CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Political Studies

Department of Political Science

Bachelor's Thesis

2024 **Johanna Reichart**

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Political Studies

Department of Political Science

Artificial Intelligence as a Challenge to Social Justice in the Light of the Theories of J. Rawls and M. Walzer

Bachelor's Thesis

Author of the Thesis: Johanna Reichart

Study programme: Politics, Philosophy and Economics

Supervisor: Janusz Salamon, Ph.D.

Year of the defence: 2024

Declaration

1. I hereby declare that I have compiled this thesis using the listed literature and resources only.
2. I hereby declare that my thesis has not been used to gain any other academic title.
3. I fully agree to my work being used for study and scientific purposes.

In Prague on 29 July 2024 Johanna Reichart

References

REICHART, Johanna. *Artificial Intelligence as a Threat to Social Justice in the Light of the Theories of J. Rawls and M. Walzer.* Praha, 2024. 61 s. Bachelor's thesis (Bc). Charles University, Faculty of Social Sciences, Institute of Political Studies, Department of Political Science. Supervisor Janusz Salamon, Ph.D.

Length of the Thesis: 104,742 characters

Abstract

The developments of Artificial Intelligence (AI) challenge the distribution of various social goods e.g. privacy, equality (of opportunity), interpersonal relationships, and the balance of power among different actors in unprecedented ways. Since the prevalence of some of their goods is among the very assumptions of various theories of justice, this thesis aims to reexamine the conceptual framework of J. Rawls' theory of *Justice as Fairness* and M. Walzer's *Justice as Complex Equality* to answer the question of whether they are conceptually equipped to be applied in the light of the developments of AI.

Using methods of analytical political philosophy this thesis identifies various social goods whose just distribution is challenged by the developments of AI. Building on those findings, the limitations of J. Rawls' and M. Walzer's theories of justice to address those challenges are uncovered by adhering to their own respective methodologies. The argument is that AI is challenging both theories in their very assumptions in a way that to protect one fundamental social good another one would have to be given up on, hence both theories' conceptual frameworks are insufficient to accommodate the challenges AI poses to them. Therefore, it concludes that in light of the developments of AI, new theories of justice must be developed.

Abstrakt

Vývoj umělé inteligence zpochybňuje distribuci různých sociálních statků, např. soukromí, rovnost (příležitostí), mezilidské vztahy a rovnováhu moci mezi různými aktéry, a to dosud nevídaným způsobem. Vzhledem k tomu, že převaha některých z jejich statků patří k samotným předpokladům různých teorií spravedlnosti, je cílem této práce znovu prozkoumat konceptuální rámec teorie J. Rawlse Spravedlnost jako spravedlnost a M. Walzera Spravedlnost jako komplexní rovnost a odpovědět na otázku, zda jsou konceptuálně vybaveny pro aplikaci ve světle vývoje UI.

S využitím metod analytické politické filosofie tato práce identifikuje různé společenské statky, jejichž spravedlivé rozdělení je vývojem UI zpochybněno. Na základě těchto zjištění jsou odhalena omezení teorií spravedlnosti J. Rawlse a M. Walzera při řešení těchto výzev, a to na základě dodržování jejich vlastních metodologií. Argumentuje se tím, že UI zpochybňuje obě teorie v jejich samotných předpokladech takovým způsobem, že k ochraně jednoho základního společenského dobra by se muselo rezignovat na jiné, tudíž

konceptuální rámce obou teorií jsou nedostatečné k tomu, aby se vyrovnaly s výzvami, které před ně UI staví. Proto dochází k závěru, že ve světle vývoje UI je třeba vytvořit nové teorie spravedlnosti.

Keywords

Artificial Intelligence, Social Justice, John Rawls, Michael Walzer

Klíčová slova

Artificial Intelligence, Social Justice, John Rawls, Michael Walzer

Title

Artificial Intelligence as a Challenge to Social Justice in the Light of the Theories of J. Rawls and M. Walzer

Název práce

Umělá inteligence jako výzva sociální spravedlnosti ve světle teorií J. Rawlse a M. Walzera

Acknowledgement

I would like to thank Dr Janusz Salamon for making this thesis possible.

Table of Contents

Introduction

Artificial Intelligence (AI) has been ascribed the potential to "reshap[e] the values and practices of government, business, and society" (De Oliveira et al., 2022, p. 2) and it has done so e.g. in public services which are distributed by an AI, in bureaucratic organisation or in transforming the role of citizens (Mergel et al., 2023, pp. 7-8). Not only does the implementation of AI subsequently threaten the perception of fairness (e.g., Al Samman & Mohamed, 2024; Yalcin et al., 2023) but can also create situations of injustice (Rafanelli, 2022, pp. 2-4).

To solve problems of justice in a systematic, thus non-arbitrary, way, various theories on how to best achieve social justice in society have been developed (Rafanelli, 2022, p. 2) and using them has been argued to "now [be] more necessary than ever" (Han, 2017, p. 50). However, as has been shown by e.g. John Rawls, received theories of justice might have to be amended to their contextual environment[1] while others e.g. Walzer (1983) have developed perfectionist theories that, by definition, do not need to be amended to a particular situation. Those theories are designed to leave enough room in their conceptual framework so that it would be possible to apply them to any given circumstance (Van Wyk, 2008, p. 258). However, in the light of AI, different authors have come to different conclusions about whether theories that take a similar approach to justice as Rawls' are better suited to address the challenges AI is posing to social justice (e.g., Ferretti, 2022) or whether it is Walzer's (e.g., Santoni De Sio et al., 2021).

Despite this, the developments of AI might be changing the very foundations of human society so that the basic assumptions of the currently received theories of justice might be challenged. It follows that for social justice to prevail, an examination of the conceptual framework of those theories is needed. Therefore, my thesis aims to answer the following question: *What are the conceptual limitations of J. Rawls' and M. Walzer's theories of justice in light of the developments of AI and will it be enough to amend those theories, or do they need to be replaced by other, new theories of justice?*

In my work, I shall adhere to a conceptualisation of AI as a digital tool that can develop human-like, autonomous behaviours from data input. From data, it is able to learn,

[1] Rawls showed this by constantly amending his theory of *Justice as Fairness* to the global political situation. For instance, what started as a domestic theory of justice was later amended by a book on its global application after the collapse of the Soviet Union (see *The Law of Peoples*).

detect recurring patterns, and act accordingly (Zuiderwijk et al., 2021, p. 2). Within AI technologies, a further distinction between strong AI and weak AI can be made whereby 'weak AI' is designed to detect recurring patterns while 'strong AI' is expected to perform more cognitively complex, and thus challenging tasks (Hermansyah et al., 2023, p. 158). In the following chapters, my argument is primarily centred on strong AI.

To be able to answer the given research question, I will first outline my methodology. The chapter that follows will be dedicated to identifying those social goods whose just distribution is most challenged in the light of AI developments starting with aspects connected to political life and continuing with aspects connected to economic life. In Chapter 2, I apply the findings of the previous chapter and shed light on how J. Rawls' and M. Walzer's theories of justice fail to protect those goods. The last section consists of a comparative analysis of both theories' frameworks and aims to answer the question of which of them is better conceptually equipped to be applied in an AI-driven context.

Methodology

Justification of Cases of Analysis

Choice of Theories

I first had to make an appropriate choice of theories whose conceptual framework should be analysed in light of the developments of AI. J. Rawls' theory of *justice as fairness* was chosen due to its incredible influence on the discourse on social justice (Johnston, 2011, p. 196) and because some scholars (e.g., Ferretti, 2022) have argued that his approach to social justice would be best suited to accommodate the challenges AI poses on the matter. Furthermore, the fact that he developed his work throughout various years leaves room to assume that his conceptual framework is not perfect.

M. Walzer on the other hand, takes a fundamentally different approach to social justice than Rawls by formulating a perfectionist theory which heavily stresses the importance of community and membership (Walzer, 1983, Chapter 2). Therefore, Walzer's theory of justice as *complex equality* seems to be quite the opposite of Rawls' *justice as fairness* in many of its conceptual provisions. Nevertheless, others have argued that Walzer's theory would be best equipped to be applied to AI-driven societies rather than Rawls' (e.g., Santoni De Sio et al., 2021).

It follows that comparing the chosen two theories to answer the second part of the research question, whether it is enough to make amendments to them or whether entirely new theories of justice must be developed, seems sensible.

Choice of Realms for the Identification of Vulnerable Social Goods

Both assessed theories are concerned with how to distribute social goods justly among members of a community. Therefore, in order to shed light on the conceptual limitations of those theories, those social goods whose just distribution is most vulnerable to the developments of AI must be identified first.

Because of the limited scope of this bachelor's thesis, I limited the analysis thereof to aspects related to political and economic life since those are commonly considered to lay the foundations for other realms of life. My analysis of the economic realm encompasses all sorts of market exchange, including the labour market, as well as the realm of education. This is because an individual's education determines their position in the labour market to a significant degree.

The political realm, on the other hand, is that in which legislation is made, thus the space where a state interacts with its citizens. Therefore, I included all sorts of political participation and power relations, although not exclusively between public and private actors, in my examination.

Methods of Analysis

Identification of Vulnerable Goods

To identify the social goods which are most vulnerable to the developments of AI, I employed methods from analytical political philosophy which means that I "rely upon intuitions when trying to determine what the rules of morality require" (McDermott, 2008, p. 15) which further entails that I have to make assumptions about some essential features of a good human (social) life (McDermott, 2008, p. 17). The very basic one of them, which my analysis is centred around, is that equality between actors and equality of opportunity for individuals is central to relations of justice. This conforms to both of the assessed theories' arguments (Rawls, 1985, p. 227; Walzer, 1983, p. 17).

Furthermore, I used concepts from *consent theory* which stresses the importance of individual autonomy and self-determination. It follows that for infringements on an

individual's liberties to be just, consent has to be given which can be done both in the form of *verbally expressed* consent and *tacit* consent (McDermott, 2008, p. 18).

Analysis of the Theories' Conceptual Frameworks

To examine the conceptual framework of J. Rawls' and M. Walzer's theories of justice in light of the developments of AI I adhered to both theories' respective own methods.

Rawls' method of *reflective equilibrium* consists of a revision of the agreements made in the *original position* with the agent's intuitive understanding of principles of justice after the *Veil of Ignorance* has been lifted. This process would be necessary since even in the most unfavourable conditions, moral "judgements are no doubt subject to certain irregularities and distortions" (Rawls, 1999a, p. 42) due to moral learning throughout the subject's life or other factors that may alter an individual's sense of justice in a different context. The reflective equilibrium is thus "reached after a person has weighed various proposed conceptions and he has either revised his judgements to accord with […] them or held fast of his initial convictions (and the corresponding conceptions)" (Rawls, 1999a, p. 43).

On the other hand, Walzer's method of *deep interpretation* (Stassen, 1994, p. 379) is one for which "[w]e don't have to discover the moral world because we have always lived there" (Walzer, 1985, p. 19). It follows that "[m]oral argument […] is interpretive [in] character, closely resembling the work of a lawyer or judge who struggles to find meaning in a morass of conflicting laws and precedents" (Walzer, 1985, p. 19). Thus, a repeated evaluation of social institutions based on moral intuitions and more abstract principles can be conducted to determine whether a specific institution is just in the given cultural and historical context (Stassen, 1994, p. 379). For Walzer's theory, the core of justice is that a privileged position in one social sphere does not translate into advantages in others (Walzer, 1983, p. 10).

Comparative Analysis of the Theories' Conceptual Limitations

The last Chapter is a comparative analysis of both theories' frameworks and their potential to accommodate the challenges posed to them by AI developments. Because this method cannot discover new theories but only examine existing ones (Jahn, 2007, p. 20) it is sufficient for the given research question. Nevertheless, because of the missing data of philosophical inquiry, it must be combined with another method (Jahn, 2007, p. 22).

Therefore, I have chosen to employ the methods of analytical political philosophy I have also used to identify the most vulnerable social goods.

1. Vulnerable Social Goods

Because both J. Rawls', as well as M. Walzer's theories of justice, are concerned with the distribution of various social goods among the members of a given community, it is crucial to identify those social goods which are most vulnerable to AI developments before delving into the conceptual limitations of the theories.

While Rawls is merely concerned with the distribution of certain *primary goods* that the proper design of the *basic structure* is aimed to achieve (Rawls, 1999a, pp. 54, 54 ff.), Walzer argues that justice itself "requires a positive structure" (Walzer, 1984, p. 322) that consists of "institutions, rules, mores, and customary practices"(Walzer, 1984, p. 322) so that all actors can be free from unjustified coercion. Nonetheless, both structures are aimed at regulating the distribution of social goods and ensuring relations of justice between members of a given community.

The following chapter attempts to identify a non-exhaustive list of such social goods whose just distribution is threatened by AI developments. Before examining the economic and political realms in more depth, some general remarks on the threats AI poses to (just) social life are made.

1.1 General Remarks

While AI developments are naturally challenging the prevalence of justice differently in each realm of human life, some of its developments are affecting most of them in similar ways, hence a separate subchapter is devoted to them.

1.1.1 Marketisation of Personal Data

The first and perhaps most important development of social life that AI enables is the platformisation of various (social) goods that had previously been publicly distributed (Filgueiras, 2023, p. 8). Although Filgueiras (2023) in his research is more concerned with the use of AI in the educational sector, his arguments can be translated into other spheres of life such as health care, politics, or the labour market by empirical observation as well.

Big data methodologies, facilitated by the existence of AI technologies, enable both

private and public actors to turn everything into data and subsequently create precise profiles of individuals to make economic profit from them (Filgueiras, 2023, p. 2). Those methodologies have enabled strict citizen surveillance which presents an infringement on fundamental human rights, especially political freedoms (Santoni De Sio et al., 2021, pp. 17-18) while also having the potential to result in a form of surveillance capitalism[2].

It is precisely this form of economic organisation that has the potential to disrupt social justice as we currently understand it and create problems of accountability in all areas relevant to social justice (e.g., Filgueiras, 2023; Hermansyah et al., 2023; Narayanan et al., 2024) since the private is made public and used for the economic benefit of a third party. An individual's vulnerabilities are thus exposed and the right to privacy, which is a fundamental building block of a dignified (human) life, is violated all for the economic interests of another actor.

1.1.2 Algorithmic Bias

It is not only for economic profit that individuals' fundamental rights are at risk of becoming potential rights (Wagner, 2019, p. 85) but the developments of AI become even more problematic considering that those technologies are more often than not centred around an already-dominant group's bias (Rafanelli, 2022, p. 1; Sloane, 2019, pp. 5-7).

This is because dominant social groups are overrepresented among both the engineers developing AI and among those whose data the mentioned technologies are trained on, compared to other social groups (Rafanelli, 2022, p. 1). Not only is it problematic because AI is political and always includes not only biases in terms of perspective taken on issues but also acts on a certain set of values (Sloane, 2019, p. 2). Therefore, it is inevitable that social injustices are amplified if no countermeasures are taken (Hermansyah et al., 2023, pp. 155, 164).

1.1.3 Attempts for democratic regulation

However, the regulations formulated in order to make AI design more inclusive and intersectional would have to be arrived at in a liberal democratic manner. Yet, I shall argue that regulating AI in developments in a way that is compatible with human rights and other ethical as well as moral standards is close to impossible if those measures are to be agreed

[2] Surveillance capitalism is a "new economic order that claims human experience as free raw material for hidden commercial practices of extraction, prediction, and sales […] that is best understood as a coup from above: an overthrow of the people's sovereignty" (Zuboff, 2019, Section THE DEFINITION).

on following the democratic process of deliberation.

This is because governments cannot possibly collect sufficient information at each stage of AI development to formulate effective regulations that respect every possible perspective taken by all actors in a given community (Néron, 2016, p. 716) which is especially problematic in democratic regime types in which deliberation and the finding of a consensus play a crucial role in the formulation and legitimising process of laws (Ferretti, 2022, pp. 257, 259). However, even if there are attempts to regulate AI in this way, those attempts are often criticised as not involving citizens enough and merely benefitting the AI industry (e.g., Kak, 2020, pp. 1-2).

It follows that (democratic) political decision-makers are presented with the following dilemma. Either they regulate AI developments promptly and thereby risk those regulations being perceived as illegitimate, or they do not regulate them and thereby risk unequal treatment of citizens by the mentioned technologies. Either way, the practice of social justice at large might be at stake.

1.2 Vulnerable Goods in the Political Realm

1.2.1 Data and Privacy

To begin with, it should be stressed that most importantly, AI and the use thereof is and will most likely continue to blur the lines between what is public and what is private (Zuboff, 2019, pp. 181-182) not only in terms of data leakage and misuse (Hermansyah et al., 2023, p. 155) but also in non-consensual data-gathering for both research but also citizens' surveillance (Crawford, 2021, pp. 104-105).

While in the early stages of AI development, the people who participated in its training "gave full consent" to the use of their personal data and possible algorithmic bias was mentioned in the section on limitations in a pioneer project's final report, this practice changed after 9/11-attacks (Crawford, 2021, p. 105). In the early 21st century, public demand for rigid security measures spiked and tracking of individuals online took off. This practice changed the standard way of gathering data needed to feed the new AI technologies with more of it. Nonetheless, it was not only infringements on individuals' right to privacy but those early developments also resulted in increasingly biased results. This is because those privileged enough to afford smart technologies at the turn of the century were able to publish their data online and voluntarily reveal more information about themselves than necessary so that even more data could be extracted without them necessarily knowing

(Crawford, 2021, pp. 105-108) which was subsequently done by both private and public actors (Crawford, 2021, pp. 109-111).

I argue that this need for an increased amount of data was handled in a way that resulted in illegitimate infringement of individuals' right to privacy. This is because their data was extracted and used for purposes other than the ones its owners[3] had given their consent to. Because the AI developed on this data has since been used to coerce citizens, the loss of privacy entails a shift in the power dynamics between those who manage data i.e. powerful private companies or even the government and citizens.

If nothing is done unseen and no action can be hidden from authorities, there is little to no room for human freedom. It is easy to interpret many kinds of actions as crimes and dismantle counterarguments before they are even voiced due to the information asymmetries that result from the ability to analyse huge amounts of data which AI makes possible.

Regardless of whether this relation of injustice arises between citizens and a public authority or between an individual and a private actor, the subject's consent would be crucial for an infringement of their privacy to be fully legitimate. Consent in this sense means that something happens to an individual on a voluntary basis (McDermott, 2008, p. 13) which also implies that for consent to be meaningful, it is necessary that an individual has the option to leave in the case tacit consent is assumed (McDermott, 2008, p. 18). Yet, this is not given in the case of monitoring and thus surveillance of various public spaces (e.g., Crawford, 2021, p. 10).

Furthermore, the argument that one is simply faced with a trade-off between whether to participate in public, thus social, life and expose oneself to surveillance or retreat oneself from it to escape surveillance, is invalid. Even if one managed to live a life completely detached from all kinds of AI, in order to live a fulfilling human life, one would have to connect with others and therefore re-enter society from time to time.

This is because of our evolutionary traits to seek meaningful interpersonal relationships with a small group of people and closely cooperate with them for our survival. It is those social relationships which are the main motivating factor behind our actions. Without them, it has been proven that we are likely to fall into depression, become physically sick, and die (Bauer, 2007, pp. 36-39). Therefore, it simply cannot be argued that there is a meaningful option to exit from places from which data is extracted and analysed, especially

[3] In the following part, I shall refer to those individuals who data is extracted from as the legitimate 'owners' of their data.

considering that even without using any digital devices, one is recorded on security cameras and subjected to leave personal data in some other public places, such as airports or the bank, that human (social) life is not possible without going to.

The entire concept of having to give up on privacy and potentially, what I shall call, 'selling one's own person' in order to take part in social life and thereby fulfil the basic human need of social interaction has gained unprecedented speed during but also after the outbreak of the Covid-19 pandemic in 2020. Already in 2020, scholars warned about the "undeniable threats to privacy, individual freedoms, and democracy" (Barriga et al., 2020, p. 2) that the introduction of quasi-mandatory, AI-based tools used for disease control would pose to the public.

Furthermore, the worries that the cause of disease control could be used as an excuse to extend citizen surveillance to areas of life which are unrelated to public health (Barriga et al., 2020, p. 2) have been proven to be legitimate. Backlashes specifically in regards to democratic accountability of policies as well as amplifications of authoritarian tendencies have been reported (Freedom House, 2023, pp. 11, 13-15, 27, 29, 34).

1.2.2 Citizen Control

Perhaps one of the most prominent examples where AI-based monitoring of citizens and hence surveillance and a subsequent "tyranny of numbers" (Filgueiras, 2023, p. 3) has already become a reality is in China's Social Credit System.

This system tracks and subsequently analyses every single one of a person's actions in all areas and aspects of life so that a specific *social score* can be determined. However, because of the huge amounts of data needed to arrive at such a score, an AI technology capable of processing the mass of data is used (Backer, 2019, p. 211). Coming back to the problem I outlined in the previous part, the judgements AI makes about ordinary citizens are unlikely to be unbiased but will rather reinforce existing biases and amplify existing injustices (Rafanelli, 2022, p. 1).

Furthermore, the necessary data is not collected without a specific purpose either but rather to determine a person's alignment with state ideology and to determine whether or not that person's actions adhere to the prevalent expected (state) norms and values. In case of a low score, a person could be deprived of basic social services or even be deprived of some of their fundamental human rights e.g. the freedom of movement, in case of official code of a violation of the official code of conduct (Xu et al., 2022, p. 2231) and thus has great

repressive potential (Xu et al., 2022, p. 2243).

While in the West this method of data collection and the subsequent evaluation of individual agents is mainly employed in business, hence the private sector (Backer, 2019, p. 214), in other parts of the world e.g. in China, the system is used for political purposes by incumbent elites (Xu et al., 2022, p. 2231) which is problematic in more than one way. For instance, it can be assumed that individuals give tacit consent to businesses to use their personal data because their freedom of movement, creditworthiness, or access to education and means of transportation does not depend on that one single business (unless that business holds significant market power).

However, as a Chinese citizen one can hardly free oneself from state surveillance because even if one wanted to, the mechanism would most likely detect dissent, lower one's social credit score, and potentially deprive one of the ability to buy transportation tickets (Xu et al., 2022, p. 2230) so that the option to exit is not given. This leads me to conclude that China's Social Credit System is a perfect illustration of how states could deprive and are already depriving individuals of their basic right to movement as well as their free development of minds i.e. their capacities to reason.

Furthermore, after having been identified, those dissidents will be repressed and coerced in a targeted way, usually causing a lot less public backlash than more overt repressive techniques (Xu et al., 2022, p. 2231) so that like-minded persons who also disagree with the incumbents' way of governing are harder to find. Subsequently, the establishment of meaningful interpersonal relations on the basis of ideological agreement, other than the official state's, might be hindered.

Not only is it hard to find them but in the process of searching the AI will already ascribe the individual a lower social score that will most likely result in isolation of persons (Xu et al., 2022, p. 2241) which can only be undesirable in light of the basic human need for interpersonal relations (Bauer, 2007, Chapter 2). Putting this together with the findings that individuals will go as far as to spread unverified or even misinformation to fit in with their peers (Špecián, 2022, pp. 85-88), it can be concluded that citizen control using AI technologies can be highly effective and secure incumbents in their office.

1.2.3 Dignity and Autonomy

Additionally, the entire process of ascribing numerical scores to persons, regardless of whether it is done by an AI or another person, goes against the common understanding of

human dignity. If that number subsequently also determines the individual's freedom of speech, movement etc. it can be argued that the equality between all citizens as well as their autonomy are threatened.

In the Chinese case, the replacement of human agents by an AI that makes 'decisions' on behalf of citizens i.e. 'the scored persons' can be said to be a non-consensual infringement on a person's autonomy. This hypothesis has been confirmed by Xu et al. (2022) who conducted a study on the perceived legitimacy of the Social Credit System in China. They found that the more a person knows about the repressive potential of the system, rather than merely being informed about its alleged social benefits, the less likely they are to support the system (Xu et al., 2022, pp. 2238, 2242).

Therefore, I argue that AI has enable systems to emerge that present a severe infringement on citizens autonomy and equality of opportunity for their personal as well as professional lives. Furthermore, the misuse of AI technologies shifts power away from citizens and makes them significantly more vulnerable to government control which in turn decreases the scope within which a free, self-determined life is possible while still trying to take part in the (political) community's life.

1.2.4 Communities and Misinformation

However, it is not only the individual's role and freedom within a given community that the developments of AI pose a challenge to, but also the strength of the existing communal ties itself. While the use of AI in Social Credit Systems such as China's might lead to more (political) stability and seemingly more "harmonious and amicable interpersonal relationships" (Backer, 2019, p. 213), others argue that AI might lead communities onto a different path.

Since most of the algorithms[4] are designed by private agents who are interested in their own economic profit (Špecián, 2022, p. 93) rather than the communal good, when operating on the assumption of free and perfect markets, the following scenario can frequently be observed.

The main objective of AI-based technologies is to generate a maximum amount of economic revenue and social media platforms are based around on AI algorithms. Those algorithms, however, are not designed for a neutral purpose but always with a certain goal

[4] When writing about algorithms in the parts that follow, I refer exclusively to those which are AI-based. The term *algorithm* is thus not to be used interchangeably with the term *process* but rather as an AI tool which gives different outputs depending on the data input of a user.

in mind and bias included, as I have elaborated on in part 1.1.2. In the case of social media, this goal would be to "maximize the platform's revenue generated by ads or sponsored links, striving against the constraint of the user's available attention" (Špecián, 2022, p. 81).

Therefore, content displayed to users would be tailored to their individual interests which effectively means that they will be thrown into one of many epistemic bubbles in which their beliefs will only be confirmed but seldomly contested (Špecián, 2022, p. 81). Because of the human need for peer recognition (Bauer, 2007, Chapter 2) and the way social media is designed (Špecián, 2022, p. 83), fake news are easily spread on those platforms. However, against the common perception of individuals merely being passive agents, they are involved in the spread of fake news themselves.

Since the online space is technically a private space governed by private businesses, legitimate authorities have no means to effectively intervene to try to stop the spread of disinformation (Špecián, 2022, pp. 79-80). Nonetheless, the spread thereof has created problems for the persistence of communal ties between members of a given traditional community.

This problem becomes apparent when one considers the spread of fake news combined with the emergence of digital echo chambers, in which people unlearn how to deal with disagreement. The combination of both phenomena separates individuals from the reality of the social circumstances of their respective communities while strengthening the ties to those in one's epistemic bubbles who do not necessarily need to be members of the same political community. Subsequently, this inevitably leads to a growing potential for (political) polarisation (Špecián, 2022, p. 83) as well as potentially unstable democracies.

Furthermore, it has been observed that communities whose members identify themselves with each other regardless of their differences are becoming less and less common. Pluralistic societies with strong ties between all members are thus on the decline (Mounk, 2023, pp. 11-14). This phenomenon is further amplified due to the decreased number of interactions in the way that the shared rituals which have traditionally kept communities together are being lost (Han, 2017, pp. 32-33, 41).

While in the immediate, physical space, interactions are governed by social norms and conventions, they are only limitedly translated into the online sphere of algorithms. Whereas in the 'offline' realm, police, hence state representatives who are bound by law and institutional practice, are in place to regulate citizens' conduct, in the digital sphere this is done by private platforms themselves. Yet, critics of this practice have argued that private

actors' regulation practices often seem to be rather arbitrary, leaving room for interpretation and cannot be considered a legitimate infringement on a person's right to freedom of speech (Ferretti, 2022, p. 246).

1.2.5 Deliberation

The effect of misinformation being easily spread online because of peer pressure is further amplified by the publicity of online statements. Because of this, the spread of unverified or even false information can be seen as proof of loyalty to a group (Špecián, 2022, pp. 91-92) so that the fraction between societal poles can be expected to deepen. The result for society would be that rational deliberation about shared social meanings or political institutions hardly remains possible.

Furthermore, it is only logical that within the existing echo chambers prejudices about other social groups might prevail and potentially lead to a dehumanisation of the 'Other' (Han, 2017, p. 13). Considering that the time spent online has increased gradually over the past 10 years, although not linearly (Kemp, 2023), the threat to social unity is growing.

1.2.6 Power Structures

1.2.6.1 Transfer of Power onto Private Agents

In addition to the increasingly easy spread of misinformation on social media platforms, another issue stands out. It is no secret that information and informed decision-making are crucial to the proper functioning of participatory political systems. Therefore, the media is sometimes even considered the fourth pillar in the separation of powers of such systems (e.g., Schneider & Toyka-Seid, 2024). Yet, others argue that the media has already amplified the concentration of power even before AI was introduced to the field (e.g., Sánchez Muñoz, 2002, p. 281). I shall follow the line of the latter argument and extend it to the concentration of power in the hands of not only government but also private agents.

Through the emergence of the digital sphere and AI, the power of channelling and verifying information is shifted away from "the numerous gatekeepers and crude audience targeting that kept fake news at bay" (Špecián, 2022, p.80) onto private agents who have no public mandate to filter and gatekeep the information which is spread among members of society. Regardless, they are left with no choice but to do so and are left to regulate content based on rather vague legal formulations (Špecián, 2022, pp. 93-94).

While some of the leading tech companies e.g. Google or Microsoft have introduced self-imposed standards on how to combat the problem of misinformation on their platforms, some critics have devaluated them as mere marketing strategies with no sincere intentions to improve the situation for the affected disadvantaged or even defamed persons or groups. Wagner (2019) for example, argues that by letting corporations decide which pieces of information to remove and what to retain, they are given the legal right to infringe on individual liberties (Wagner, 2019, p. 2) which should usually be reserved for legitimate state authorities (Ferretti, 2022, p. 246).

However, not only is the possible infringement on individual rights made by private agents problematic but those infringements cannot be expected to be of the same magnitude for all social groups. Coming back to the problem of algorithmic bias I have explained at the beginning of this chapter, those problems will affect the most vulnerable groups of domestic societies disproportionately (Sloane, 2019, p. 5) and amplify existing (global) power imbalances in international politics (Kak, 2020, pp. 309-310).

1.2.6.2 Global Balance of Power

In addition to the weakening of the bonds among members of traditional political communities, the lines between what is perceived to be a threat from within a given society and what is considered an external threat which would have to be fought collectively are blurring as well.

One could argue that the use of AI and probably even more so the digital sphere is greatly contributing to those developments since it facilitates transnational communication (Adamson, 2005, pp. 36-37). Furthermore, it clears the way for individuals to find like-minded individuals in the echo chambers algorithmic social media design is likely to throw an individual into and reassure them of their potentially radical or even extremist ideology (Špecián, 2022, p. 81).

Moreover, Adamson (2005) concludes that "[t]he international security environment inhabited by states looks less and less like a system of unitary state actors operating in anarchy, and more and more like an emerging, yet unevenly developed and weakly institutionalised, global polity" (Adamson, 2005, p. 45). This means that not even democratically elected people's representatives would be fully capable of protecting their communities.

This effect is further amplified by the rise of authoritarian states on the global stage.

While in the early days of the internet, it was mainly used to solve collective action problems of democratic resistance movements in authoritarian states, since the Arab spring, the internet has been increasingly instrumentalised and censored by authoritarian incumbents to stay in power (Hellmeier, 2016, p. 1160). Nonetheless, not only can the internet be censored but it is governed by algorithms that incumbents can implement their bias into so that propaganda which is tailored to an individual's irrational tendencies can be spread easily.

Taking this together with the fact that the countries who are globally leading in the development of AI are either classified as authoritarian[5] e.g. China or do not heavily regulate AI e.g. the United States (Ulnicane, 2022, pp. 254-255), it becomes even more problematic. Since communication across political and language barriers has become easier (Adamson, 2005, pp. 36-37), authorities have been relying heavily on the use of AI technologies in order to identify possible threats (Crawford, 2021, p. 105). Considering the inevitable bias in all AI systems explained above, it becomes clear that the most prominent AI technologies are not centred around principles that emphasise human freedom, equality or autonomy.

Furthermore, communities who emphasise those principles are in a disadvantaged position to keep up with their own developments of said technologies because of the time needed to gather enough data to include various perspectives (Ferretti, 2022, p. 246). It follows that, in order to keep up internationally, liberal communities are becoming dependent on those who disregard some of their fundamental values, hence power is increasingly shifted to more authoritarian states in respect of international negotiations.

1.2.6.3 Consolidation of Incumbents' Power

Not only is the global prevalence of political freedoms endangered by the developments of AI but that of an individual's freedom of choice as well as human autonomy too. This is because incumbent elites are trying to gather valuable information about an individual in order to make sure a person's behaviour is in line with their objective to stay in power.

To do so, nudges are the most cost-efficient, both in economic and political terms since they do not create a need to change anything about the given incentive structure, provision or access to information or alternatives offered. They are a mere reframing of the choices available to an agent and ideally intended to help an individual make a rational decision which is in line with their own interests (Špecián, 2022, pp. 111-112).

[5] see Freedom House (2023)

However, it is easy to see how this method would be misused by those who hold more information over the other i.e. by those who have access to AI technologies to collect and analyse data over those who do not. Regardless, even the well-intended form of nudges presents a paternalist method (Špecián, 2022, p. 111) and paternalism can always be seen as an infringement of an (adult) person's autonomy if no meaningful alternative choice is provided (Laitinen & Sahlgren, 2021, p. 5).

I further argue that, in case of severe information asymmetry between two agents that clearly exists in light of AI, nudges can be designed in a way that, objectively speaking, a meaningful alternative is provided that the individual, from a subjective point of view, cannot see (Krupiy, 2020, p. 22). Therefore, it can be said that nudges do not necessarily preserve individual freedom (cf. Button, 2018, p. 1036).

The threat of nudges which interfere with an individual's freedom presents such kind of asymmetry of power Durkheim had already warned about as "the most dangerous form of inequality [….] that make 'conflict itself impossible' by 'refusing to admit the right of combat'" (Zuboff, 2019, p. 184). It follows that incumbent elites' access to AI technologies is not only infringing on individual freedom and autonomy, but also minimises the subjects' means to defend themselves from injustices in relationships of unequal power.

1.3 Vulnerable Goods in the Economic Realm

Because AI interferes with social justice in the economic realm in similar ways as it does in the political one, I shall not delve into the details here but rather outline the main challenges AI presents to a society's economic life with an emphasis on the aspects of labour, education, and information available to market participants.

1.3.1 Labour Market

1.3.1.1 Employment

Perhaps one of the most repeated arguments related to AI and the economic sphere is that AI technologies might lead to mass unemployment because of an expected productivity gain when choosing a machine or, respectively, an AI to do a human's job (Zuboff, 2019, p. 181). However, analyses of how AI is currently changing the labour market arrives at different results.

Here, the argument is that the recent developments of AI are merely presenting a

continuation of the changes the labour market has been undergoing since the start of industrialisation. The earliest of such changes was most likely that humans resorted to operating the machines which had been designed to do manual tasks e.g. weaving instead of the humans themselves. In times of AI, the shift that is occurring is that human workers are increasingly engaged in non-routine tasks in which unexpected occurrences can happen. This is because it is hard to train AI for unexpected events, hence human workers are still better at some jobs than any AI can ever be imagined to be (Moradi & Levy, 2020, pp. 272-275).

For those jobs which are, and most likely will remain, more suitable to be done by humans rather than AIs, specialised training and knowledge are needed. It is precisely in this regard that new power imbalances and thus injustices in terms of equality (of opportunity) and freedom of choice arise.

"[D]ilemmas of *knowledge, authority, and power*" (Zuboff, 2019, p. 180, emphasis in the original) refer to the problem of how to determine which role an individual will take in a society classified by a *division of learning*[6]. This system of social organisation distorts equality of opportunity to learn new skills and thus qualify oneself for better jobs. The question of who holds the authority to decide who is to learn more and who is not as well as what kind of information to share with others and which pieces of information to withhold becomes central to a society's organisational structure (Zuboff, 2019, pp. 180-181).

Coming back to the biases which are inherent in all existing AI technologies, it is only natural that already disadvantaged social groups will not be given the opportunity to learn as much as others (Filgueiras, 2023, p. 7), possibly due to perceived lower (cognitive) abilities. This has the potential to result in a worsening of working conditions for already disadvantaged social groups.

1.3.1.2 Balance of Power in Contractual Relationships

Furthermore, the use of and reliance on AI in human resource decision-making is creating some new, perceived injustices (Al Samman & Mohamed, 2024, p. 19) and other problems in the power dynamics between employer and their employees (Crawford, 2021, pp. 54-57).

A blind reliance on AI technologies has the potential to lead to a "tyranny of numbers" (Filgueiras, 2023, p. 3) in all areas of life. If an AI is fed with inconsistent and

[6] Zuboff (2019) sees a division of learning as rather problematic. In such a form of social organisation the interdependence of individuals prevalent in systems of a division of labour is replaced by significant asymmetries of power (Zuboff, 2019, pp. 180-186).

unstructured data, the use thereof results in a reinforcement of existing injustices or can even create new forms of discrimination, inequality and oppression (Filgueiras, 2023, pp. 7-8).

Applying this argument to the labour market, one can see how groups who have not traditionally been represented in a given sector are disproportionally disadvantaged when it comes to fully automated human resource decision-making (e.g., Crawford, 2021, pp. 129-130). This phenomenon could show both regarding job security and thus income stability but also in discriminatory practices in the recruiting process (Moradi & Levy, 2020, pp. 282-284), amplifying the existing patterns of domination within a given society (Sloane, 2019, p. 7). Subsequently, societies might become even more segregated if one considers economic insecurity as a threat multiplier.

Additionally, when processes are entirely outsourced to AI i.e. no human agent is involved in the process, the subject's identity is depersonalised (Filgueiras, 2023, p. 8) so that their human experience is dispossessed (Zuboff, 2019, pp. 232-233). In this case and the case of AI-based performance monitoring individuals are "dehumanized as just more data" (Crawford, 2021, p. 94) which inarguably goes against their human dignity.

Moreover, AI is enabling strict and precise worker surveillance which makes it possible to "*redistribute the risks and costs of* […] *inefficiencies to workers* while serving a firm's bottom line" (Moradi & Levy, 2020, p. 279, emphasis in the original) by, for instance, staffing according to acute demand (which entails that employees are called into work on short notice) or only paying for the execution of a specific task (leaving the preparations therefor uncompensated) (Moradi & Levy, 2020, pp. 279-284). Thus, neither a balance of power between employer and employee nor a stable income is given. Furthermore, it can be argued that through increased workers' surveillance that AI makes possible, the information asymmetries which have traditionally protected workers to some extent, are minimised, if not eradicated.

I argue that a form of limited information asymmetries between employee and their employers favouring the former would be necessary to maintain a balance of power between the two parties. This is because employees often have no choice but to take on a job and thus have less bargaining power than the employer in the design of the contract. This power imbalance can later be brought back into equilibrium once the employee is on the job since the employer will have to try their best to prevent an agency slack[7]. However, the occurrence

[7] Agency slack in relation to the labour market refers to an employee's decreased efforts in completing the assigned tasks (Hawkins et al., 2006, p. 8).

thereof is much less likely in case of perfect surveillance because of which the employee is in constant fear of underperforming and losing their job.

The illustrated case of the balance of power between employer and employee serves as a prime example of how AI contributes to a disequilibrium of power in various types of contractual relationships in which information asymmetries occur and can be translated into other realms of life as well.

1.3.2 Education

Nevertheless, it is not only in the realm of labour that AI developments and the use thereof are amplifying existing injustices but taking a step back to examine the realm of education, one can observe similar developments. Yet, in this realm, some scholars argue that the employment of AI technologies in order to tailor education more closely to a person's needs and talents to craft an individualised curriculum in accordance with those talents is not inherently bad, although not without problems either (Filgueiras, 2023).

The benefits of individualised school curricula are also intuitive because when children are given the chance to discover their passions early on and are later provided with adequate opportunities to delve deeper into their areas of interest without having to slow down for their classmates to catch up, a given society's economy has the potential to be transformed into a highly specialised one that each individual is capable of contributing greatly to. Respectively, if too little is demanded of students, they might lose passion and interest in what they are doing, deem education a necessary evil and thus end up in places where their potential is not fully realised. Therefore, some have gone as far as to argue that it is a moral obligation to individualise a child's schooling curriculum (Adelsberger-Hoss, 2021, pp. 14-17).

Nevertheless, the realisation of such a project centred around an AI-based tool to assess each person's potential is not without problems. Firstly, there would most likely be privacy concerns since basic education is mandatory in most countries (Heymann, 2014) so no opt-out option is provided. Secondly, algorithmic biases would reinforce existing injustices and create new ones (Filgueiras, 2023, p. 7) which, in this case, would mean that not all are assessed equally. Some individuals would subsequently be provided with fewer opportunities to learn and be deprived of opportunities to educate themselves further.

For the above-mentioned reasons, I conclude that forcing the division of learning[8] on

[8] see Zuboff (2019, p. 180)

a society, the consequences of which are explained in more detail in the previous section, are not inherently bad. However, from the perspective of equality between persons as well as freedom of choice, it is highly undesirable since it amplifies existing injustices in those realms.

1.3.3 Market and Information Asymmetries

In the economic realm as such, AI is also posing threats in terms of information asymmetries on market participants. Until now, one of the assumptions of neoclassical economics was that human economic behaviour follows the homo oeconomicus model that is completely rational and unbound by the contextual situation. However, this assumption does not hold under real-world conditions. Therefore, psychological explanations of human behaviour are more precise in predicting actions as well as choices due to their ability to predict irrational behaviour (Špecián, 2022, p. 22) such as the "moralization of markets" (Stehr & Adolf, 2010, p. 217).

Through the availability of a perfect AI, this could be changed, though, or the ideal could at least be approximated. What weak AI is doing is analysing amounts of data that are too big for any human agent to grasp and checking it for recurring patterns (Moradi & Levy, 2020, p. 272) to predict future occurrences which might include market developments and shifts in either the supply or the demand curve or market failures such as externalities, market power etc. It follows that AI has the potential to decrease the need for government intervention in the market and lead to a more just allocation of market goods among a group of fully rational agents.

However, the problem of the marketisation of non-market goods outweighs this argument since real-world markets are imperfect markets. It follows that there should be realms of life which are not exposed to the market exchange of goods (Anderson, 1995, Chapter 3).

When an institution is exposed to market forces, it will have to be administered like a business and be incentivised to operate as cost-efficiently as possible and to "extract[…] resources [from persons in order] to transform them into various forms of products and services" (Filgueiras, 2023, p. 2). The individual human experiences are subsequently rendered from the data it generates and are transferred to agents who are monetarily incentivised. It is those agents objective to sell more of a product or service so that the 'owner' of data merely becomes its creator (Zuboff, 2019, pp. 232-234).

Therefore, I argue that the marketisation of various, traditionally publicly provided social goods, is threatening human freedom, autonomy, as well as the right to self-determination. Furthermore, information asymmetries can be exploited in order to make more profit off of an individual who did not necessarily consent to their data being used for the mentioned purposes. While AI is not the sole cause of it, it constitutes the technological tools necessary to facilitate those developments.

1.4 Chapter Summary: Vulnerable Social Goods

To summarise my findings, I conclude that AI is amplifying existing social injustices in the following ways. Firstly, it is infringing on individuals' privacy which is problematic because it is in private that citizens have traditionally been free from (state) domination (e.g., Krupiy, 2020, p. 9) and been able to exercise their human irrationality. The infringement of privacy is thus also an illegitimate infringement of a person's freedom.

It follows that, secondly, existing imbalances of power are reinforced and a fight against those made significantly harder (e.g., Krupiy, 2020, p. 22). Not only might crucial information be withheld but incumbent elites can frame discourses in a way that makes every act of resistance appear to be a fight against the community (Zuboff, 2019, p. 177) and thus result in the quasi-expulsion of those members who slightly disagree with the incumbent elites. Considering that interpersonal human recognition is a basic human need (Bauer, 2007, Chapter 2), acting on the principles of rationality and moral convictions is increasingly becoming undesirable.

However, the developments of AI are not only amplifying power imbalances within a given society but also lead to a transfer of power onto transnational actors as well as private actors who are involved in the design and provision of AI (e.g., Adamson, 2005, pp. 33-37; Kak, 2020, p. 2). This raises concerns because an individual usually transfers some of their rights onto a state, consisting of, ideally, democratically elected peoples' representatives, who will in turn provide protection and security (Ferretti, 2022, p. 247) while the new incumbents of power do not have the citizens' consent to infringe on their liberties. Because they do so regardless, human agents have become less autonomous and participatory political processes have lost their importance.

Furthermore, through the emergence of epistemic bubbles, interpersonal ties which have traditionally kept various members of a given society together, are lost (Mounk, 2023, pp. 13-16). Since one cannot say that the employment of AI technologies results in a

consideration of all kinds of a person's strengths and weaknesses, inequality of opportunity is being amplified as well (Filgueiras, 2023, pp. 7-9). Not only because some strengths might not be noticed but also because less talented individuals are deprived of their chance to 'prove themselves' through hard work and determination.

Most importantly, however, the increasing influence of AI in all areas of life has led to the marketisation of various non-market goods. The suppliers thereof are subsequently motivated by monetary rather than moral motives (Filgueiras, 2023, pp. 2, 6) which results in the dispossession of an individual's human experiences (Zuboff, 2019, Chapter 8) and represents a dehumanising practice in itself (Crawford, 2021, p. 94).

It follows that social and political institutions are increasingly designed on the assumptions of economic models that have proven to not hold under real-world conditions. This phenomenon entails that inevitable market failures have even more futile consequences for the prevalence of just social practices.

2. Conceptual Limitations of J. Rawls and M. Walzer's Theories

Now that those social goods whose just distribution is most vulnerable to the developments of AI have been identified, the following chapter is devoted to shedding light on where the conceptual limitations of J. Rawls' and M. Walzer's theories of justice to protect the just distribution thereof lay.

This chapter aims to precisely investigate where and why the given theories fail to accommodate the challenges AI poses to social justice. Without the results of this chapter, the research question of whether it will be enough to amend those theories or whether they will have to be replaced by new theories cannot be answered.

2.1 J. Rawls' Justice as Fairness

To begin with, Rawl's theory of justice belongs to the category of *institutionalist* approaches to ethics. Institutionalists "defend[…] a 'division of moral labor between governments and the private sector" (Ferretti, 2022, p. 240). As I have found in Chapter 1, however, the balance of power between private and public agents i.e. governments and businesses is being tilted by AI. It follows that private businesses are most likely to largely influence governments which means that the division of moral labour Ferretti (2022) elaborates on is, contrary to his argument, not a division of moral labour between two

independent agents anymore.

Furthermore, power is being diffused from traditional units of states and their respective political incumbents' power towards transnational actors (Adamson, 2005, p. 44) so that political and social institutions might lose their influence on the lived experience of human life. That is because experiences are translated into data before governments can take action to protect their citizens (cf. Ferretti, 2022; Zuboff, 2019, pp. 232-253).

2.1.1 Limitations of Justice as Fairness in its Global Application

In terms of funding for and regulation of research projects on AI, a comparison between the EU and other global actors such as China or the United States has shown an interesting phenomenon[9]. What the EU describes as its aim of "promoting a human-centric approach and ethics-by-[AI-]design principles" (European Commission, 2018) might backfire in the long run since it presupposes that the ethical principles for research and developments should be agreed on beforehand. In this way, I argue that the way EU authorities are trying to deal with and regulate AI developments is in line with most aspects of Rawls' theory of *justice as fairness.* This is because a framework consisting of the basic principles is arrived at in a deliberative process before any agent can act.

Although EU policymakers do not fully rid themselves from all their personal attributes, they deliberate about their shared institutional framework as free and equal agents. Despite keeping parts of their personal characteristics, I argue that in a way they do find themselves behind a translucent *Veil of Ignorance* since they cannot predict how AI will continue to shape the social reality they will find themselves in soon.

Once those basic shared institutions are designed, they are adjusted to meet the ever-evolving challenges AI is posing to social justice. Because of the practice of constant revisions and subsequent adjustments, it can be said that they are arrived at according to the Rawlsian method of *reflective equilibrium*. If AI developments were not a transnational threat to justice, this approach would work, and the EU's approach could be seen as a prime example of how to regulate AI in line with Rawls' theory of justice. However, this is not the case.

[9] This refers to the resources allocated to developing AI and thus the speed at which technological progress can be made. See (Ulnicane, 2022, pp. 254-255) for more information.

2.1.1.1 Global Order

In *The Law of Peoples* J. Rawls stresses that his theory can only be applied to certain kinds of peoples who are assumed to have agreed on a *basic structure* by consulting their capacity to reason (Rawls, 1999b, pp. 59-70). Subsequently, their institutions would protect individuals' freedom as well as the equality of all members so that fairness can prevail (Rawls, 1999b, p. 59). To classify as a liberal well-ordered people, "citizens [would have to be] represented fairly (reasonably), in view of the symmetry (or the equality) of their representatives' situation" (Rawls, 1999b, p. 31) both domestically and globally. Furthermore, it is a liberal people's moral duty to ensure that fundamental human rights are not heavily violated anywhere. This means that liberal peoples would have two options regarding non-liberal ones.

If a society is well-ordered but decently hierarchical it can be represented in the Society of Peoples (Rawls, 1999b, p. 63). However, the principle of non-intervention and thus communal autonomy as well as the self-determination of a people "will obviously have to be qualified in the general case of outlaw states and grave violations of human rights" (Rawls, 1999b, p. 37).

It can be seen that through the strengthening of authoritarian tendencies, AI constitutes a powerful tool for, the *duty of assistance*[10] can and must be applied to more cases i.e. well-ordered people intervene in hierarchical societies in order to fight the violation of human rights by "help[ing] a burdened society to change its political and social culture" (Rawls, 1999b, pp. 108-109). However, I shall argue that the effectiveness of such interventions is declining since hierarchical peoples are gaining in strength and number and fights for democracy are most effective from within a community (Freedom House, 2023).

Perhaps, rather than exclusion of those hierarchically organised peoples from the Society of Peoples, a demonstration of what it means for individuals to be secure in their rights would be more effective in incentivising a change in political culture than mere isolation or external intervention. Especially in times of digital echo chamber of information and algorithms which help incumbent elites to stay in power, one cannot assume that individuals who grow up under a repressive regime and have been nudged as well as manipulated all their lives are capable of rationally designing just institutions for the mutual benefit of all members[11].

[10] explained in more detail in *The Law of Peoples* (Rawls, 1999b, Sections 15-16)
[11] cf. Rawls (1999a, pp. 3-6)

This problem becomes apparent when one compares the case of the EU to that of leading AI developers. As explained above, in Rawls' conception of social justice the EU would be an exemplary actor when it comes to dealing with AI. Nonetheless, in other, non-exemplary countries e.g. China, much more funding is allocated to developing AI, hence it is realised at a much faster speed than in the EU (Ulnicane, 2022, pp. 254-255).

Not only is it faster but also more cost-efficient and less legally complex but the EU even becomes dependent on the import of such technologies (Varnholt, 2023) since the use of AI presents a significant advantage in keeping up with the fast-paced environment of contemporary global politics as well as economics (Bundesregierung, 2024). This is because of its capacity to absorb vast amounts of data and detect recurring patterns and thus predict possible future events allowing users to prepare adequately or to "achieve previously unattainable levels of accuracy" (Hermansyah et al., 2023, p. 156) in various realms of life where data is involved or generated. Therefore, it can be assumed that non-liberal, hierarchical peoples[12] are much more likely to influence well-ordered peoples rather than vice versa.

This dominance of hierarchical peoples over liberal, thus well-ordered, peoples is perhaps the main threat AI developments are posing to the global application of J. Rawls' theory of justice. His conceptual framework to deal with human rights abuses and other developments is solely based on an assumption of liberal hegemony and does not account for what happens in the case of well-ordered peoples' dependency on hierarchical peoples. The *principle of toleration* is only explained in one way, with respect to liberal peoples' duty to tolerate *decent nonliberal peoples* (Rawls, 1999b, pp. 59-60).

For the reasons explained above, Rawls' condition for when a given society should be tolerated and accepted in the *Society of Peoples* has been rendered unrealistic by AI developments. This is because the power to exclude or accept peoples is increasingly transferred into the hands of those actors who do not need to adhere to privacy rules and ethical codes of conduct so they have a significant epistemic advantage over liberal people who must adhere to the lengthy process of democratic deliberation that is often to slow to keep up with the fast-paced global AI environments (Ferretti, 2022, pp. 242-243).

However, examining the domestic application of Rawls's theory more conceptual limitations stand out.

[12] I classify those countries based on the Freedom House (2023) report which reports violations of human rights.

2.1.2 Limitations of Justice as Fairness as a Theory of Domestic Justice

Generally, Rawls' theory can be described as a rather minimal one that merely operates on the assumption of free, rational, and equal individuals (Rawls, 1958, p. 166) who come together to agree on a set of principles they would like to live by and design their institutions accordingly. What is in the back of Rawlsian moral agents is the aim of designing a structure for their society whose rules create a mutual advantage of cooperation for all members (Rawls, 1985, pp. 227-230). For the realisation thereof, it is necessary to be able to identify oneself with the other and have some trust in institutions (Rawls, 1958, pp. 187-189).

2.1.2.1 Common Sympathies and the Community

However, the transfer of power into the hands of agents who do not satisfy the conditions of being controlled by a liberal-democratic electorate (Rawls, 1999b, p. 24) is something that Rawls did not account for. Furthermore, "the Law of Peoples start[ing] with the need for common sympathies, no matter what their source may be" (Rawls, 1999b, p. 24), while going in the right direction of not being limited to national myths or longstanding political communities, is being challenged in its real-world application by the developments of AI.

Furthermore, Rawls (1999b) himself distinguished between the global and the domestic applications of his theory and argues that as a global theory of justice, the principles arrived at between different peoples' representatives in the *Society of Peoples* can only be thin ones and create a need for toleration of non-liberal forms of social and political organisation (Rawls, 1999b, pp. 59-70). However, this entails that on the domestic level, the principles must be thicker and, ideally, be strictly liberal (Rawls, 1958, p. 166).

If one takes, what Rawls in his later works, refers to as *common sympathies* (e.g., Rawls, 1999b, p. 24) it becomes clear that a community is by no means a given entity. I argue that on this account of what makes a community a relevant unit in which the *Veil of Ignorance* can be applied, it is not necessary to be physically close to one's peers but only to abstractly identify with them. Bringing AI into the picture and considering the problem of the creation of epistemic bubbles whose emergence is facilitated by the employment of AI-based algorithms in the digital sphere (Špecián, 2022, p. 81), other central building blocks of Rawls' theory of justice are challenged.

Rawls relies on a procedural account of justice which is liberal in nature and calls for

a focus on domestic institutions (Rawls, 1999b, p. 59). On the other hand, the internet has enabled increasing communication across larger physical distances and enabled the emergence of *common sympathies* on new grounds (Adamson, 2005, p. 36) that Rawls did not account for. This effect is amplified by adding AI to the picture which easily connects individuals with like-minded persons regardless of their geographical location (Adamson, 2005, p. 34; Špecián, 2022, p. 81).

The mentioned phenomenon is likely to escalate the fragmentation of traditional, geographically contained communities of peoples (Adamson, 2005, pp. 36-37) who he argues should be governed by strong institutions (Rawls, 1999a, pp. 6-7) and result in the prevalence of what he calls the "associationist social form […] which sees persons first as members of groups – associations, cooperations, and estates" (Rawls, 1999b, p. 68). This form of social organisation is not inherently bad since human rights can also include various associations' social rights (Rawls, 1999b, p. 80) and sometimes even necessary (Rawls, 1958, pp. 166, 170).

Nonetheless, the new forms of associations are gradually replacing common sympathies towards immediate others, who traditional institutions rooted in a common political culture are shared with (Adamson, 2005, pp. 36-37). One can see how the ties to online peers within the mentioned epistemic bubbles can be closer than to co-nationals because of the confirmation bias AI is trained to satisfy (Bauer, 2007, pp. 33-36; Špecián, 2022, p. 81).

Subsequently, strong common sympathies might result in a strong sense of moral duties (Walzer, 1983, p. 33) according to moral principles in line with the Rawlsian reflective equilibrium to others outside one's own community, hence persons whom the strong, domestic institutions are not shared with. Nonetheless, the institutional environment of others is not to be interfered with unless there are human rights violations (Rawls, 1999b, p. 36).

It follows that Rawls leaves little room for the realisation of partial ties to others outside of one's own institutional community that goes beyond the protection of peoples' basic human rights. Yet, AI is promoting the fostering of precisely those relationships between persons, in which at least the way people treat each other must be regulated (Rafanelli, 2022, p. 2).

2.1.2.2 Partial ties and information

Furthermore, the existence of epistemic bubbles makes it easier to withhold or downplay all sorts of information, including human rights abuses (Hellmeier, 2016, p. 1160). Those controlling socially influential AI also hold the power to decide which information is shared or retained even within the epistemic bubbles they allow to emerge (Rafanelli, 2022, p. 5). It follows, that within a political entity, different persons will be exposed to different information which influences their moral development as well as value system in a way. This might result in a blurring of distinct political cultures that Rawls wanted to retain (Rawls, 1999b, Chapter 2) to a certain extent and result in higher levels of polarisation within a society.

I argue that an application of the Veil of Ignorance in a highly polarised and unequally informed political environment is not possible in the way Rawls intended it to be done. Even though he emphasises that his thought experiment was a mere hypothetical procedure, it is not even hypothetically applicable since those epistemic bubbles are making it harder to reflect on "the recognition of the aspirations and interests of the others to be realized by their joint activity" (Rawls, 1958, p. 182). Therefore, the individuals might be unable to rid themselves of their personal attributes.

It follows that an individual who is brought up in an environment shaped by mis- and disinformation, in which one is hardly confronted with disagreement, will find it hard to rationally deliberate about principles of justice considering other persons' perspectives. This leads me to conclude that attempts to arrive at the principle of justice for a given society from behind the Veil of Ignorance will be governed by emotions rather than reason in the light of AI developments in the information sphere.

2.1.2.3 AI Decision-Making and the Reflective Equilibrium

The recent developments of AI not only threaten the prevalence of rational discourse in the original position but, as I shall argue, limit the variety of situations in which the Rawlsian method of *reflective equilibrium* can be applied as such.

Given the weakness of human nature combined with the convenience of outsourcing decision-making to presumably 'rational', 'intelligent', and 'omniscient' AI technologies, various problems arise. Those problems include a potential decrease in the necessity to adjust one's own moral principles to the given contextual situation so that moral learning becomes irrelevant (if no relevant moral decisions are made by the human agents themselves anyway).

I assume that human moral agents will increasingly rely on AI to make (moral) decisions for them, especially when it comes to designing their respective society's institutions. This is because when individuals do not feel like their actions can change much, they will prioritise something else (Špecián, 2022, pp. 56-57). Additionally, AI might be perceived as being able to make decisions on behalf of a person or at least give recommendations they will certainly follow since it is no secret that huge amounts of personal data are collected in many everyday actions.

However, humans' reliance of AI technology to make decisions for cannot be desirable in all areas of life. While relying on said technologies in some areas of life such as the decision making might be desirable under perfect world conditions of unbiased and complete data (e.g., Kak, 2020; Sloane, 2019) such as for the comparison of prices or other technical tasks allows for human agents to take care of more pressing issues, reliance on AI in more emotionally-complex decision-making is generally not perceived as fair (e.g., De Oliveira et al., 2022; Narayanan et al., 2024; Santoni De Sio et al., 2021; Yalcin et al., 2023) and would thus not be accepted in the original position nor after consulting one's moral intuitions in the revision process.

Furthermore, Rawls' (1999a, pp. 42-45) method of reflective equilibrium is naturally challenged if persons assume AI as being more suitable to design a given society's moral principles than themselves because AI has no reflective capacities (Brożek & Janik, 2019, p. 105). The problem is that does not undergo a process of moral learning but can only change its behaviour if its developer changes the code (Laitinen & Sahlgren, 2021, p. 3). However, Rawls' method is one that relies on the intrinsic motivation to adjust the principles of one's acting.

It follows that AI cannot act on any reflective equilibrium but only adhere to the principles which were first installed in it (parallel to the results of the first round of negotiations behind the Veil of Ignorance). While agents in the original position must be rational, Rawls' method of reflective equilibrium is meant to correct those fully rational principles to a lived human reality in which 'rationality' is imperfect.

2.1.2.4 Distribution After Principles of Contribution and Desert

Not only is the real-world application of Rawls' method made obsolete but his principle after which social goods ought to be distributed, *the contribution principle*, is becoming more problematic too.

This is because, in his theory, everything that goes beyond the basic goods needed for self-sustenance is to be distributed in accordance with how much an individual contributes to the community as such (Rawls, 1958, p. 166). Considering how AI is already being used to classify and assess children's performance in school and subsequently adjust their curriculum in line with how much confidence an AI has in their abilities (e.g., Filgueiras, 2023, pp. 4-5) combined with the epistemic biases which are built into AI (e.g., Rafanelli, 2022, p. 3; Sloane, 2019), the principle becomes problematic because the equality of opportunity to contribute to a community is threatened by AI developments.

I go as far as to argue that individuals in the original position would not agree on a distribution of social goods in line with this principle. Because AI does not possess 'human' attributes of irrationality or emotional intelligence (Brożek & Janik, 2019, p. 105) it can only evaluate persons based on their hard i.e. quantifiable, measurable performance. However, talents and motivation to work hard to overcome disadvantages in comparison to others cannot be quantified and thus not considered. Therefore, not only is AI limited in its abilities to assess all facets of a person's character and thus talents as well as natural aptitudes, but it also affects the prevalence of equality of opportunity as well as the freedom to work harder than others in order to overcome one's disadvantages adversely.

Furthermore, even if the contribution argument were to be agreed on in the original position it would be rejected after consulting one's moral intuitions in an AI-driven context. At first sight, it might seem that the contribution argument would still be fair. However, by becoming aware of one's personal attributes, individuals will realise that distribution according to contribution might not be desirable in all cases (e.g., Walzer, 1983, pp. 21-26) or even regard 'hard work' that goes beyond one's natural aptitudes as a social good worthy of protection in itself (e.g., Nussbaum, 2002, p. 457; Walzer, 1983, Chapter 6).

2.2 M. Walzer's Justice as Complex Equality

In contrast to J. Rawls, M. Walzer is often referred to as someone who has a hands-on approach to ethics and is concerned with a pragmatic application of his theory. His theory leaves room for value pluralism between different communities (Van Wyk, 2008, p. 258; Walzer, 1983, pp. 4 ff.) and allows for other values to prevail in different spheres of life within one community even (Walzer, 1983, p. 7). While those aspects go in the right direction given the developments of AI, Walzer's theory is not without problems either.

2.2.1 Community and Membership

One central building block of Walzer's theory is that of membership. He claims that everyone would be entitled to membership in a (political) community that would be hard, although not impossible, to exit. Membership would be central because only by being assigned to a specific group of people can injustices in one sphere be walled off in another (Walzer, 1983, Chapter 2). Yet, if one adds AI to the picture of Walzer's understanding of membership, the real-world application of his account of membership is challenged. This is problematic because relations of complex equality, hence of justice, can only prevail within a given community (Walzer, 1983, p. 5).

Not only is AI capable of "depersonalizing identities" but also of "compromising an idea of citizenship" (Filgueiras, 2023, p. 8) in a given (political) community. Through the algorithmic workings which facilitate the emergence of epistemic bubbles, the process of which I have explained in more detail in Chapter 1, individuals are less and less confronted with views, opinions, and perspectives that contradict their own. This leads to the reinforcement of social segregation based on externally ascribed identities even within existing communities (Mounk, 2023, pp. 2-4, 84-93).

The result is a strengthening of various sub-communities within one state who each feel powerless and unheard when it comes to changing the political landscape in the bigger picture. Because of this feeling of powerlessness to change the course of larger political events, groups will start searching for the enemy within their own political community (Mounk, 2023, Chapter 7). Thus, the degree of solidarity which has traditionally persisted within nation-states (Adamson, 2005, pp. 33 ff.) is being limited further to those to who one has immediate personal ties (Mounk, 2023, pp. 197-199).

Therefore, the very foundation, that of a right to membership (Walzer, 1983, pp. 31-32), of Walzer's account of what social justice consists of, is challenged by AI. More specifically, this phenomenon is enabled by the AI used for algorithms on social media platforms and what users make of it (Mounk, 2023, Chapter 5). However, Walzer's conceptual framework accommodates the existence of distinct moral duties between members of different communities (Walzer, 1983, pp. 31-34) but no satisfying solution for what to do in case a historically developed community breaks apart.

That is because the argument Walzer gives for what to do with newborns is insufficient to accommodate the challenges the developments of AI are posing to communal ties. It makes sense that newborns will shape their respective communities according to their

own moral principles and values while still identifying with the larger collective (Walzer, 1983, pp. 34-38). Nevertheless, AI is contributing to the strengthening or new formation of sub-societies, that can replace a common national identity with that group's one. Therefore, I argue that the existence of epistemic bubbles online has led to a situation in which conflicting *social meanings*[13] within one political community are possible. This results in a situation where the community in which social understandings are shared, do not coincide with the historically emerged political communities, the importance of which Walzer stresses for his theory of justice to be applicable (Walzer, 1985, Chapter 2).

2.2.2 State Behaviour

While membership in a community can be regarded as the most important condition for justice, Walzer states that "there is no other social good whose possession and use is more important than [power]" (Walzer, 1983, p. 285). As I have concluded in Chapter 1, however, the current developments of AI are threatening the balance of power in various ways.

2.2.1 Guards of Rights as Predators for Persons

Through the mentioned problems of fragmentation of traditional (political) communities, and the breaking off of many political boundaries, as well as the transfer of power onto private agents, I shall argue that social life is increasingly taking place in clubs rather than traditional states.

Walzer himself describes clubs as a form of association in which "only founders choose themselves (or one another); all other members have been chosen by those who were members before them" (Walzer, 1983, p. 41). Although he is aware that "we might imagine states as perfect clubs", he further emphasises that for moral life "states are like families rather than clubs" (Walzer, 1983, p. 41) by drawing an analogy of both the household as well as the state being a safe space for individuals that members can seek refuge in (Walzer, 1983, pp. 41-42). In this part of his theory, he completely disregards the possibility of a state's behaviour (or family structures) being the source of injustices.

Furthermore, intellectuals have found that even within families, existing injustices are not to be ignored in any examination of the prevalence of social justice in a community (e.g., Nussbaum, 2002). In this case, a specific personal attribute e.g. ranks in birth hierarchy,

[13] i.e. the value each community places on a certain good (Walzer, 1983, pp. 6-10)

gender etc. would influence every sphere Walzer mentions in the same way. Yet, for complex equality between all members to be realised, this is unacceptable (Walzer, 1983, p. 6).

Now, bringing AI into the picture this issue becomes even more pressing. Not only can the process of admission to a community be outsourced to dehumanising AI-based procedures but power imbalances are also amplified within a given community.

2.2.1.1 Treatment of Refugees

Each community has the right to decide on who to admit and who to reject (Walzer, 1983, p. 34). To exercise their right, however, administrative units have started relying on AI in their entry policies which is often regarded as a mere reinforcement of existing discriminatory practices in the admission process because of low-quality data. Those technologies are being used in regards to various types of migrants, refugees included (Vavoula, 2021, p. 483). However, Walzer argues that while there is no moral obligation to accept migrants, the case is different for refugees who must be taken in by a community if their fundamental rights are being threatened in their place of origin and the receiving community has enough resources to do so (Walzer, 1983, pp. 49-51).

Given that "AI systems used in migration, asylum and border control management affect people who are often in a particularly vulnerable position and who are dependent on the outcome of the action of the competent public authorities" (European Commission, 2021, para. 39), the naturalness of the practice of Walzer's account of moral duties towards refugees is threatened by AI.

When dealing with refugees, states are participating in a community's moral life. Through the use of AI-based admission policies, however, they are organised like a club rather than a family[14], where the dominant group instils their perspectives into the respective AI and thus influences whose application for asylum to admit and who to reject. I argue that by doing so, a migrant's appeal to admission e.g. on the grounds of unused resources (Walzer, 1983, p. 50) becomes ineffective.

Furthermore, AI merges all spheres of life by collecting data in all of them simultaneously and subsequently analysing it centrally. However, the conceptual framework of Walzer's theory is built on the assumption that the spheres can be separated. In light of AI developments and the use of those technologies, the realisation thereof is unrealistic. This

[14] see *Spheres of Justice* (Walzer, 1983, p. 41)

is because AI cannot differentiate between different (social) spheres but will extract as much data as possible from all of them at the same time (Zuboff, 2019, Chapter 5).

2.2.3 Limitations within a Community

However, not only is AI interfering with power imbalances with respect to moral questions toward outsiders but also within the communities themselves. This is because AI is creating huge asymmetries of knowledge, thus power, within a given community (Zuboff, 2019, p. 180). Those imbalances are amplified by the surveillance mechanisms AI enables allowing incumbents to effectively control citizens via systems e.g. the Chinese Social Credit System I have elaborated on in Chapter 1.

It follows that not only money can be considered a *dominant good*[15] but information and knowledge too. I shall argue that having the latter in the position of the dominant good in a society is even more problematic than the former occupying the same position. This is because money could still be kept out of citizens' private lives where personal ties and emotions have traditionally been more important than a person's financial situation.

2.2.3.1 Loss of the Private Sphere

The extent of the problems the dominant position of information and knowledge that the developments of AI have resulted in becomes clear when one considers the importance Walzer places on the separation of a person's private and public life (e.g., Walzer, 1984, p. 317). As I have found in Chapter 1, AI has infiltrated private spaces, extracted data from them, and thus publicised private life.

Yet, for Walzer, it is in the private sphere where individuals can enjoy their "individual and familial freedom, privacy and domesticity" (Walzer, 1984, p. 317). It is "[o]ur homes [that we treat as] our castles, and there we are free from official surveillance" (Walzer, 1984, p. 317) which AI ought not to interfere with. As such, privacy is a good "we greatly value" (Walzer, 1984, p. 317).

Because AI, in its current use, does not identify situations in which no data should be collected i.e. in homes or other rather intimate environments or situations (Zuboff, 2019, pp. 128-130), I argue that privacy is being disregarded and one of the central assumptions of Walzer's theory, that of the public and the private being separated.

[15] Walzer himself "call[s] a good dominant if the individuals who have it, because they have it, can command a wide range of other goods" (Walzer, 1983, p. 10).

It becomes problematic considering that he goes as far as to equate privacy with freedom (Walzer, 1984, p. 317). Given that freedom[16] is central to Walzer's theory of justice, I claim that blurring the lines between the public and the private challenges his conceptual framework in various ways.

For his theory of justice, Walzer regarded the mentioned distinction between the public and the private as given. Going off of this assumption, he emphasises not only the importance of communal autonomy and self-determination (e.g., Walzer, 1983, pp. 6-10) but also citizens' participation in the political process (Walzer, 1983, pp. 306-309).

Nonetheless, no solution is provided for how to solve a lack of freedom other than by mutual assistance from another community (Stassen, 1994, p. 388). Given that AI is dominating all social spheres of most if not all existing communities, assistance might not be enough and potentially not even possible if citizens are not aware of being unfree and can thus not provide the needed assistance to others either.

Thus, the surveillance which AI has enables goes not only against Walzer's idea of self-determination and the evolution of communal values over time but also against the concept of complex equality as such. It is in private that we form our (social) identities. In public life, however, we are to separate between the different parts of our identity in their respective spheres.

Regardless, for human life, an individual must be able to have multiple affiliations to live a dignified life (Sen, 2002, pp. 42-43). Because Walzer gives precedence to the separation of social spheres over the importance of privacy[17], I conclude that if one applied his theory to an environment shaped by the current AI developments, a contradiction would occur.

If a private sphere is to exist but a separation of spheres is more important than the former, the private sphere would inevitably disappear. However, Walzer argues that it is the private sphere that "creates the sphere of individual and familial freedom" (Walzer, 1984, p. 317). This freedom would subsequently have to be used in a way that the social meanings of the larger, public community can prevail. For instance, citizens must be free enough to be passive in some social processes (e.g., Walzer, 1983, p. 308) which is an option not given if

[16] While in *Spheres of Justice* (Walzer, 1983) different accounts of freedom are addressed i.e. communal freedoms to determine their own social meanings ascribed to social goods as well as individual freedoms, the following part shall focus on that of individual freedom.

[17] I conclude this from the chronological order of the publications of his works. While the separation of social spheres is mentioned extensively in his 1983 work *Spheres of Justice*, the importance of privacy is only heavily emphasised in his 1984 publication entitled *Liberalism and the Art of Separation.*

the private sphere is not to be separated from the public one. Therefore, Walzer's theory of justice presents us with a choice between which crucial building block of his theory to save in light of AI developments which is not enough since for the rest of his conceptual framework, both of those aspects must be given.

2.2.3.2 Representative Political Systems

Another central assumption in Walzer's theory is that of a representative political system (Walzer, 1983, p. 303). He argues that in a just society, everyone has to abide by rules the authority i.e. the state makes but in return can influence them to the same degree as every other member of that community (Walzer, 1983, p. 61).

While in the writing of his work, the author himself was concerned more with (im-) migration, membership in a community and traditional political representation of members (Walzer, 1983, Chapters 2, 12), the developments of AI give rise to this issue on a larger scale, hence a necessity to reexamine representative political systems.

While AI is not greatly threatening political representation as such, it is, however, threatening the second assumption of Walzer's ideal political system i.e. that of all citizens "ultimately [having] an equal say"(Walzer, 1983, p. 61).

As has been found in Chapter 1, AI and other big data methodologies have the potential to analyse citizens' actions and subsequently classify them in a certain way. Through nudging and, in some cases, outright manipulation, individual citizens are provided with different opportunities for all sorts of things. When this is put together with some sort of a social credit system, this presents us with a highly unequal society in which each individual is ascribed a number as their 'social worth' (Xu et al., 2022, p. 2241). It is only one step further until individuals are provided with an unequal number of votes based on their *social score*, especially when propositions such as *Quadratic Voting*[18] become a reality.

If this is actualised, abstractly, one could still see how every citizen would theoretically have the same say since everyone is born with the same 'score' which is only later changing. Nevertheless, I argue that that through the changing of a social score, personalisation of all sorts of offers, as well as nudges, individuals might be perceived to be punished for their way of leading a life rather than a single act itself which goes directly against Walzer's account of what makes a punishment just (Walzer, 1983, p. 268).

[18] An explanation of Quadratic Voting can be found in Špecián (2022, pp. 143-150).

2.2.3.3 Distribution of Offices

Furthermore, if a jury is already informed about all of the individuals' other acts, it is hard to separate which act happened in which social sphere, so it is only possible to punish someone for a specific act but with information in the back of one's mind about the individual's other actions. This information obtained through surveillance enabled by AI is thus likely to result in undifferentiated judgment (Ignor, 2012, pp. 231-232).

Yet, in his theory, Walzer did not accommodate irrational judges and deemed undifferentiated judgements unjust in themselves (Walzer, 1983, p. 268). Yet, in the case of AI-supported judgements, the biases instilled in those technologies will result in an undifferentiated judgment regarding the spheres but also entail more severe punishments for non-dominant social groups (Rafanelli, 2022, pp. 1, 5).

This is especially problematic considering that through the collection of data and the reorganisation of society around it, the distribution of offices through competition (Walzer, 1983, p. 132) is made unrealistic. Nevertheless, the distribution of offices through competition would constitute simple equality and thus be undesirable anyway. Still, taking into account e.g. the possible individualisation of school curricula (Filgueiras, 2023, p. 6) the playing field for competition for qualifications relevant to a given office is severely tilted and influenced by other spheres of a citizen's life so that there is no "fair equality of opportunity" (Walzer, 1983, p. 135).

Regardless, Walzer himself advocates the distribution of offices according to citizens' qualities which AI might be able to measure and thus evaluate, the biases instilled in AI judgement challenge the corrective function of state institutions. In them, officeholders, ought to represent society, especially in courts. It is precisely this representation that AI is threatening because Walzer himself reserves the use of quotas and the reservation of certain offices for a specific social group to bi-national or highly pluralistic societies (Walzer, 1983, p. 149). While AI is inarguably creating more pluralistic societies[19], in Walzer's conceptual framework of justice, they would not classify as such due to their shared history and thus make the reservation of some offices an unjust practice (Walzer, 1983, p. 149).

Furthermore, he argues that the Selection Committee's decision would not be the sole criterion relevant to an individual's chance to office. This would be because individuals

[19] in terms of shared social meanings which emerge in the epistemic bubbles whose emergence AI enables

are assumed to make free decisions prior to application the consequences of which would weigh heavily on their qualifications (Walzer, 1983, p. 145). There are two problems with this: firstly, there is the issue of who is on the Committee (e.g., Zuboff, 2019, p. 180) and secondly, there is the problem that AI developments are interfering with individuals' freedom of choice, as I have outlined in the previous chapter.

Therefore, I argue that because the distribution of office, which is supposed to be representative of a given society and correct existing injustices (Walzer, 1983, pp. 268-270), cannot be regarded as a separate social sphere anymore and must thus be considered an unjust practice in itself that Walzer's conceptual framework of justices provides no solution for. This essentially eradicated either the prevalence of equality of opportunity or that of equal representation so that a hierarchically organised society is bound to emerge. Both options are undesirable.

3. Comparison of the Limitations in the Assessed Theories' Framework

As I have uncovered in Chapter 2, both J. Rawls' as well as M. Walzer's theories of justice are conceptually flawed in light of the developments of AI. Nevertheless, there are certain differences in how their conceptual frameworks fail to accommodate the threats AI poses to their real-world applicability.

Therefore, depending on the perspective taken on social justice, some scholars have argued that Rawls' theory would be better suited to the given contextual situation (e.g., Ferretti, 2022) while others have argued that Walzer's would be the best to do so (e.g., Santoni De Sio et al., 2021). Yet, in the following section, I shall argue that neither of the theories is suitable to be applied in light of the developments of AI.

3.1 Breakdown of Communities

By comparing the conceptual limitations of both theories in light of AI developments, it stands out that the breakdown of boundaries of traditional communities is affecting both theories, although differently.

For Rawls' theory, the problem lies in the emergence of common sympathies to persons outside one's own community which entail moral duties to them. Subsequently, there exist moral duties to persons who only limited institutions are shared with.

Yet, the satisfaction of moral duties can only take place within an institutional framework. Nevertheless, the establishment of institutions outside of one's own political culture contradicts Rawls' own principle of non-interference and is therefore an act of injustice in itself.

For Walzer's theory, on the other hand, the problem lies more in declining levels of identification with the larger political community. This is because through the strengthening of sub-groups, the bonds that have traditionally kept communities together are weakening. Therefore, citizens' trust in institutions which wall off inequalities in one sphere with another sphere can be expected to be low.

It follows that applying Walzer's theory of *justice as complex equality* to an AI-driven environment is likely to result in low levels of public trust which in turn, reinforces the phenomenon of an increasingly fragmented society consisting of various competing groups rather than one unified community, which is part of the basic assumptions of Walzer's theory.

3.2 Treatment of Strangers

Additionally, both assessed theories fail to provide a sufficient framework how to justly treat strangers i.e. non-members of a given community.

For Rawls' theory, the problem is to be found more in the international order, while for Walzer's the problem lies in the treatment of an immediate person. In contrast to Walzer, Rawls' theory fails to protect individuals in the way that his theory is based on the assumption of a global liberal hegemony. The current global trend amplified by AI developments, points to the decline thereof, though, so there is no remaining effective mechanism that would protect basic human rights in Rawls' theory. However, it is those rights whose importance for justice to prevail Rawls stresses.

In comparison to Walzer, this problem is rather huge and more encompassing. This is because, in Walzer's framework, I have merely found that the treatment of refugees solely according to their status is being made more difficult by AI-decision making and the large-scale availability of personal data as well as surveillance.

3.3 Theory Specific Issues

3.3.1 Justice as Fairness

For justice as fairness, it is perhaps most important to state that AI is weakening existing social institutions. Nevertheless, it is precisely those institutions that the theory is aimed to address by formulating a set of principles those institutions ought to protect. Yet, even the first step of Rawls' procedural theory of justice is challenged. Because of the analytic capacities AI technologies possess, individuals are exposed to different information and nudged into different directions early on, so that an abstraction from the self in the way Rawls imagined it to happen is impossible.

Furthermore, decision-making competencies are increasingly being transferred onto AI agents who have no capacity for moral learning and thus reflective abilities. Therefore, I argue that both Rawls' point of departure i.e. the *original position* as well as his methodology of *reflective equilibrium* are not applicable in the context of AI developments.

Finally, Rawls' argument on the distribution of primary goods in accordance with how much an individual is contributing to a community, is unlikely to be realised in a fair manner considering AI developments. This is because, as I have found in Chapter 1, AI is challenging the freedom of choice, equality of opportunity as well balance of power within a society so that it is predetermined how much an individual can contribute, regardless of their willingness to hard work or sacrifice.

3.3.2 Justice as Complex Equality

For Walzer on the other hand, the main problem lays in the merging of what is public and what is private which AI enables. Inequalities are acceptable in his conceptual framework not only because they are walled off in another sphere but also because "[o]ur homes are our castles" (Walzer, 1984, p. 317) in which we are supposed to be protected from external interference of all sorts.

Since data is becoming a dominant good which severely influences the distribution of information among various members of society and therefore has the potential to create new imbalances of power in several ways. In doing so, the dominance of data goes as far as to penetrate into even the private sphere, resulting in an essential abolishment thereof.

Furthermore, on the political level as well as in the sphere of office, equal representation is being challenged because of inequality of opportunity in the qualification

process as well as the dominant group's possible infringement on individual choice. It follows that even applications for offices may be hindered and the dominance of certain groups in offices be kept. Subsequently, it becomes harder to correct existing injustices.

Conclusion

The research question of this thesis was what the conceptual limitations of J. Rawls' and M. Walzer's in the light of the developments of AI are and whether it will be enough to amend those theories or if they will have to be replaced by new ones.

My general hypothesis was that the current developments of AI are challenging the relevance of both J. Rawls' as well as M. Walzer's theories in their very assumptions. Therefore, I further hypothesised that it would not be enough to solely amend those theories in the light of the developments of AI but that entirely new theories of justice might have to be invented. This hypothesis has been proven right.

Even though, M. Walzer's theory of *justice as complex equality* seems to be slightly better conceptually equipped to accommodate the challenges AI poses to social justice, it is still challenged in its foundational assumptions so that, in case amendments are made in those regards, the rest of the theory might collapse. Furthermore, I have shown that even if one of the basic social goods whose importance Walzer stresses is to be protected, the protection of another good would contradict this enterprise.

I, therefore, conclude that in light of the developments of AI, both J. Rawls' and M. Walzer's theories are severely conceptually limited so new theories of justice must be developed.

Summary

The research question of this thesis was the following: *What are the conceptual limitations of J. Rawls' and M. Walzer's theories of justice in light of the developments of AI and will it be enough to amend those theories, or do they need to be replaced by other, new theories of justice?* The two theories were chosen due to their incredible influence on the discourse on social justice as well as the argument in the literature that either of them would be best equipped to be applied considering AI developments[20].

In order to find an answer to the given research question, I started by identifying

[20] see e.g. Ferretti (2022) and Santoni De Sio et al. (2021)

those social goods whose just distribution is challenged in the light of AI developments. To do so, I relied on existing literature on real-world applications of AI and applied methods from analytical political philosophy to show how they would create or amplify social injustices. I found that the most threatened social goods are privacy, the prevalence of reason, a just distribution of power, equality, (human) freedom and the freedom of choice as well as the breakdown of traditional (political) communities and interpersonal relations to immediate others.

In Chapter 2, I used those findings to shed light on how the assessed theories fail to protect the just distribution of those goods. In doing so, I adhered to Rawls' and Walzer's own methodologies. However, both have been found to be inapplicable in an AI-driven environment.

About Rawls' theory, I found that the global application of his theory is challenged since he assumes a liberal hegemony which is threatened by AI. Furthermore, the realm in which he argues that institutions ought to work does not coincide with the emergence of new forms of common sympathies that AI facilitates. This results in a contradiction between his principle of non-interference and the development of personal ties which entail moral duties to others. Additionally, his own method of reflective equilibrium is challenged by the loss of rationality and the inability of AI to reflect on its actions. Moreover, the distribution of social goods in accordance with the contribution principles is unfair due to inequalities of opportunity caused by the developments of AI.

Concerning Walzer's theory, on the other hand, I uncovered that his basic frame of analysis in which justice can be analysed, the traditional political community, is challenged. Additionally, the separation of the public and the private sphere is endangered by knowledge and information becoming a dominant good. Since Walzer himself does not account for a separate sphere in which a dominance of information would be acceptable so one can only assume it to dominate all spheres and the private one. On top of that, his conceptual framework leaves room for political systems to be classified as just which he did not intend to and his account of how offices ought to be distributed in a society contradicts his other arguments if one considers how AI is challenging the equality of opportunity to obtain qualifications.

The analysis of both theories' conceptual limitations in light of AI developments was followed by a comparison of them. This part showed that although Walzer's framework is slightly better equipped to meet the challenges AI is posing to social justice, it is insufficient

to ensure it.

Therefore, in light of the developments of AI, new theories must be developed.

Limitations

It is important to stress that one of the main assumptions of my analysis was that AI is operating in an unregulated way. This is because of its global application so that those who develop it faster have a comparative (economic) advantage over others. If AI were to be regulated globally, it might be possible that both assessed theories would still be applicable in the given framework.

Besides this, J. Rawls' and M. Walzer's theories were chosen because of their fundamentally different approach to social justice. Nonetheless, there are disagreements in the literature on whether theories like Rawls' are better suited to be applied in light of AI (e.g., Ferretti, 2022) or whether it is Walzer's (e.g., Santoni De Sio et al., 2021). Even so, not all currently received theories of justice have been assessed so the solution to their conceptual limitations which were identified in this thesis might have already been published.

Závěr

Výzkumnou otázkou této práce bylo, jaká jsou koncepční omezení teorií J. Rawlse a M. Walzera ve světle vývoje umělé inteligence a zda bude stačit tyto teorie pozměnit, nebo zda budou muset být nahrazeny novými.

Moje obecná hypotéza byla, že současný vývoj umělé inteligence zpochybňuje relevanci teorií J. Rawlse i M. Walzera v jejich samotných předpokladech. Proto jsem dále vyslovil hypotézu, že nebude stačit pouze pozměnit tyto teorie ve světle vývoje umělé inteligence, ale že bude možná nutné vymyslet zcela nové teorie spravedlnosti. Tato hypotéza se ukázala jako správná.

I když se zdá, že teorie spravedlnosti jako komplexní rovnosti M. Walzera je konceptuálně o něco lépe vybavena k tomu, aby se vyrovnala s výzvami, které AI představuje pro sociální spravedlnost, stále je zpochybněna ve svých základních předpokladech, takže v případě, že by v těchto ohledech byly provedeny změny, zbytek teorie by se mohl zhroutit. Navíc jsem ukázal, že i když je třeba chránit jeden ze základních společenských statků, jehož důležitost Walzer zdůrazňuje, ochrana jiného statku by tomuto podniku odporovala.

Dospěl jsem proto k závěru, že ve světle vývoje UI jsou teorie J. Rawlse i M. Walzera značně konceptuálně omezené, takže je třeba vytvořit nové teorie spravedlnosti.

Souhrn

Výzkumná otázka této práce zněla: Rawlse a M. Walzera ve světle vývoje umělé inteligence a bude stačit tyto teorie pozměnit, nebo je třeba je nahradit jinými, novými teoriemi spravedlnosti? Tyto dvě teorie byly vybrány vzhledem k jejich neuvěřitelnému vlivu na diskurz o sociální spravedlnosti a také vzhledem k argumentaci v literatuře, že některá z nich by byla nejvhodnější pro aplikaci s ohledem na vývoj UI .

Abych našel odpověď na danou výzkumnou otázku, začal jsem identifikací těch sociálních statků, jejichž spravedlivé rozdělování je ve světle vývoje UI zpochybněno. Za tímto účelem jsem se opíral o existující literaturu o reálných aplikacích UI a aplikoval jsem metody z analytické politické filosofie, abych ukázal, jak by vytvářely nebo zesilovaly sociální nespravedlnost. Zjistil jsem, že nejvíce ohroženými sociálními statky jsou soukromí, převaha rozumu, spravedlivé rozdělení moci, rovnost, (lidská) svoboda a svoboda volby, jakož i rozpad tradičních (politických) společenství a mezilidských vztahů k nejbližším druhým.

V kapitole 2 jsem na základě těchto zjištění osvětlil, jak posuzované teorie selhávají při ochraně spravedlivého rozdělení těchto statků. Přitom jsem se držel Rawlsovy a Walzerovy vlastní metodologie. Obě však byly shledány jako nepoužitelné v prostředí řízeném umělou inteligencí.

Pokud jde o Rawlsovu teorii, zjistil jsem, že globální aplikace jeho teorie je zpochybněna, protože předpokládá liberální hegemonii, která je ohrožena umělou inteligencí. Navíc oblast, v níž by podle něj měly instituce fungovat, se neshoduje se vznikem nových forem společných sympatií, které AI umožňuje. To má za následek rozpor mezi jeho zásadou nevměšování a rozvojem osobních vazeb, z nichž vyplývají morální povinnosti vůči druhým. Navíc je jeho vlastní metoda reflexivní rovnováhy zpochybněna ztrátou racionality a neschopností UI reflektovat své jednání. Rozdělování společenských statků podle principů příspěvku je navíc nespravedlivé kvůli nerovnosti příležitostí způsobené rozvojem UI.

Co se týče Walzerovy teorie, zjistil jsem, že jeho základní rámec analýzy, v němž lze spravedlnost analyzovat, tradiční politické společenství, je zpochybněn. Navíc je ohroženo oddělení veřejné a soukromé sféry tím, že se dominantním statkem stává vědění a informace. Jelikož sám Walzer nepočítá s oddělenou sférou, v níž by byla přijatelná dominance informací, tak lze pouze předpokládat, že dominují všem sférám i té soukromé. Navíc jeho pojmový rámec ponechává prostor pro klasifikaci politických systémů jako spravedlivých, což neměl v úmyslu, a jeho popis toho, jak by měly být úřady ve společnosti rozdělovány, je v rozporu s jeho dalšími argumenty, pokud uvážíme, jak UI zpochybňuje rovnost příležitostí k získání kvalifikace.

Po analýze koncepčních omezení obou teorií ve světle vývoje umělé inteligence následovalo jejich srovnání. Tato část ukázala, že ačkoli je Walzerův rámec o něco lépe vybaven k řešení výzev, které AI představuje pro sociální spravedlnost, k jejímu zajištění nestačí.

Proto jsem dospěl k závěru, že ve světle vývoje UI je třeba vytvořit nové teorie.

Omezení

Je důležité zdůraznit, že jedním z hlavních předpokladů mé analýzy bylo, že umělá inteligence funguje neregulovaně. Důvodem je její globální uplatnění, takže ti, kdo ji vyvíjejí rychleji, mají komparativní (ekonomickou) výhodu oproti ostatním. Pokud by UI byla regulována globálně, bylo by možné, že by obě posuzované teorie byly v daném rámci stále

použitelné.

Kromě toho byly teorie J. Rawlse a M. Walzera vybrány kvůli jejich zásadně odlišnému přístupu k sociální spravedlnosti. Nicméně v literatuře se objevují neshody ohledně toho, zda jsou pro aplikaci ve světle UI vhodnější teorie jako Rawlsova (např. Ferretti, 2022), nebo zda je to Walzerova (např. Santoni De Sio et al., 2021). I tak ale nebyly posouzeny všechny v současnosti přijímané teorie spravedlnosti, takže řešení jejich koncepčních omezení, která byla identifikována v této práci, již mohla být publikována.

List of References

Adamson, F. B. (2005). Globalisation, Transnational Political Mobilisation, and Networks of Violence. *Cambridge Review of International Affairs*, *18*(1), 31–49. https://doi.org/10.1080/09557570500059548

Adelsberger-Hoss, E. (2021). *ETHISCHE UND MORALISCHE ASPEKTE DER BEGABTENFORDERUNG UNTER BESONDERER BEZUGNAHME AUF DAS BERUFSBILDENDE VOLLZEITSCHULWESEN*. INNSBRUCK UNIVERSITY PRESS.

Al Samman, A. M., & Mohamed, A. (2024). Artificial Intelligence, Organizational Justice and Organizational Trust: Towards a Conceptual Framework. *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, 17–22. https://doi.org/10.1109/ICETSIS61505.2024.10459667

Anderson, E. (1995). *Value in ethics and economics* (2. print). Harvard Univ. Press.

Backer, L. C. (2019). China's Social Credit System: Data-Driven Governance for a 'New Era'. *Current History*, *118*(809), 209–214. https://doi.org/10.1525/curh.2019.118.809.209

Barriga, A. D. C., Martins, A. F., Simões, M. J., & Faustino, D. (2020). The COVID-19 pandemic: Yet another catalyst for governmental mass surveillance? *Social Sciences & Humanities Open*, *2*(1), 100096. https://doi.org/10.1016/j.ssaho.2020.100096

Bauer, J. (2007). *Prinzip Menschlichkeit: Warum wir von Natur aus kooperieren* (5. Aufl). Hoffmann und Campe.

Brożek, B., & Janik, B. (2019). Can artificial intelligences be moral agents? *New Ideas in Psychology, 54*, 101-106. https://doi.org/10.1016/j.newideapsych.2018.12.002

Bundesregierung. (2024, May 22). *AI Act verabschiedet Einheitliche Regeln für Künstliche Intelligenz in der EU*. https://www.bundesregierung.de/breg-de/themen/digitalisierung/kuenstliche-intelligenz/ai-act-2285944. Retrieved July 24, 2024.

Button, M. E. (2018). Bounded Rationality without Bounded Democracy: Nudges, Democratic Citizenship, and Pathways for Building Civic Capacity. *Perspectives on Politics*, *16*(4), 1034–1052. https://doi.org/10.1017/S1537592718002086

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

De Oliveira, L. F., Da Silva Gomes, A., Enes, Y., Castelo Branco, T. V., Pires, R. P., Bolzon, A., & Demo, G. (2022). Path and future of artificial intelligence in the field of justice: A systematic literature review and a research agenda. *SN Social Sciences*, *2*(9), 180. https://doi.org/10.1007/s43545-022-00482-w

European Commission. (2018, December 7). *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Coordinated Plan on Artificial Intelligence.* https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52018DC0795#:~:text=The%20Commission%20pro. Retrieved April 17,2024.

European Commission. (2021, April 21). *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206. Retrieved July 17, 2024.

Ferretti, T. (2022). An Institutionalist Approach to AI Ethics: Justifying the Priority of Government Regulation over Self-Regulation. *Moral Philosophy and Politics*, *9*(2), 239–265. https://doi.org/10.1515/mopp-2020-0056

Filgueiras, F. (2023). Artificial intelligence and education governance. *Education, Citizenship and Social Justice*, 174619792311606. https://doi.org/10.1177/17461979231160674

Freedom House. (2023). *FREEDOM IN THE WORLD 2023*. https://freedomhouse.org/sites/default/files/202303/FIW_World_2023_DigtalPDF.pdf. Retrieved June 28, 2024.

Han, B.-C. (2017). *The agony of eros*. MIT Press.

Hawkins, D., Lake, D. A., Nielson, D. L., & Tierney, M. J. (2006). Delegation under anarchy: States, international organizations, and principal-agent theory. In D. G. Hawkins, D. A. Lake, D. L. Nielson, & M. J. Tierney (Eds.), *Delegation and Agency in International Organizations* (1st ed., pp. 3–38). Cambridge University Press. https://doi.org/10.1017/CBO9780511491368.002

Hellmeier, S. (2016). The Dictator's Digital Toolkit: Explaining Variation in Internet

Filtering in Authoritarian Regimes. *Politics & Policy*, *44*(6), 1158–1191. https://doi.org/10.1111/polp.12189

Hermansyah, M., Najib, A., Farida, A., Sacipto, R., & Rintyarna, B. S. (2023). Artificial Intelligence and Ethics: Building an Artificial Intelligence System that Ensures Privacy and Social Justice. *International Journal of Science and Society*, *5*(1), 154–168. https://doi.org/10.54783/ijsoc.v5i1.644

Heymann, J. (2014, November 20). *Children's right to education: Where does the world stand?* https://www.right-to-education.org/fr/blog/children-s-right-education-where-does-world-stand. Retrieved July 22, 2024.

Ignor, A. (2012). *Befangenheit im Prozess*. 228–237.

Jahn, D. (2007). Was ist Vergleichende Politikwissenschaft? Standpunkte und Kontroversen. *Zeitschrift für Vergleichende Politikwissenschaft*, *1*(1), 9–27. https://doi.org/10.1007/s12286-007-0001-y

Johnston, D. (2011). The Theory of Justice as Fairness. In *A Brief History of Justice* (1st ed.). Wiley. https://doi.org/10.1002/9781444397550

Kak, A. (2020). 'The Global South is everywhere, but also always somewhere': National Policy Narratives and AI Justice. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 307–312. https://doi.org/10.1145/3375627.3375859

Kemp, S. (2023, January 26). *DIGITAL 2023 DEEP-DIVE: UNDERSTANDING THE DECLINE IN TIME SPENT ONLINE*. Datareportal. https://datareportal.com/reports/digital-2023-deep-dive-time-spent-online. Retrieved July 8, 2024.

Krupiy, T. (Tanya). (2020). A vulnerability analysis: Theorising the impact of artificial intelligence decision-making processes on individuals, society and human diversity from a social justice perspective. *Computer Law & Security Review*, *38*, 105429. https://doi.org/10.1016/j.clsr.2020.105429

Laitinen, A., & Sahlgren, O. (2021). AI Systems and Respect for Human Autonomy. *Frontiers in Artificial Intelligence*, *4*, 705164. https://doi.org/10.3389/frai.2021.705164

McDermott, D. (2008). Analytical political philosophy. In D. Leopold & M. Stears (Eds.), *Political theory: Methods and approaches* (pp. 11–21). Oxford University Press.

Mergel, I., Dickinson, H., Stenvall, J., & Gasco, M. (2023). Implementing AI in the public sector. *Public Management Review*, 1–14.

https://doi.org/10.1080/14719037.2023.2231950

Moradi, P., & Levy, K. (2020). The Future of Work in the Age of AI: Displacement or Risk-Shifting? In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI* (pp. 269–288). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.17

Mounk, Y. (2023). *The identity trap: A story of ideas and power in our time*. Penguin Press.

Narayanan, D., Nagpal, M., McGuire, J., Schweitzer, S., & De Cremer, D. (2024). Fairness Perceptions of Artificial Intelligence: A Review and Path Forward. *International Journal of Human–Computer Interaction*, *40*(1), 4–23. https://doi.org/10.1080/10447318.2023.2210890

Néron, P. (2016). Rethinking the Ethics of Corporate Political Activities in a Post-Citizens United Era: Political Equality, Corporate Citizenship, and Market Failures. *Journal of Business Ethics*, 136(4), 715-728. https://doi.org/10.1007/s10551-015-2867-y

Nussbaum, M. C. (2002, November). *Beyond the Social Contract: Capabilities and Global Justice*. THE TANNER LECTURES ON HUMAN VALUES. https://tannerlectures.utah.edu/_resources/documents/a-to-z/n/nussbaum_2003.pdf. Retrieved June 4, 2024.

Rafanelli, L. M. (2022). Justice, injustice, and artificial intelligence: Lessons from political theory and philosophy. *Big Data & Society*, *9*(1), 205395172210806. https://doi.org/10.1177/20539517221080676

Rawls, J. (1958). *Justice as Fairness*. *67*(2), 164–194.

Rawls, J. (1985). *Justice as Fairness: Political not Metaphysical*. *14*(3), 223–251.

Rawls, J. (1999a). *A theory of justice* (Rev. ed). Belknap Press of Harvard University Press.

Rawls, J. (1999b). *The law of peoples: With, The idea of public reason revisited*. Harvard University Press.

Sánchez Muñoz, Ó. (2002). Review of: JAHRBUCH DES ÖFFENTLICHEN RECHTS, núm. 50, «Die Macht der Medien in der Gewaltenteilung» by Walter Schmitt Glaeser. *Centro de Estudios Políticos y Constitucionales*, 66, 277-282.

Santoni De Sio, F., Almeida, T., & Van Den Hoven, J. (2021). The future of work: Freedom, justice and capital in the age of artificial intelligence. *Critical Review of International Social and Political Philosophy*, 1–25. https://doi.org/10.1080/13698230.2021.2008204

Schneider, G., & Toyka-Seid, C. (2024). *Vierte Gewalt*. Das junge Politik-Lexikon.

https://www.bpb.de/kurz-knapp/lexika/das-junge-politik-lexikon/321342/vierte-gewalt/. Retrieved July 22, 2024.

Sen, A. (2002). Justice across Borders. In P. De Greiff & C. P. Cronin (Eds.), *Global Justice and Transnational Politics* (pp. 37–52). The MIT Press. https://doi.org/10.7551/mitpress/3302.003.0003

Sloane, M. (2019). Inequality Is the Name of the Game: Thoughts on the Emerging Field of Technology, Ethics and Social Justice. *Weizenbaum Conference*. https://doi.org/10.34669/WI.CP/2.9

Špecián, P. (2022). *Behavioral political economy and democratic theory: fortifying democracy for the digital age*. Routledge.

Stassen, G. (1994). *Michael Walzer's Situated Justice*. *22*(2), 375–399.

Stehr, N., & Adolf, M. (2010). Consumption between Market and Morals: A Socio-cultural Consideration of Moralized Markets. *European Journal of Social Theory*, *13*(2), 213–228. https://doi.org/10.1177/1368431010362287

Ulnicane, I. (2022). Artificial intelligence in the European Union. In T. Hoerber, G. Weber, & I. Cabras, *The Routledge Handbook of European Integrations* (1st ed., pp. 254–269). Routledge. https://doi.org/10.4324/9780429262081-19

Van Wyk, M. W. (2008). *Equal opportunity and liberal equality* [DPhil, University of Johannesburg]. https://hdl.handle.net/10210/1354. Retrieved May 25, 2024.

Varnholt, H. (2023, June 15). KÜNSTLICHE INTELLIGENZ So reguliert sich Europa in die Abhängigkeit. *WirtschaftsWoche*. https://www.wiwo.de/unternehmen/it/kuenstliche-intelligenz-so-reguliert-sich-europa-in-die-abhaengigkeit/29208524.html. Retrieved July 15, 2024.

Vavoula, N. (2021). Artificial Intelligence (AI) at Schengen Borders: Automated Processing, Algorithmic Profiling and Facial Recognition in the Era of Techno-Solutionism. *European Journal of Migration and Law*, *23*(4), 457–484. https://doi.org/10.1163/15718166-12340114

Wagner, B. (2019). Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping? In E. Bayamlioglu, I. Baraliuc, L. A. W. Janssens, & M. Hildebrandt (Eds.), *BEING PROFILED* (pp. 84–89). Amsterdam University Press. https://doi.org/10.1515/9789048550180-016

Walzer, M. (1983). *Spheres of justice: A defense of pluralism and equality* (Nachdr.). Basic Books.

Walzer, M. (1984). *Liberalism and the Art of Separation*. *12*(3), 315–330.

Walzer, M. (1985). *Interpretation and Social Criticism*. THE TANNER LECTURES ON HUMAN VALUES, Harvard University. https://tannerlectures.utah.edu/_resources/documents/a-to-z/w/walzer88.pdf. Retrieved June 23, 2024.

Xu, X., Kostka, G., & Cao, X. (2022). Information Control and Public Support for Social Credit Systems in China. *The Journal of Politics*, *84*(4), 2230–2245. https://doi.org/10.1086/718358

Yalcin, G., Themeli, E., Stamhuis, E., Philipsen, S., & Puntoni, S. (2023). Perceptions of Justice By Algorithms. *Artificial Intelligence and Law*, *31*(2), 269–292. https://doi.org/10.1007/s10506-022-09312-z

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power* (First edition). PublicAffairs.

Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, *38*(3), 101577. https://doi.org/10.1016/j.giq.2021.101577