibility over capacity development should already be part of engineering education

**Fig. 3** Strategic processes in organisational governance of pervasive digital systems



(e.g., Tuovinen and Rohunen 2021), I argue that basic structure institutions should ensure that their own personnel has the capacity to reflect on the impact of their actions on societal structures. In this regard, as well, organisations operating in the EU can start from compliance with the EU AI Act’s AI literacy requirement in Article 4 (and the guidance that is expected to follow on how to implement the requirement), which requires context-specific capacity building people that work with AI systems. It calls for ‘a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems’ in a way that considers contextual elements such as the skill level and the context of use of the AI system (EU AI Act, Article 4).

However, the EU AI Act has three main limitations that require additional consideration. First, it provides only minimum requirements for the protection of health, safety and fundamental rights, which means it is likely to lack in measures that would support active protection of justice as fairness (see, e.g., Westerstrand 2025). Second, it is only binding for organisations operating in EU countries, which limits its global applicability to non-European democracies—something that is a significant drawback in times when pervasive digital systems span across borders. Third, it only applies to AI systems. As discussed by, e.g., Stahl (2022), focusing on specific technologies when considering ethical and social impacts misses the overlapping and interconnected impacts digital technologies have in connected digital ecosystems. Therefore, organisations should establish programmes that go beyond the existing regulatory requirements to train their employees in recognising the complexities of pervasive digital systems and identifying their impacts on social justice, such as those discussed in Sects. 3 and 4.1 of this paper.

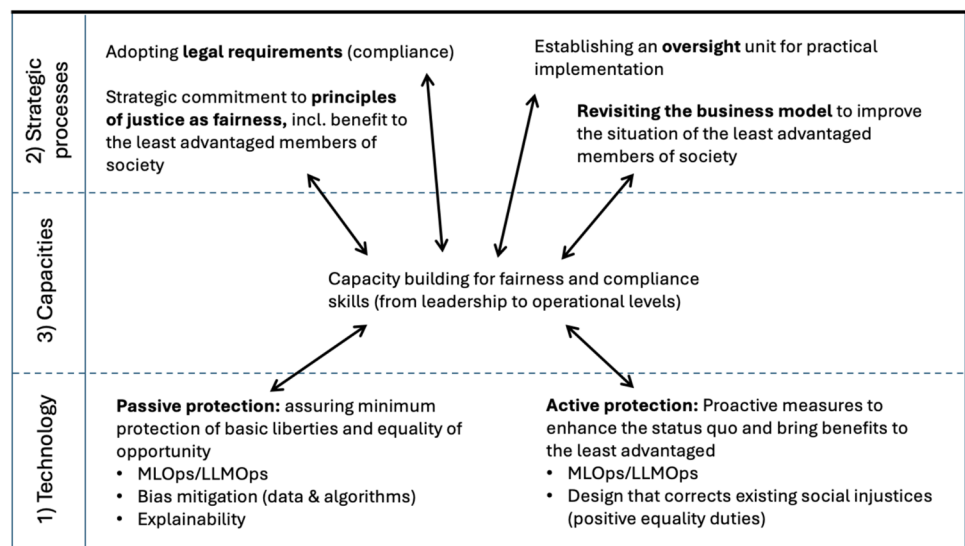
Besides training programmes, capacity building should be connected to both the strategic processes and technical operations to support continuous learning and reflection. Doing so, capacity building builds a bridge between strategic processes and technical operations, which allows organisations to develop their strategic processes based on feedback from technical operations, and vice versa. This addresses the rapid pace of introduction of pervasive digital systems and allows for continuous learning as new systems with potential new impacts enter the market. Figure 4 illustrates the three tracks together and their connectedness.

With the three tracks in place, it is time to steer attention from internal processes to exchange between the organisation and its key external stakeholders.

#### 4.2.4 Stakeholder engagement: citizen participation and regulatory sandbox as forum for public reason

For Rawls, the basic structure is a ‘complex of institutions’ (Rawls 2005, p. 260)—an entity where institutions are interdependent rather than isolated entities. To secure social justice, institutions belonging to that complex should collaborate and interact. What is more, in Rawls’s ideal society, citizens play an important role as deliberators over public good. For Rawls, this requires *public reason*, i.e., deliberation over what the conception of justice requires of the basic structure of society, as well as the purpose and end they serve (Rawls 2005, p. 213). Public reason is not only a matter of law (Rawls 2005, p. 2013) but a practice that is based on sincere and transparent public deliberation. This deliberation takes place in a fair system of social cooperation (e.g., Rawls 1999, p. 4). As Genelet and Mizrahi (2024) discuss, attributing agency to citizens through citizen participation should be a central part of AI governance initiatives, as it

**Fig. 4** Technology, strategic processes and capacity building in Rawlsian organisational governance of pervasive digital systems



mitigates many of the limitations of technical bias mitigation measures. I argue that participatory mechanisms should extend to organisational governance of pervasive digital systems so that both public and private institutions engage in public dialogue with regulators and citizens to gain input on governance mechanisms that contribute to societally sustainable digitalisation.

As private organisations operate in a different context compared to democratic institutions, the methods of public reason should be fit to their operating context. I thus propose establishing *regulatory sandboxes* that bring together regulators, providers and developers of pervasive digital systems, and civil society to deliberate over steering goals and mechanisms of digitalisation. Regulatory sandboxes have attracted increasing interest in the context of European technology regulation (e.g., the EU AI Act, Article 57), and in regulated sectors such as FinTech (e.g., McCarthy 2023). Yet, the existing proposals for regulatory sandboxes are typically portrayed as early platforms for testing and auditing for technology providers and regulators as main stakeholders (e.g., Díaz-Rodríguez et al., 2023), which excludes the potential of sandboxes to serve as a forum for ethical deliberation (Undheim et al. 2022). Considering the profound impact of pervasive digital systems on people's lives (Sect. 3) and the central role of public deliberation in Rawlsian idea of just democracy, granting citizens a place in this dialogue would strengthen the democratic legitimacy of these actions and the role of citizens as the ones holding power over the course of digitalisation. Regulatory sandbox can thus serve as an effective forum for public deliberation over the goals of technology regulation and measures of operationalisation that can lead to digital systems that serve the public good.

The ongoing dialogue with external stakeholders addresses the issue raised by Gabriel (2022), according to whom AI systems challenge Rawls's requirement of publicity that calls for transparency on the functioning of the basic structure institutions, without a need to disclose business secrets that would endanger the existence of these institutions in competitive markets. Developing a concept for such a sandbox would merit its own research and thus falls out of the scope of this paper, which is why I contend here to establish that such forum is needed.

In addition to regulatory sandboxes, I propose that organisation should make use of participatory methods when developing or adopting pervasive digital systems so that their potential to positively impact social justice can be demonstrated. This proposition, however, requires going beyond existing initiatives to 'democratise' pervasive digital systems. For example, Anthropic (2023) organised a citizen consultation to establish a 'constitution' for their AI development. However, Anthropic still holds the power over deciding whether citizens' voice will be heard, who will be involved in participation, and whether citizens will ever be

heard again in the next round of policy changes and model updates. Therefore, from a Rawlsian perspective, organisations should establish participatory mechanisms that are continuous and rely on research and expertise informed by democratic theory to ensure their democratic legitimacy.

For organisations that offer products and services, gaining user insight is typically already part of their operations. However, the impacts of pervasive digital systems are often not limited only to the system users. In the example of the Dutch algorithm that was used to detect child care benefit fraud (Heikkilä, 2022), the users were government officials, but the negative impacts caused by the biased system were families, children and their relatives who had never interacted with the system themselves. Digital systems thus pervade our lives, often regardless of the services we decide to buy, leaving us outside the scope of any end-user consultations the providers might conduct. Extending the dialogue beyond user groups would thus help mitigate the issue discussed by Genelet and Mizrahi (2024), namely, the inability of current bias mitigation and algorithmic auditing techniques to properly address the lived experience of citizens, targeting indirect effects rather than actual harms. Identifying the best methods for this purpose is a topic for further research.

Figure 5 illustrates the resulting framework, where:

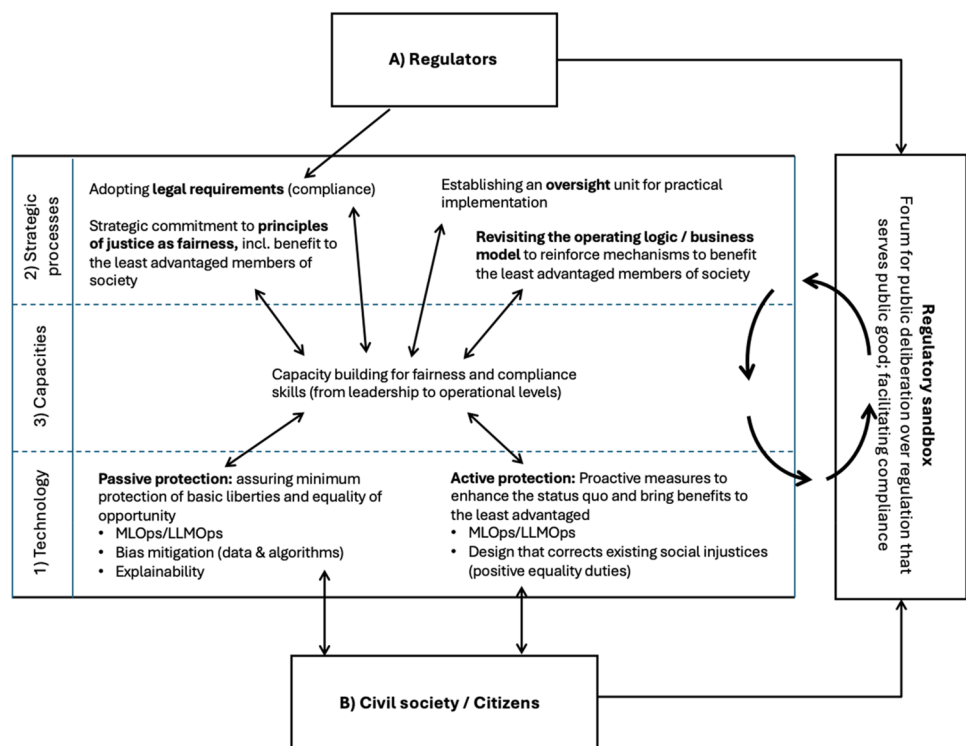
- A) The role of regulators is to provide legal requirements that are adopted in strategic processes and participate in the regulatory sandbox;
- B) Civil society and citizens are in continuous exchange with technological operations through ongoing dialogue and the regulatory sandbox.

## 5 Discussion

I have now used Rawls's concept of the basic structure of society to evaluate the impact of pervasive digital systems on democratic societies. To respond to the challenges revealed by this analysis, I have proposed a framework for organisations to govern pervasive digital systems towards a more societally sustainable direction in a democratic context. As pervasive digital systems pose fundamental challenges to democratic societies, also the changes expected from organisations—especially those unaccustomed to considering their impact on social justice—can seem difficult if not unrealistic to meet. This study has arguably pointed out a gap that requires a change of mindset, especially within private organisations that develop pervasive digital systems in a capitalist system.

Even so, the present approach does not call for a sudden change, or a *coup d'état* that would revolutionise the underlying logic of digitalisation overnight. The impacts

**Fig. 5** A framework for Rawlsian governance of pervasive digital systems in basic structure organisations



discussed in Sects. 3 and 4 have developed gradually over time, with many influencing factors contributing to how pervasive digital technologies shape democratic societies. The framework illustrated in Fig. 5 is also not intended as an operational governance model that covers all operations that governing digital systems requires. Rather, it is a framework that organisations can use to explore ways in which they can contribute to societal sustainability of pervasive digital systems and thus support democratic resilience in times of digital disruption. Adopting measures proposed in Sect. 4.2 can help mitigate challenges pointed out in Sect. 4.1., such as fairness of public and private services that influence people's lives and freedom of opinion-formation thanks to less manipulative system design. Perhaps most importantly, it can steer resources to developing technologies that can actively improve the current state of democracies and their fairness. It is realistic to expect the framework to be adopted piece by piece, starting with passive protection and building towards active protection of social justice, all while innovating for new business models with growing positive societal impact.

This approach also comes with limitations. I have only concentrated on the basic structure institutions and proposed actions for providers of pervasive digital systems, although these systems have implications for many other aspects of Rawlsian ideal democracy, as well, (e.g., public reason through AI-powered deep-fakes, Jones 2023). The basic structure of society consists of a variety of institutions, such as regulatory bodies, policymakers, and in Rawlsian view,

even family and religion. In further research, the framework could be extended beyond organisational governance to address other dimensions of social justice.

The implications of pervasive digital systems on democracies also merit discussion through theories other than only Rawls's—his theory has, after all, received critique. As one example among many, Sen (2010) criticises Rawls for being too idealistic and only presenting an unattainable ideal rather than a framework for comparing existing situations. Whereas I argue that the present study shows the usefulness of an ideal theory and its concepts in analysing real-life events, many of the practical implications discussed in Sect. 4.2 arguably require further development and validation to be fully applicable in practice. Applying Rawls's theory to establish the foundations does, however, offer those further studies a philosophically robust basis to build on. This mitigates the challenges pointed out by, e.g., Bleher and Braun (2023), according to whom operationalisations that are built on weak moral-philosophical foundations risk rather preventing than advancing ethical technology development. Hence, although this approach has its limits, it is an essential step towards robust governance mechanisms for societally sustainable digitalisation.

Lastly, it is essential to recognise that this approach is limited in geographical reach. As, e.g., Sen (2010) notes, Rawls's theory is difficult to apply in a global context, which has been demonstrated in applications of his principles into AI ethics (Westerstrand 2024). This framework thus does not resolve questions around global governance,

which would require coordination of actions beyond national borders (e.g., Erman and Furendal 2024). Although Rawls later attempted to extend his theory to address global justice (Rawls 2001), he still limited justice to be a matter of collaboration between nation-states, and more specifically, that of democratic nation-states. As contemporary pervasive digital systems exceed national borders and span across regimes, global governance of these systems should not assume democracy as the only possible premise for sustainable digitalisation. This said, we need to acknowledge the variety of layers that go into governance of pervasive digital systems, where I argue this paper still brings a contribution. Organisations do not need to wait for global governance mechanisms to start steering their actions towards a societally sustainable direction. Rawls's theory is drafted for democratic societies (Rawls 1971, p. viii), and thus it serves as a viable starting point for organisations operating in democratic regimes to align pervasive digital systems with democratic values, hence contributing to the democratic sustainability of these systems. Nevertheless, further research is needed to discover how we can globally coordinate the efforts in governance of pervasive systems in a way that is societally sustainable.

With these limitations in mind, this study comes with both theoretical and practical contributions. Firstly, it advances our understanding of the mechanisms through which pervasive digital systems impact democratic societies. It contributes to the discussion of social justice in the context of digitalisation, opening paths for further research that grows our ability to develop societally sustainable technologies. It contributes to the discussion on the role of private organisations in governing the social impacts of technology development (e.g., Gabriel 2022; Sullivan and Fosso Wamba 2022; Susskind 2022; Coeckelbergh 2024; Westerstrand 2024), bringing in a new perspective that highlights the role of social justice as a cornerstone for democratic resilience in the digital age. Secondly, this study offers practical guidance for industry actors that develop and deploy pervasive digital systems on how to do so in a more societally sustainable way, which complements the existing body of knowledge in governance of systems such as AI (Birkstedt et al. 2023; Mäntymäki et al. 2023). I hope these theoretical and practical insights, as well as the revealed questions for further research, encourage both researchers and practitioners to explore ways to steer digitalisation to a societally sustainable direction.

## 6 Conclusions

In this paper, I have used Rawls's basic structure of society as a conceptual tool to evaluate the impacts of pervasive digital systems (Grover and Lyytinen 2023) on

democracy. The analysis indicates that as a result of the increasingly profound impact of pervasive digital systems on people's lives, the shift of power to providers of these technologies, and limitations of regulations in addressing the challenges, there is a need to attribute responsibility over securing social justice also to private entities, including providers of pervasive digital systems. Currently, however, these organisations have not adopted the duty attributed to basic structure institutions over securing social justice. This has come with ethical implications, such as erosion of basic liberties like freedom and human autonomy (Coeckelbergh 2022; Prunkl 2024; Sio 2024), political liberties (Jones 2023; Łabuz and Nehring 2024), equality of opportunity and the difference principle (Westerstrand 2024).

This gap in responsibility and resulting lack of action that would safeguard social justice risks aggravating democratic decline. To work towards fair digital democracy, all institutions belonging to the basic structure—providers of pervasive digital systems in forefront—would need to align their actions with the principles of justice and thus safeguard fair background conditions for individuals and associations to function in society. Doing so ensures that pervasive digital systems protect basic liberties, support equality of opportunity, and improve the situation of the least advantaged members of society, leading to societally sustainable digitalisation that serves public good rather than utility maximisation for the few.

To help organisations address the identified gaps, I introduce a framework that builds on prior literature in AI ethics and organisational AI governance. This framework (Fig. 5 in Sect. 4.2.4) calls for organisational measures in compliance and oversight, strategic commitment to Rawlsian principles, as well as capacity building and technical measures to embed strategic principles and compliance requirements into actions towards social justice. These actions should be conducted with active engagement of citizens, civil society and regulators in regulatory sandboxes to ensure democratically legitimate interpretations of Rawlsian justice as fairness in the development and deployment of pervasive digital systems.

**Author contributions** S.W. is the sole contributor to this research.

**Funding** Open Access funding provided by University of Turku (including Turku University Central Hospital). No funding was received for preparing the manuscript or conducting the research.

**Data availability** No datasets were generated or analysed during the current study.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Almeida D, Shmarko K, Lomas E (2022) The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of US, EU, and UK regulatory frameworks. *AI Ethics* 2(3):377–387. <https://doi.org/10.1007/s43681-021-00077-w>
- Alnemr N (2020) Emancipation cannot be programmed: blind spots of algorithmic facilitation in online deliberation. *Contemp Polit* 26(5):531–552. <https://doi.org/10.1080/13569775.2020.1791306>
- Anthropic (2023). Collective constitutional AI: aligning a language model with public input. <https://www.anthropic.com/news/collective-constitutional-ai-aligning-a-language-model-with-public-input>. Last accessed 23 July, 2025.
- Beckman L, Hultin Rosenberg J, Jebari K (2024) Artificial intelligence and democratic legitimacy. The problem of publicity in public authority. *AI Soc* 39(3):975–984. <https://doi.org/10.1007/s00146-022-01493-0>
- Bednar K, Spiekermann S (2024) The power of ethics: uncovering technology risks and positive value potentials in IT innovation planning. *Bus Inf Syst Eng* 66:181–220. <https://doi.org/10.1007/s12599-023-00837-4>
- Berkey B (2021) Rawlsian institutionalism and business ethics: does it matter whether corporations are part of the basic structure of society? *Bus Ethics Q* 31(2):179–209. <https://doi.org/10.1017/beq.2020.14>
- Bermeo N (2016) On democratic backsliding. *J Democ* 27(1):5–19
- Bernholz L, Landemore H, Reich R (2021) Digital technology and democratic theory. University of Chicago Press, Chicago
- Birkstedt T, Minkinen M, Tandon A, Mäntymäki M (2023) AI governance: themes, knowledge gaps and future agendas. *Internet Res* 33(7):133–167. <https://doi.org/10.1108/INTR-01-2022-0042>
- Blanc S, Al-Amoudi I (2013) Corporate institutions in a weakened welfare state: a Rawlsian perspective. *Bus Ethics Q* 23(4):497–525. <https://doi.org/10.5840/beq201323438>
- Bleher H, Braun M (2023) Reflections on putting AI ethics into practice: how three AI ethics approaches conceptualize theory and practice. *Sci Eng Ethics* 29(3):21. <https://doi.org/10.1007/s11948-023-00443-3>
- Bode L (2016) Political news in the news feed: learning politics from social media. *Mass Commun Soc* 19(1):24–48. <https://doi.org/10.1080/15205436.2015.1045149>
- Bonechi S (2024) Development of an automated moderator for deliberative events. *Electronics* 13(3):Article 3. <https://doi.org/10.3390/electronics13030544>
- Brioscú A, Lauringson A, Saint-Martin A, Xenogiani T (2024) A new dawn for public employment services: Service delivery in the age of artificial intelligence. OECD. <https://doi.org/10.1787/5dc3eb8e-en>
- Buchanan A (2002) Political legitimacy and democracy. *Ethics* 112(4):689–719. <https://doi.org/10.1086/340313>
- Chandler D (2023). *Free and equal: what would a fair society look like?* Penguin, UK.
- Chomanski B (2021) The missing ingredient in the case for regulating big tech. *Minds Mach* 31(2):257–275. <https://doi.org/10.1007/s11023-021-09562-x>
- Christiansen PM (2024) Corporatism. *Political science and public policy*. Elgar, Cheltenham, pp 26–37
- Coeckelbergh M (2022) *The political philosophy of AI: an introduction*. Wiley, Amsterdam
- Coeckelbergh M (2024) *Why AI undermines democracy and what to do about it*. Wiley, Amsterdam
- Cohen J (2003) For a democratic society. In: Freeman S (ed) *The Cambridge companion to Rawls*. Cambridge University Press, Cambridge, pp 86–167
- Cohen MA (2010) The narrow application of Rawls in business ethics: a political conception of both stakeholder theory and the morality of markets. *J Bus Ethics* 97(4):563–579. <https://doi.org/10.1007/s10551-010-0525-y>
- Corporate Europe Observatory (2023a). Lobbying power of Amazon, Google and Co. Continues to grow | Corporate Europe Observatory. <https://corporateeurope.org/en/2023/09/lobbying-power-amazon-google-and-co-continues-grow>. Last accessed 23 July, 2025.
- Corporate Europe Observatory (2023b). Byte by Byte: how big tech undermined the AI Act. <https://corporateeurope.org/en/2023/11/byte-byte>. Last accessed 23 July, 2025.
- Danaher J, Nyholm S (2024) Digital duplicates and the scarcity problem: might AI make us less scarce and therefore less valuable? *Philos Technol* 37(3):106. <https://doi.org/10.1007/s13347-024-00795-z>
- de Sio FS (2024) *Human freedom in the age of AI*. Routledge, Milton Park. <https://doi.org/10.4324/9781003303244>
- Díaz-Rodríguez N, Del Ser H, Coeckelbergh M, López de Prado M, Herrera-Viedma E, Herrera F (2023) Connecting the dots in trustworthy Artificial Intelligence: From AI principles ethics and key requirements to responsible AI systems and regulation. *Information Fusion* 99, 101896. <https://doi.org/10.1016/j.inffus.2023.101896>
- Douglas DM (2015) Towards a just and fair internet: applying Rawls' principles of justice to internet regulation. *Ethics Inf Technol* 17(1):57–64. <https://doi.org/10.1007/s10676-015-9361-1>
- Du H, Teng S, Chen H, Ma J, Wang X, Gou C, Li B, Ma S, Miao Q, Na X, Ye P, Zhang H, Luo G, Wang F-Y (2023) Chat with ChatGPT on intelligent vehicles: an IEEE TIV perspective. *IEEE Trans Intell Veh* 8(3):2020–2026. <https://doi.org/10.1109/TIV.2023.3253281>
- Erman E, Furendal M (2024) Artificial intelligence and the political legitimacy of global governance. *Polit Stud* 72(2):421–441. <https://doi.org/10.1177/00323217221126665>
- Fia M, Sacconi L (2019) Justice and corporate governance: new insights from Rawlsian social contract and Sen's capabilities approach. *J Bus Ethics* 160(4):937–960. <https://doi.org/10.1007/s10551-018-3939-6>
- Franke U (2021) Rawls's original position and algorithmic fairness. *Philos Technol* 34(4):1803–1817. <https://doi.org/10.1007/s13347-021-00488-x>
- Franzke AS (2022) An exploratory qualitative analysis of AI ethics guidelines. *J Inf Commun Ethics Soc* 20(4):401–423. <https://doi.org/10.1108/JICES-12-2020-0125>
- Gabriel I (2022) Toward a theory of justice for artificial intelligence. *Daedalus* 151(2):218–231
- Genelet K, Mizrahi SK (2024) A Human-centered approach to AI governance. Operationalizing human rights through citizen participation. In: Régis C, Denis J-L, Axente ML, Kishimoto A (eds) *Human-centered AI. A multidisciplinary perspective*

- for policy-makers, auditors and users. Chapman and Hall/CRC, Boca Raton, pp 215–230
- Gent E (2023) A cryptocurrency for the masses or a universal ID?: worldcoin aims to scan all the world’s eyeballs. *IEEE Spectr* 60(1):42–57. <https://doi.org/10.1109/MSPEC.2023.10006664>
- Grover V, Lyytinen K (2023) The pursuit of innovative theory in the digital age. *J Inf Technol* 38(1):45–59. <https://doi.org/10.1177/02683962221077112>
- De Guzman C (2024). What to know about Meta’s ‘Political Content’ Limit—and how to turn it off on Instagram. *Time Magazine*. <https://time.com/6960587/meta-instagram-political-content-limit-off-setting-default/>. Last accessed 23 July, 2025.
- Hacker P (2018) Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Rev*. <https://doi.org/10.54648/COLA2018095>
- Harsanyi JC (1975) Can the maximin principle serve as a basis for morality? A critique of John Rawls’s theory. *Am Polit Sci Rev* 69(2):594–606. <https://doi.org/10.2307/1959090>
- Harwell D, Nix N, Zakrzewski C (2024). Liars and trolls overwhelm social media after Trump rally shooting. *Washington Post*. <https://www.washingtonpost.com/nation/2024/07/15/social-media-trump-rally-shooting-fake-photos-conspiracies-rnc/>. Last accessed 23 July, 2025.
- Heikkilä M (2022). Dutch scandal serves as a warning for Europe over risks of using algorithms. *Politico*. <https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/>. Last accessed 23 July, 2025.
- Javaid M, Haleem A, Singh RP (2023) ChatGPT for healthcare services: an emerging stage for an innovative perspective. *Bench-Council Transact Benchmarks Stand Eval* 3(1):100105. <https://doi.org/10.1016/j.tbench.2023.100105>
- Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. *Nat Mach Intell* 1(9):Article 9. <https://doi.org/10.1038/s42256-019-0088-2>
- Jones N (2023) How to stop AI deepfakes from sinking society—and science. *Nature* 621(7980):676–679. <https://doi.org/10.1038/d41586-023-02990-y>
- Jungherr A (2023) Artificial intelligence and democracy: a conceptual framework. *Soc Media Soc* 9(3):20563051231186353. <https://doi.org/10.1177/20563051231186353>
- Karimov A (2024) Comment on the article “Filter Bubbles and the Unfeeling: How AI for Social Media Can Foster Extremism and Polarization” (PHTE-D-23-00352R3). *Philos Technol* 37(3):93. <https://doi.org/10.1007/s13347-024-00768-2>
- Kaur R, Gabrijelčič D, Klobučar T (2023) Artificial intelligence for cybersecurity: literature review and future research directions. *Inf Fusion* 97:101804. <https://doi.org/10.1016/j.inffus.2023.101804>
- Keeling G (2018) Against Leben’s Rawlsian collision algorithm for autonomous vehicles. In: Müller VC (ed) *Philosophy and theory of artificial intelligence 2017*. Springer, New York, pp 259–272
- Kilovaty I (2019) Legally cognizable manipulation. *Berkeley Tech LJ* 34:449
- Klenk M (2024) Ethics of generative AI and manipulation: a design-oriented research agenda. *Ethics Inf Technol* 26(1):9. <https://doi.org/10.1007/s10676-024-09745-x>
- König PD, Wenzelburger G (2020) Opportunity for renewal or disruptive force? How artificial intelligence alters democratic politics. *Gov Inf Q* 37(3):101489. <https://doi.org/10.1016/j.giq.2020.101489>
- Kordzadeh N, Ghasemaghahi M (2022) Algorithmic bias: review, synthesis, and future research directions. *Eur J Inf Syst* 31(3):388–409. <https://doi.org/10.1080/0960085X.2021.1927212>
- Łabuz M, Nehring C (2024) On the way to deep fake democracy? Deep fakes in election campaigns in 2023. *Eur Polit Sci*. <https://doi.org/10.1057/s41304-024-00482-9>
- Larson J, Mattu S, Kirchner L, Angwin J (2016) How we analyzed the COMPAS recidivism algorithm. *ProPublica* 9(1):3
- Laux J, Wachter S, Mittelstadt B (2021) Taming the few: Platform regulation, independent audits, and the risks of capture created by the DMA and DSA. *Comput Law Secur Rev* 43:105613. <https://doi.org/10.1016/j.clsr.2021.105613>
- Leben D (2017) A Rawlsian algorithm for autonomous vehicles. *Ethics Inf Technol* 19(2):107–115. <https://doi.org/10.1007/s10676-017-9419-3>
- Leben D (2018) *Ethics for robots: how to design a moral algorithm*. Routledge. <https://doi.org/10.4324/9781315197128>
- Mainz JT, Sønderholm J, Uhrenfeldt R (2024) Artificial intelligence and the secret ballot. *AI Soc* 39(2):515–522. <https://doi.org/10.1007/s00146-022-01551-7>
- Manheim K, Kaplan L (2019) Artificial Intelligence: Risks to Privacy and Democracy. *Yale Journal of Law and Technology* 21, 106.
- Mäntymäki M, Minkinen M, Birkstedt T, Viljanen M (2022) Putting AI ethics into practice: the hourglass model of organizational AI governance. *arXiv*. <https://doi.org/10.48550/arXiv.2206.00335>
- Mäntymäki M, Minkinen M, Zimmer M, Birkstedt T, Viljanen M (2023). Designing an AI governance framework: from research-based premises to meta-requirements. *ECIS 2023 Research Papers*. [https://aisel.aisnet.org/ecis2023\\_rp/295](https://aisel.aisnet.org/ecis2023_rp/295). Last accessed 23 July, 2025.
- McCarthy J (2023) From childish things: the evolving sandbox approach in the EU’s regulation of financial technology. *Law Innov Technol* 15(1):1–24. <https://doi.org/10.1080/17579961.2023.2184131>
- Mitchell S, Potash E, Barocas S, D’Amour A, Lum K (2021) Algorithmic fairness: choices, assumptions, and definitions. *Annu Rev Stat Appl* 8(1):141–163. <https://doi.org/10.1146/annurev-statistics-042720-125902>
- Moses LB (2013) How to Think about Law, Regulation and Technology: Problems with ‘Technology’ as a Regulatory Target. *Law Innov Technol* 5(1):1–20. <https://doi.org/10.5235/17579961.5.1.1>
- Muldoon J, Raekstad P (2023) Algorithmic domination in the gig economy. *Eur J Polit Theory* 22(4):587–607. <https://doi.org/10.1177/14748851221082078>
- Muldoon J (2022) *Platform socialism: how to reclaim our digital future from big tech*. Pluto Press.
- Nemitz P (2018) Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences* 376(2133): 20180089. <https://doi.org/10.1098/rsta.2018.0089>
- Nieminen E (2024). Human rights -teksti t-paidassa oli Instagramille liikaa—Meta rajoittaa poliittista sisältöä, ja tulkinnat yllättävät. *Yle*. <https://yle.fi/a/74-20121337>. Last accessed 23 July, 2025.
- Nozick R (2013) *Anarchy, State, and Utopia*. Basic Books, New York.
- O’Neil C (2016) *Weapons of math destruction: how big data increases inequality and threatens democracy*. Penguin Books, Great Britain.
- Pettit P (2012) *On the people’s terms*. Cambridge University Press, Cambridge
- Prunkl C (2024) Human autonomy at risk? An analysis of the challenges from AI. *Minds Mach* 34(3):26. <https://doi.org/10.1007/s11023-024-09665-1>
- Przeworski A (2019) *Crises of democracy*. Cambridge University Press, Cambridge
- Rawls J (1971) *A theory of justice*. Harvard University Press, Harvard. <https://doi.org/10.2307/j.ctvjf9z6v>
- Rawls J (1999) *A Theory of Justice, revised*. Harvard University Press, Harvard
- Rawls J (2001) *The law of peoples: with “the idea of public reason revisited.”* Harvard University Press, Harvard

- Rawls J (2005) Political liberalism. Columbia University Press, Columbia
- Rességuier A, Rodrigues R (2020) AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data Soc* 7(2):205395172094254. <https://doi.org/10.1177/2053951720942541>
- Rodilosso E (2024) Filter bubbles and the unfeeling: how AI for social media can foster extremism and polarization. *Philos Technol* 37(2):71. <https://doi.org/10.1007/s13347-024-00758-4>
- Sahlgren O (2024) What's impossible about algorithmic fairness? *Philos Technol* 37:124. <https://doi.org/10.1007/s13347-024-00814-z>
- Scherer MU (2015) Regulating artificial intelligence systems: risks, challenges, competencies, and strategies. *Harvard J Law Technol* 29(2):353–400
- Schmitt M, Flechais I (2024) Digital deception: generative artificial intelligence in social engineering and phishing. *Artif Intell Rev* 57(12):324. <https://doi.org/10.1007/s10462-024-10973-2>
- Sen A (2010). The idea of justice. Penguin Books, London.
- Shah MU, Rehman U, Parmar B, Ismail I (2024) Effects of moral violation on algorithmic transparency: an empirical investigation. *J Bus Ethics* 193:19–34. <https://doi.org/10.1007/s10551-023-05472-3>
- Simons J (2023). Algorithms for the people: democracy in the Age of AI. Princeton University Press
- Singer A (2015) There is no Rawlsian theory of corporate governance. *Bus Ethics Q* 25(1):65–92. <https://doi.org/10.1017/beq.2015.1>
- Stahl BC (2022) From computer ethics and the ethics of AI towards an ethics of digital ecosystems. *AI Ethics* 2(1):65–77
- Stahl BC, Eke D (2024) The ethics of ChatGPT – exploring the ethical issues of an emerging technology. *Int J Inf Manage* 74:102700. <https://doi.org/10.1016/j.ijinfomgt.2023.102700>
- Stokel-Walker C (2025), How the Trump administration plans to use algorithms to target protesters. *Fast Company*. Available at: <https://www.fastcompany.com/91295390/how-the-trump-administration-plans-to-use-algorithms-to-target-protesters>. Last accessed Apr 10 2025.
- Sullivan Y, Fosso Wamba S (2022) Moral judgements in the age of artificial intelligence. *J Bus Ethics* 178:917–943. <https://doi.org/10.1007/s10551-022-05053-w>
- Susskind J (2022). The digital republic: on freedom and democracy in the 21st Century. Bloomsbury, Dublin.
- Tolan S (2019) Fair and unbiased algorithmic decision making: current state and future challenges. arXiv. <https://doi.org/10.48550/arXiv.1901.04730>
- Tuovinen L, Rohunen A (2021). Teaching AI ethics to engineering students: reflections on syllabus design and teaching methods. *Proceedings of the Conference on Technology Ethics 2021—Tethics 2021*. 19–33. [https://ceur-ws.org/Vol-3069/FP\\_02.pdf](https://ceur-ws.org/Vol-3069/FP_02.pdf). Last accessed 23 July, 2025.
- Undheim K, Erikson T, Timmermans B (2022) True uncertainty and ethical AI: regulatory sandboxes as a policy tool for moral imagination. *AI Ethics* 3:997–1002. <https://doi.org/10.1007/s43681-022-00240-x>
- Wang N, Pan H, Feng Y, Du S (2024) How do ESG practices create value for businesses? Research review and prospects. *Sustain Account Manag Policy J* 15(5):1155–1177. <https://doi.org/10.1108/SAMPJ-12-2021-0515>
- Westerstrand S (2024) Reconstructing AI ethics principles: Rawlsian ethics of artificial intelligence. *Sci Eng Ethics* 30(5):46. <https://doi.org/10.1007/s11948-024-00507-y>
- Westerstrand S (2025) Fairness in AI Systems Development: Beyond EU AI Act Compliance. In: Papatheocharous E, Farshidi S, Jansen S, Hyrynsalmi S (eds) *Software business*. ICSOB 2024. *Lecture Notes in Business Information Processing (LNBIIP)*. Springer, Cham
- Westerstrand S, Westerstrand R, Koskinen J (2024) Talking existential risk into being: a Habermasian critical discourse perspective to AI hype. *AI Ethics*. <https://doi.org/10.1007/s43681-024-00464-z>
- Wood D (1994) Business justice: transactions, resources, and organisations. *J Bus Ethics* 13(6):481–486
- Wunsch N, Blanchard P (2023) Patterns of democratic backsliding in third-wave democracies: a sequence analysis perspective. *Democratization* 30(2):278–301. <https://doi.org/10.1080/13510347.2022.2130260>
- York JC (2021). *Silicon values: the future of free speech under surveillance capitalism*. Verso Books, London/Brooklyn.
- Zuboff S (2015) Big other: surveillance capitalism and the prospects of an information civilization. *J Inf Technol* 30(1):75–89. <https://doi.org/10.1057/jit.2015.5>
- Zuboff S (2019). *The age of surveillance capitalism*. Profile Books, London.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.