

Chapter Title: Introduction: Why We Need Critical Perspectives on AI

Chapter Author(s): Pieter Verdegem

Book Title: AI for Everyone?

Book Subtitle: Critical Perspectives

Book Editor(s): Pieter Verdegem

Published by: University of Westminster Press. (2021)

Stable URL: <https://www.jstor.org/stable/j.ctv26qjjhj.3>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



This book is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0). To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>.



University of Westminster Press is collaborating with JSTOR to digitize, preserve and extend access to *AI for Everyone?*

CHAPTER I

Introduction: Why We Need Critical Perspectives on AI

Pieter Verdegem

Introduction

The renewed interest in Artificial Intelligence (AI) has made it the most recent hype in the world of technological innovation. In the business world, AI is seen as a catalyst for growth, which will manifestly transform the economy and the future of work (Agrawal, Gans and Goldfarb 2018; Lee 2018; McAfee and Brynjolfsson 2017). Policymakers and civil society are putting their hopes on AI for tackling global challenges such as pandemics and even climate change (Dobbe and Whittaker 2019; Dananjayan and Raj 2020). AI also seems to be the subject of an arms race between China, Russia and the USA for equipping their armies with automated weaponry (Asaro 2018).

Whenever we are confronted with a hype, it is of utmost importance to untangle what exactly is at stake and who is behind the discourses and myths created. We are being told stories about AI as the ultimate innovation, transforming the ways we live and work – often started in corporate circles and distributed by their supportive popular outlets. At the same time, however, analysis is revealing that AI itself is one reason behind intensifying societal problems and harms. Researchers and thinkers have observed and/or predicted that AI leads to discrimination (Zuiderveen Borgesius 2018), is the engine behind growing inequalities (Korinek and Stiglitz 2017), can bring about technological unemployment (Ford 2015) and may even contribute to the end of humanity (Bostrom 2014).

How to cite this book chapter:

Verdegem, P. 2021. Introduction: Why We Need Critical Perspectives on AI.

In: Verdegem, P. (ed.) *AI for Everyone? Critical Perspectives*. Pp. 1–18. London: University of Westminster Press. DOI: <https://doi.org/10.16997/book55.a>. License: CC-BY-NC-ND 4.0

Amidst this doom and gloom, what we desperately need is a more nuanced debate about AI's risks and opportunities. This can – must – be a serious and informed discussion that goes beyond hyperbole and polarisation, fuelled by popular media and thus feeding into public debate. What we need is critical perspectives on AI: what it is and what it is not; what type of AI we need, what visions exist about this and who is behind them; and ultimately, how to think and talk about AI power and inequalities.

In one word, it is *power* that must be at the centre of our conversations about AI and that is what this book is about. If we want to talk about critical perspectives on AI, formulating a critique on AI, how it is currently being developed and discussed, and yes, if we are serious about making sure that AI will benefit everyone, we need to talk about power. Power refers to the capacity to influence the actions, beliefs or behaviour of others. Ultimately, this comes down to 'the question of who can influence what society looks like and who controls the means that allow such influence' (Fuchs 2017: 86). Power decides who will benefit from new technologies such as AI, but a concentration of power will likely result in growing inequalities and other negative outcomes. The current critiques about AI centre on AI ethics (Coeckelbergh 2020), which is valuable and important to shape policy discussions. AI ethics, however, also has serious limitations when it comes to bringing about real change and making sure that the benefits of AI are accessible for everyone. Further in this introduction, I elaborate on this and I make the case for a radical democratisation of AI, and why we need to put power at the centre for achieving this.

The contributions in this book braid discussion of power and critique with three strands: AI – Humans vs. Machines, Discourses and Myths About AI and AI Power and Inequalities.

Part 1: *AI – Humans vs. Machines* – deals with the history and conceptualisation of AI and what is at stake in its development. This section looks at different perspectives about what characterises machine intelligence and how it might be important to further radical humanism in the era of automation and AI.

Part 2: *Discourses and Myths About AI* – analyses how AI is framed in popular and scholarly discussions and investigates the normative projections of what AI should be and what it should do. This section poses critical questions about how AI needs to debunk the myths surrounding it.

Part 3: *AI Power and Inequalities* – advances the debate around AI by critically examining what 'AI for Everyone?' means. This is dealing with the root of the problem: who will benefit from AI is ultimately down to who has the power to decide. These contributions look at how AI capitalism is organised, what (new) inequalities it might bring about and how we can fight back.

Why do we need a book on AI for Everyone? and why do we need it now? The 2007–2008 financial crisis, and the resulting global economic crisis, has not only brought about a decade of austerity in large parts of the Western world; it has also been the context in which social media and digital platforms have transformed into behemoths. Tech companies are now dominating the top 10

of the most valuable companies in the world (Verdegem 2021). Austerity has also led to growing inequalities and political polarisation, bringing right-wing authoritarian politics into power in a number of countries (Fuchs 2018). A world already cracked by economic uncertainty and the looming threat of climate change was then shaken in 2020 by a global pandemic. COVID-19 has massively impacted the global economy, on a much larger scale than the 2007–2008 crisis. On top of this, the pandemic has also resulted in an even bigger dependence and dominance of tech platforms such as Amazon, Alibaba, Google and Tencent. These companies are, not surprisingly, also leading AI companies. Only a small number of corporations have the necessary computational power to develop AI systems, are financially strong enough to hire the brightest AI talent and have access to the gigantic datasets that are needed to train machine learning and deep learning (AI) models. This context makes it very clear why we need to ask critical questions about AI and power.

Conceptualising AI – What AI Are We Talking About?

Before understanding what type of AI we *want*, we need to understand what AI we *have*. This is an area of significant debate, and the book opens by exploring the varying approaches to how we define AI.

The Origins of AI

It is easy to forget that AI has been with us for more than 60 years. Despite the flash of excitement and anxiety that feels so recent, AI itself is not a new phenomenon. The name *Artificial Intelligence* (AI) was coined in the mid-1950s at a series of academic workshops organised at Dartmouth College, New Hampshire (USA). A group of scientists, led by mathematics professor John McCarthy, gathered to investigate the ways in which machines could simulate aspects of human intelligence: the ability to learn and make decisions. Their core assumption was that human reasoning could be reconstructed using mathematical techniques and, as a consequence, problem-solving could be formalised into algorithms (McCarthy et al. 1955/2006).

What is more recent is a reflexive, if not critical, and social-scientific, understanding of not just AI's capabilities, but its impacts on human life and social organisation (Elliott 2019). It took decades for AI research to move from what it could do *for* us to what it could do *to* us, or enable us to do to each other. These first critical insights came along with observations that AI can not only supercharge innovation and bring about economic prosperity but also lead to inequalities and unfairness.

This book contributes to this debate by critically reflecting on how we should think about AI and the relationship between humans and machines. It analyses the discourses and myths that exist around AI; what it will enable

and what not. And it looks at issues about AI, power and inequalities, investigating where the risks of exclusion are and how we should deal with this.

The book also brings diverse and critical voices to this debate. Whereas AI as a discipline has been dominated by white, male, predominantly older scientists from mathematical disciplines, this collection brings perspectives that are characterised by a strong diversity in authorship and discipline. And threading through all, the contributions offer a discussion of different tangents of power and political economy in the field of AI and society.

The first task is to name our terms. For a concept that has been with us for so long, there is little consensus on how to define it. The history of debating AI is almost as old as AI itself. There is more debate than agreement about what AI is and what it is not, and the only thing generally agreed is that there is no widely accepted definition (Russell and Norvig 2016). The first definition comes from that gathering of scientists in 1955: McCarthy et al. (1955/2006) then defined AI as: ‘Making a machine behave in ways that would be called intelligent if a human were so behaving.’ This only raises the challenge of how exactly to define *intelligence*. Russell and Norvig (2016: 2) define different approaches to AI to serve different goals. AI can refer to systems that: (1) think like humans; (2) think rationally; (3) act like humans; and (4) act rationally. Each of the approaches requires different disciplinary expertise, thus requiring an inter-, or at least cross-disciplinary discussion. The human-centred approaches will depart from social science studying human behaviour, while the rationalist approaches will involve a combination of mathematics and engineering. From the four approaches, *acting like humans* is closest to how we define and understand contemporary AI.

We can see the roots of *acting like humans* in the *Turing test*, developed by Alan Turing in 1950. This test, originally designed to provide a satisfactory definition of intelligence, has been central to conceptualising AI. According to the test, if a human interrogator cannot distinguish a machine from a human through conversation, then the machine can be considered *intelligent*. Russell and Norvig (2016) argue that for a computer to be intelligent – to pass the Turing test – it needs to possess the following capabilities: *natural language processing* (being able to communicate successfully), *knowledge representation* (being able to store what it knows or hears), *automated reasoning* (being able to use the stored information to answer questions and to draw new conclusions) and *machine learning* (being able to adapt to new circumstances and to detect and extrapolate patterns).

Towards an Operational Definition – For Now

It is helpful to first distinguish between *strong* and *weak* AI (Bostrom 2014). Strong AI, also called *AGI* (Artificial General Intelligence) refers to computational systems with *general* cognitive abilities which have the future

potential to surpass human intellectual capacities. This can be seen as the attempt to mechanise human-level intelligence. Computer scientists and philosophers disagree on whether this is at all possible (Coeckelbergh 2020): some directly reject this scenario while others think if theoretically possible, it is not likely to happen (soon) in practice (Boden 2016). This is why it might be better to focus on advancements in weak AI or *ANI* (Artificial Narrow Intelligence), as this is the type of AI already impacting everyday life on a massive scale. Weak/narrow AI performs *specific* tasks which would normally require intelligence in a human being – machines aiding human thought and action. This type of AI is a mathematical method for prediction (Agrawal et al. 2018). Such systems can be extremely powerful but are limited in the range of tasks they can perform.

Russell and Norvig (2016) see *machine learning* as a prerequisite for intelligent machines. Machine learning is a paradigm that allows programs to automatically improve their performance on a particular task by learning from vast amounts of data (Alpaydin 2016). It seeks and uses statistical patterns and correlation in enormous datasets. Unlike older types of AI (e.g. expert systems, that are based on rules which are inputted by humans), machine learning algorithms learn not from humans but from data. The availability of significant amounts of real-world data (that we produce by using the internet, social media, sensors or other Internet-of-Things applications), combined with the availability of powerful and almost limitless computing capacity and advancements in machine learning and deep learning is why we are currently in another period of AI optimism and hype (Elish and boyd 2018).

Given the concepts and the brief discussion above, how can we agree on an operational definition of AI? A basic definition would be to refer to AI as computer programming that learns from and adapts to data. A more elaborate version of this, as Elliott (2019: 4) puts it, defines AI as ‘any computational system that can sense its relevant context and react intelligently to data’ in order to perform highly complex tasks effectively and to achieve specific goals, thereby mimicking intelligent human behaviours. The discussion about how to define AI cannot be settled in one definition, let alone one book. It is an important starting point, however, and Part 1 and Part 2 of this book will unpack several approaches to defining AI.

The Realities of AI for Some vs. the Ideals of AI for Everyone

Visions of AI in Policies and Ethics

Examining AI policies and ethics helps us to explore questions of what type of AI we want/need, how its development should look like and how we deal with its impact. Policy development happens at several levels and includes a number of stakeholders: national governments, intergovernmental organisations, corporations, professional associations and academics.

While AI policies reflect the priorities of the stakeholders involved, ethical guidelines project a vision of what type of AI is preferred, what benefits it should deliver and how we should deal with potential risks. Obviously, this is part of a normative debate but we can learn a lot from who is involved in these discussions and how they aim to shape the future of AI.

Given the projections about the role of AI in economic development, AI is high on the policy agenda. Putin famously said that the nation that leads in AI would be the ruler of the world (Vincent 2017). Major nations are rushing to create AI initiatives, unsurprisingly led by China and the USA (Lee 2018). What is surprising, however, is how much overlap there is in their strategic vision.

China's national strategy for AI, the *New Generation Artificial Intelligence Development Plan* (NGAIDP), was released in 2017 (State Council of China 2017). China wants to become the world leader in AI by 2030 and has formulated strategic goals to achieve this, such as making China the superpower of fundamental and applied AI research and development in order to dominate the global AI market. The main focus of China's AI policy is on economic development and competition, even though it also discusses some concerns in terms of economic security and social stability.

The Trump administration launched the *American AI Initiative* in 2019 (White House 2019). This strategic policy is all about a nationalist vision of American leadership in AI. The US government wants to invest in AI R&D, set AI standards and build the AI workforce. The Trump AI strategy not surprisingly has an intense national focus, highlighting AI for American innovation, industry, workers and values, aimed at promoting and protecting national AI technology and innovation. There is some discussion of public trust and confidence in AI as well as the protection of civil liberties, privacy and American values but this is subordinated to leadership and protecting American AI technology. With the election of Biden, it remains to be seen what the shift in AI policies will be, but given his track record the US will continue to pursue US capitalist interests, although maybe in a less outspoken nationalist way.

Most European nations where we see AI policy development, including France, Germany and the UK, are taking a different approach, which more explicitly offers a normative vision of how AI should contribute to social progress. France, for example, has entitled its vision *AI for Humanity* and aims for the development of an ethical framework for transparent and fair use of AI applications (Villani 2018). Germany also wants to guarantee responsible development and deployment of AI which serves the good of society. The UK sits somewhere between the continental European visions and the US vision, with goals contributing to global AI development, tempered with nationalist objectives focusing on specific benefits for the UK.

It is clear that China and the US are in an intense battle for global AI leadership and their policies are dominated by nationalist goals. European countries want to engage in AI innovation and boost their competitiveness while also

ensuring that the societal impact of AI is not forgotten. But still, this does not tell us a lot about what type of AI we want/need; it rather explains what countries expect AI to do for them. The European Union, however, has done more to develop a vision of what type of AI needs to be pursued and what aspects need to be dealt with in this.

The EU situates itself between China (state capitalism) and the US (market capitalism) and seeks to shift the debate in terms of the impact on society and its citizens. This positioning is aligned with how they have approached General Data Protection Regulation (GDPR) in the context of data protection and privacy. The EU has put forward *trustworthy AI* as the key term highlighting what type of AI it likes to see developed. This concept is the result of an open consultation and its ethics guidelines have been presented by the *High-Level Expert Group of AI*. According to these guidelines, trustworthy AI should be: (1) lawful (respecting all applicable laws and regulations); (2) ethical (respect ethical principles and values); and (3) robust (both from a technical perspective while taking into account its social environment) (European Commission 2019). These aspects are vague (*how* is something ethical or robust exactly?) as well as self-evident (very few people would favour *unlawful* AI). The EU, however, has made these guidelines more explicit by formulating specific aims: human agency/oversight, technical robustness/safety, privacy/data governance, transparency, diversity/non-discrimination/fairness, societal/environmental well-being and accountability. This is helpful as the list of specific aims can be read as values we would like to attribute to AI.

Not only governments or governmental organisations are active in putting forward a vision for AI. Companies also have a stake in this debate so it is instructive to examine how leading tech companies talk about what type of AI they want to build. Google (2020) has developed a vision it calls *Advancing AI for Everyone*, which can be summed up as applying AI to improve their products and developing tools to ensure that everyone can access AI. Google also has an *AI for Social Good* project, similar to Microsoft's *AI for Good* program. The latter aims to use AI expertise to solve humanitarian and environmental challenges: AI for earth, health, accessibility, humanitarian action and cultural heritage. While seemingly well-intended at first glance, these AI programs are carefully designed to support goals of *corporate social responsibility* (Sandoval 2014) and are undeniably textbook examples of what Morozov (2013) has called *techno-solutionism*. The problem with these corporate AI visions is that they lack substance and therefore do not reveal anything about what they intend specifically and how they actually can and should benefit society.

More substance can be found in how professional associations propose a vision of what *Good AI* exactly means. Organisations such as the Institute of Electrical and Electronics Engineers (IEEE) and Association for Computing Machinery (ACM) have produced codes that propose ethical principles for computer science in general and AI in particular. ACM (2020), for example,

talks about AI that needs to ‘contribute to society and to human well-being’, while IEEE (2020) has come up with principles for *ethically aligned design*. General principles include human rights, well-being, data agency, effectiveness, transparency, accountability, awareness of misuse and competence.

Often cited are the *Asilomar AI Principles*. The *Asilomar Conference on Beneficial AI* was organised by the Future of Life Institute (2017) and brought together more than 100 AI researchers from academia and industry and thought leaders in economics, law, ethics and philosophy to address and formulate principles of *beneficial AI*. The resulting Asilomar AI principles are organised around (1) research issues, (2) ethics and values and (3) longer-term issues (Future of Life Institute 2017). The first category, *research issues*, sets out some guidelines in terms of research goals, funding and culture. Secondly, thirteen specific *ethics and values* are listed, dealing with transparency, safety, privacy, etc. and they also address aspects such as *shared benefit* (‘AI technologies should benefit and empower as many people as possible’) and *shared prosperity* (‘the economic prosperity created by AI should be shared broadly, to benefit all of humanity’). Last, under *longer-term issues*, cautionary aspects and risks are addressed, including the principle of *common good*, which states: ‘Superintelligence [Artificial General Intelligence, as discussed above] should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organisation.’

While the Asilomar AI principles are valid, they leave unclear who can and should take ownership and what mechanisms can be developed to enforce them. One specific concern of the Asilomar AI initiative is the heavy involvement of corporate stakeholders, given that it is backed by tech giants including Google, Facebook and Apple. These are not non-profit organisations but companies that are among the most wealthy and profitable organisations in the world. They might say they want to develop AI applications that are beneficial for society but can we trust them not to use their power to shift the direction of AI development to their corporate benefit and the return on investment for their investors and shareholders?

The *AI4People* initiative, set up by the non-profit organisation Atomium-EISMD (European Institute for Science, Media and Democracy), is the European response to the Asilomar AI initiative. AI4People also brings together academics, business partners (e.g. Facebook, Intel and Microsoft), and civil society organisations. The ambition of AI4People (Atomium-EISMD, 2020) is ‘to draft a set of ethical guidelines aimed at facilitating the design of policies favourable to the development of a “Good AI Society”’.

AI4People has developed an ethical framework of principles that should underpin the adoption of AI and offers a list of specific recommendations and action points that should help to establish a *Good AI Society* (Floridi et al. 2018). AI4People proposes five core ethical principles: (1) *beneficence* (promoting well-being, preserving dignity and sustaining the planet); (2) *non-*

maleficence (privacy, security and capability caution); (3) *autonomy* (the power to decide/whether to decide); (4) *justice* (promoting prosperity and preserving solidarity); and (5) *explicability* (enabling the other principles through intelligibility and accountability). While the first four overlap with traditional bioethics principles, the last one is ‘a new enabling principle for AI’ (Floridi et al. 2018, 700). There might be overlap with the Asilomar AI principles, but AI4People has come up with a comprehensive list of ethical principles, recommendations and action points that can help policymakers to develop and support AI projects and initiatives that benefit society. However, they are not without gaps and flaws.

What is Missing in AI Policies and Ethics: Introducing Capitalism

The overview of AI policies and initiatives aimed at formulating AI ethics, helps us understand the debate about what AI we want/need and what it should deliver (or what should be avoided). However, something crucial is missing: power. This brings us to the crux of the book and the possibilities of critical analysis of AI. To bring power into the debate, we must first understand two points: (1) the problem of AI ideology and (2) the limitations of ethics.

Let me start with *AI ideology*. National policies clearly illustrate that AI is seen as an important instrument for positioning countries in terms of what type of future society they aim to develop. Here comes the role of ideology. While a contested notion, ideology can refer to: ‘worldviews and ideas on the one end, to the process of the production of false consciousness on the other end of the spectrum’ (Fuchs 2020, 180). In other words, it can have a neutral meaning but ideology can also be used to manipulate human consciousness. In the latter meaning, ideology is seen as a typical characteristic of capitalism and class societies, and it is being used to serve the material interests of the ruling class (Fuchs 2020). As discussions of AI often include visions about its potential to radically alter societies and economies, we need to be alert to and critical towards AI ideology.

Berman (1992) wrote almost three decades ago that the growing interest in AI in capitalist societies can be understood not only in terms of its practical achievements but also in the ideological role it plays as a technological paradigm for the continuation and reinvention of capitalism. AI as an ideology means that it can be seen as: ‘a potential hegemonic principle within the sphere of formal organizations which facilitates the “fit” of human beings into the revised structures of a capitalism based on micro-electronic and information technology, and ideologically contains, and significantly mutes, resistance and social conflict’ (Berman 1992, 104). The technological paradigm is thus a major component of hegemonic ideology that helps to maintain the essential structures of the current capitalist system and makes coherent and viable

alternatives increasingly difficult to envision. AI ideology thus propagates one specific vision of what AI is and what it should do – including serving the interests of the ruling class – and discourages alternative visions from materialising.

Second, we need to be aware of the limitations of AI ethics. Computer ethics, the broader field to which AI ethics belongs, is a philosophical field of study that deals with the question of ‘how computer technology should be used’ (Moor 1985, 266). It investigates social impact but also how policies for ethical use of computer technology can be formulated and justified. This is important and is why I discuss not only ethical guidelines but also AI policies. AI ethics are important as they let us think about what a *good society* constitutes, how we – as members of that society – can live a meaningful and fulfilling life, and especially what the role of technology, in general, and AI, in particular, in this is (Coeckelbergh 2020). There are, however, problems and limitations with AI ethics and how they get linked to policy.

When it comes to developing AI ethical guidelines, the first question to ask is: who is involved? The issue of diversity and inclusion plays out on multiple levels. Research by Jobin, Ienca and Vayena (2019). (2019) demonstrates that developing AI ethics is concentrated in North America, the European Union, Japan and a small handful of other countries. The absence or underrepresentation of countries from Africa, Central and South America and Central Asia means that large global regions are not invited to contribute to this debate, illustrative of a geopolitical power imbalance. There are also questions about who exactly is involved in developing the guidelines and whether the panels of experts who produce ethical guidelines, are – or are not – representative of society. This undermines the plurality that AI ethics aim for.

Another problem of establishing AI ethics is the speed at which technologies are developing (Boddington 2017). Formulating ethical guidelines takes time and there is a question of whether or not ethics can keep up with the rapid development of technologies. AI policies, just as any policies, face a similar challenge and as a consequence they are often reactive rather than proactive. AI ethical guidelines also face the problem of *ethics washing* (Wagner 2018). This refers to the practice of exaggerating a company’s interest in promoting beneficial AI systems (Google’s *Advancing AI for Everyone* (2020) and Microsoft’s *AI for Good* (2020) programs, cfr. supra, are often used as examples for this) but also when ethics is used as a substitute for regulation, meaning that companies highlight how ethically they are acting while simultaneously abandoning their legal obligations (for example, not respecting principles in terms of data protection).

The vulnerability of ethics advocates and researchers is illustrated by the case of Timnit Gebru. Gebru is well-known for her work on racial bias in technology, such as facial recognition, and has criticised systems that fail to recognise black faces. She was fired by Google in December 2020 after sending an internal email that accused Google of silencing marginalised voices (Hao 2020).

It is clear that we need to be aware of AI ideology and acknowledge that AI ethics alone, despite their value and contributions, will not save the world. The other problem is about how we move from AI ethics to concrete policies. There is no roadmap for what exactly should be done, no precise course of action to be taken in policy development (Coeckelbergh 2020). It comes down to who has the capacity to influence the actions, beliefs or behaviour of others. Or who can influence what type of society we want, and what the role of technology such as AI should be in it. Ultimately, this is a question about power and who is in control to make decisions.

We Have to Talk about AI and Power

The problem of AI ideology and – more broadly – the question of whether we need AI and if so, what type of AI we need, illustrates why we need critical perspectives on AI. What do I mean by *critical*? The Frankfurt School has been pivotal in the development of critical thinking and theory. According to Max Horkheimer (2002), one of the leading figures of the Frankfurt School, critical theory distinguishes itself from traditional theory because of its focus on human emancipation. The goal of critical theory is to scrutinise and understand systems of domination and oppression and to look for ways of how to increase liberation and freedom.

If we make human emancipation central, we need to ask questions about AI and power. And this is exactly what is missing in AI policies and ethics: power. *Power* is a contested concept in social theory. In a pragmatic way, Wright (2010, 111) defines it as: ‘the capacity of actors to accomplish things in the world.’ This is a positive take on power, whereas a lot of definitions of power are negative – coercive power, preventing others to act in a certain way (Fuchs 2017). In addition to coercive power, Thompson (1995) also talks about economic power, political power and symbolic power. Economic power refers to how certain individuals and groups in society can accumulate resources for productive activity; political power is about the authority to coordinate individuals and their interaction; and symbolic power refers to meaning making and influencing the actions of others. AI ideology has raised issues of *symbolic* power, so I now turn to economic and political power in the context of AI.

We need to be aware that AI simultaneously refers to technical approaches, social practices and industrial infrastructures (Crawford 2018). The *technical approaches* are straightforward: these are computational systems that use data for training machine learning and deep learning algorithms (Alpaydin 2016). The other two elements need more clarification. The *social practices* of AI refer to the classification systems, developed by humans, which are behind the machine/deep learning algorithms and models. Political power asks who is involved in developing these classification systems and who decides what they

look like (Crawford 2018). Questions about inclusion and representation are inherently political questions. AI also refers to *industrial infrastructures*: the infrastructure does not only entail the possibilities of collecting vast amounts of data, but also the computational power needed to develop machine/deep learning models. Very few companies have simultaneously the computational power, access to data and AI expertise (human resources) at their disposal, which means that the economic power of these organisations is crucial for the development of AI and is highly concentrated (Dyer-Witthof, Kjosen and Steinhoff 2019).

The Case for a Radical Democratisation of AI

Asking critical questions about AI with the objective to foster human emancipation requires us to investigate the political and economic power dynamics of AI. My point here is that we need to move beyond discussions of what beneficial AI means and what opportunities and risks exist in its development. We urgently need to think instead about what radical approaches to AI are and how we can enable them. Why *radical*, and what does that mean? *Radical* originates in the Latin word *radix*, which means *root* and that is why it has been popularised as *grasping things at the root*. Radical can mean many things but here I refer to it as in *radical politics* (Fenton 2016). Radical politics is characterised by its intention of transforming the fundamental principles of a political system or a society, often by making use of structural change or radical reform – change at the root.

A radical perspective to AI thus means we need to examine AI through the lens of power. Ultimately this comes down to the question of how AI is shifting power. This is about bringing real change for the better, disrupting power dynamics and avoiding an unequal power distribution. We could repeat (and slightly revise) William Gibson's (2003) seminal quote 'AI is already here; it's just not evenly distributed'. The question then remains: how can we redistribute power in AI?

My proposal is that if we want to establish AI that transforms society for the better and enables human emancipation, we need a radical democratisation of AI. This radical democratisation is necessary to avoid power inequalities, in other words, to avoid a situation whereby only a few organisations, whether governmental or corporate, have the economic and political power to decide what type of AI will be developed and what purposes it will serve.

This is vital in the data and AI sector, which is characterised by a strong tendency to establish monopolies. Network effects intensify competition between data platforms: the more users on their platform, the more valuable they become (Srnicsek 2017). More data then also generates more users, which allows for the creation of better services. This is called a *data-feedback loop*. Data giants will therefore acquire competitors, which leads to a situation of

an oligopoly or even monopoly. This is even more crucial in the AI industry where few companies have access to data to train machine/deep learning algorithms, possess the computing power to deal with massive data sets and also to hire the AI talent that is necessary to build AI systems and applications (Dyer-Witthford et al. 2019).

So, what does a radical democratisation of AI actually mean? First, AI and the benefits it offers, should be *accessible to everyone*. Second, AI and the different services that are being developed should also *represent everyone*. Third and last, AI should be *beneficial to everyone*. These three principles are inspired by the late Erik Olin Wright's critique of capitalism. Wright (2019) proposes the principles of equality/fairness, democracy/freedom and community/solidarity as normative foundations for establishing a society that allows its members to live a decent life. In the following paragraphs, I briefly unpack these three guiding principles.

#Principle 1: AI Should Be Accessible to Everyone

This first principle proposes equal access to AI and the benefits it can offer. In a decent society, all persons should have broadly equal access to the advantages and possibilities being created by digital technologies such as AI. This means that we need to make sure that all groups in society have access to and can use AI. The egalitarian ideal is at the centre of nearly all concepts of social justice, including *data justice* (Taylor 2017), although there are different opinions about what it means exactly. An important nuance here is to distinguish between equal *access* and *opportunity*. The former is chosen over the latter as it 'is a sociologically more appropriate way of understanding the egalitarian ideal' (Wright 2019, 11). Given the current economic, social and environmental crisis we are living in, there should be particular attention to intergenerational and environmental justice. The first aspect points to the consequences of technological developments for the future generations, whereas the second aspect asks for attention for IT and sustainability. This is controversial as AI is both seen as a source of and solution for environmental degradation (Dauvergne 2020).

#Principle 2: AI Should Represent Everyone

The second principle is centred around democracy and inclusion. In a decent society, all members should have a say about what type of AI is being developed and what services are being offered. The production and implementation of AI must be democratised so that all groups in society are consulted and represented, avoiding exclusion. This element of democracy entails two aspects: everyone is involved and everyone is represented. The latter aspect highlights that when fairness fails, there is a risk of discrimination (Hoffmann 2019). The

history of AI is full of examples of how technology is being developed by (predominantly) white middle-class men, thereby excluding people of colour and minority communities. Wright (2019) also connects democracy and freedom in order to reflect the value of self-determination. In this sense, members of society should be given the possibility to participate meaningfully in decisions that affect their lives. As AI becomes more omnipresent, people should have a say about this. Principles such as fairness, accountability and transparency (ACM 2020) are key when we want technological development not only to represent the people but also guaranteeing control by the people to counterbalance the power of the state and corporations.

#Principle 3: AI Should Be Beneficial to Everyone

The third and last principle states that developments in AI should contribute to the well-being of everyone in society. This matches with Wright's (2019) ideas of community and solidarity, which are crucial because of their connection to human flourishing and of their role in fostering equality and democracy (see also principles 1 and 2). Central is the idea that if people cooperate, they can achieve more than if they compete, and cooperation also contributes to the well-being of all members of society. This means that AI development must be organised in such a way that all members of society are able to reap the benefits. Another aspect of this principle is the question about how to develop beneficial machines, in other words, how can we ensure that AI serves the objectives of humanity. Stuart Russell (2019, 11) states: 'machines are beneficial to the extent that their actions can be expected to achieve our objectives.' According to him, this is at the centre of the problem of control in AI and his interpretation focuses on the human-machine relationship, as part of being beneficial to everyone. Developing AI that is beneficial for everyone, thus includes thinking about how to create beneficial machines that serve humanity.

AI for everyone risks becoming yet another hype, if we let the tech giants take over the debate with their slogans such as *AI for social good*. What they are missing is a real vision of democratising technology because they fail to understand what *AI for everyone* really means: putting the human at the centre (Pasquale 2020). In one word, this is about power. If we are not talking about power, we are not talking about *AI for everyone*. Critical perspectives require us to talk about the human and society. By bringing together diverse critical contributions to the debate, this book presents one thing they have in common: the idea of putting society first.

Chapter Overview

Part 1: *AI – Humans vs. Machines* consists of four contributions. Andreas Kaplan (Chapter 2) goes deeper into the history and definition of AI and

elaborates on how humans and machines have to coexist in the age of AI. Wolfgang Hofkirchner (Chapter 3) continues the discussion about humans versus machines by analysing what Digital Humanism exactly entails. He proposes dialectical models in order to overcome the human–machine dualism. Jenna Ng (Chapter 4) adds to this discussion by elaborating on the rationalisation of AI and what this means for creativity. Dan McQuillan (Chapter 5) has a different take on humanism and proposes how people’s councils for AI can serve solidarity and mutual aid in times of crisis.

Part 2: *Discourses and Myths About AI* is comprised of five chapters. Rainer Rehak (Chapter 6) stresses the importance but also limitations of metaphors when talking about AI and intelligent systems. Angela Daly, S. Kate Devitt and Monique Mann (Chapter 7) introduce and discuss their Good Data approach in order to overcome the limitations of AI ethics and governance. James Steinhoff (Chapter 8) critically analyses the social reconfiguration of AI and discusses the central questions about utility and feasibility. Benedetta Brevini (Chapter 9) analyses AI policies in Europe and unpacks some of the myths around AI that legitimate capitalism. Alkim Almila Akdag Salah (Chapter 10) reflects on how the discourses of artistic computational production have changed and how myths about AI need to be uncovered in this context.

Part 3: *AI Power and Inequalities* involves five contributions. Carrie O’Connell and Chad Van de Wiele (Chapter 11) revisit Wiener’s cybernetic prediction as the theoretical foundation of AI and make a plea how we need to uncover the black box of what is behind prediction and simulation. Jernej A. Prodnik (Chapter 12) critically analyses algorithmic logic in digital capitalism, its characteristics and social consequences. Asvatha Babu and Saif Shahin (Chapter 13) investigate biometrics and biopolitics and apply their analysis to a case study of the ban on facial recognition in California. Rafael Grohmann and Willian Fernandes Araújo (Chapter 14) turn to a discussion of human labour that is behind global AI platforms and report about their empirical research on the Mechanical Turk in Brazil. Last, Lina Dencik (Chapter 15) also reflects on the relationship between labour and AI and proposes the concept of data justice unionism to rethink the governance of AI.

References

- ACM. 2020. *ACM Code of Ethics and Professional Conduct*. Retrieved from: <https://www.acm.org/code-of-ethics>
- Agrawal, A., Gans, J. and Goldfarb, A. 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Boston, MA: Harvard Business School Publishing.
- Alpaydin, E. 2016. *Machine Learning*. Cambridge, MA: MIT Press.
- Asaro, P. 2018. What Is the Artificial Intelligence Arms Race Anyway? *I/S: A Journal of Law and Policy for the Information Society*, 15(1–2).

- Atomium-EISMD. 2020. AI4People, Europe's First Global Forum on AI Ethics, Launches at the European Parliament. Retrieved from: <https://www.eismd.eu/ai4people-europes-first-global-forum-ai-ethics-launches-at-the-european-parliament/>
- Berman, B.J. 1992. Artificial Intelligence and the Ideology of Capitalist Reconstruction. *AI & Society*, 6, 103–114.
- Boden, M. 2016. *AI: Its Nature and Future*. Oxford: Oxford University Press.
- Boddington, P. 2017. *Towards a Code of Ethics for Artificial Intelligence*. Cham: Springer.
- Bostrom, N. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Coeckelbergh, M. 2020. *AI Ethics*. Cambridge, MA and London: MIT Press.
- Crawford, K. 2018. *The Politics of AI. Royal Society, You and AI*. Retrieved from: <https://www.youtube.com/watch?v=HPopJb5aDyA>
- Dananjayan, S. and Raj, G.M. 2020. Artificial Intelligence During a Pandemic: The COVID-19 example. *The International Journal of Health Planning and Management*, 35(5), 1260–1262.
- Dauvergne, P. 2020. *AI in the Wild. Sustainability in the Age of Artificial Intelligence*. Cambridge, MA: MIT Press.
- Dobbe, R. and Whittaker, M. 2019. *AI and Climate Change: How They're Connected, and What We Can Do about It*. AI Now Institute. Retrieved from: <https://medium.com/@AINowInstitute/ai-and-climate-change-how-theyre-connected-and-what-we-can-do-about-it-6aa8d0f5b32c>
- Dyer-Witheford, N., Kjosen, A.M. and Steinhoff, J. 2019. *Inhuman Power: Artificial Intelligence and the Future of Capitalism*. London: Pluto Press.
- Elish, M.C. and boyd, d. 2018. Situating Methods in the Magic of Big Data and AI. *Communication Monographs*, 85(1), 57–80.
- Elliott, A. 2019. *The Culture of AI. Everyday Life and the Digital Revolution*. London: Routledge.
- European Commission. 2019. *Ethics Guidelines for Trustworthy AI*. Retrieved from: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
- Fenton, N. 2016. *Digital, Political, Radical*. Cambridge: Polity Press.
- Floridi, L., Cows, J., Beltrametti, M. et al. 2018. AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Mind and Machines*, 28, 689–707.
- Ford, M. 2015. *The Rise of the Robots*. London: Oneworld Publications.
- Fuchs, C. 2017. *Social Media: A Critical Introduction* (2nd Edition). London: SAGE.
- Fuchs, C. 2018. *Digital Demagogue: Authoritarian Capitalism in the Age of Trump and Twitter*. London: Pluto.
- Fuchs, C. 2020. *Marxism: Key Ideas in Media and Cultural Studies*. New York: Routledge.

- Future of Life Institute. 2017. *Principles Developed in Conjunction with the 2017 Asilomar Conference on Beneficial AI*. Retrieved from: <https://futureoflife.org/ai-principles>
- Gibson, W. 2003. The Future Is Already Here – It's Just Not Evenly Distributed. *The Economist*, 4 December 2003.
- Google. 2020. *Advancing AI for Everyone*. Retrieved from: <https://ai.google>
- Hao, K. 2020. *We Read the Paper that Forced Timnit Gebru Out of Google. Here's What it Says*. Retrieved from: <https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru>
- Hoffmann, A.L. 2019. Where Fairness Fails: Data, Algorithms, and the Limits of Antidiscrimination Discourse. *Information, Communication & Society*, 22(7), 900–915.
- Horkheimer, M. 2002. *Critical Theory*. New York: Continuum.
- IEEE. 2020. *IEEE Ethics in Action in Autonomous and Intelligent Systems*. Retrieved from: <https://ethicsinaction.ieee.org/#resources>
- Jobin, A., Ienca, M. and Vayena, E. 2019. The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence*, 1, 389–399.
- Korinek, A. and Stiglitz, J.E. 2017. *Artificial Intelligence and Its Implications for Income Distribution and Unemployment*. Working Paper 24174, National Bureau of Economic Research. Retrieved from: <https://www.nber.org/papers/w24174>
- Lee, K.-F. 2018. *AI Superpowers: China, Silicon Valley and the New World Order*. Boston, MA: Houghton Mifflin.
- McAfee, A. and Brynjolfsson, E. 2017. *Machine, Platform, Crowd: Harnessing the Digital Revolution*. New York: WW Norton & Company.
- McCarthy, J., Minsky, M., Rochester, N. and Shannon, C. 1955/2006. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine*, 27(4), Winter 2006.
- Microsoft. 2020. *AI for Good*. Retrieved from: <https://www.microsoft.com/en-us/ai/ai-for-good>
- Moor, J.H. 1985. What is Computer Ethics? *Metaphilosophy*, 16(4).
- Morozov, E. 2013. *To Save Everything, Click Here*. London: Penguin Books.
- Pasquale, F. 2020. *New Laws of Robotics. Defending Human Expertise in the Age of AI*. Cambridge, MA: Harvard University Press.
- Russell, S. 2019. *Human Incompatible. AI and the Problem of Control*. London: Allen Lane.
- Russell, S. and Norvig, P. 2016. *Artificial Intelligence: A Modern Approach*. Harlow: Pearson Education Limited.
- Sandoval, M. 2014. *From Corporate to Social Media. Critical Perspectives on Corporate Social Responsibility in Media and Communication Industries*. London and New York: Routledge.
- Srnicek, N. 2017. *Platform Capitalism*. Cambridge: Polity Press.

- State Council of China. 2017. *New Generation of Artificial Intelligence Development Plan* (translated by F. Sapio, W. Chen and A. Lo). Retrieved from: <https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf>
- Taylor, L. 2017. What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally. *Big Data & Society*, 4(2), 1–14.
- Thompson, J. 1995. *The Media and Modernity. A Social Theory of the Media*. Cambridge: Polity Press.
- Turing, A.M. 1950. Computing Machinery and Intelligence. *Mind*, LIX(236), 433–460.
- Verdegem, P. 2021. Social Media Industries and the Rise of the Platform. In P. McDonald (Ed.) *Routledge Companion to Media Industries*. New York: Routledge.
- Villani, C. 2018. *AI for Humanity. French Strategy for Artificial Intelligence*. Retrieved from: <https://www.aiforhumanity.fr/en/>
- Vincent, J. 2017. *Putin Says the Nation that Leads in AI 'Will be the Ruler of the World'*. Retrieved from: <https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world>
- Wagner, B. 2018. Ethics as an Escape from Regulation. From 'Ethics-Washing' to 'Ethics-Shopping'? In E. Bayamlioglu, I. Baraliuc, L.A.W. Janssens and M. Hildebrandt (Eds.) *Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen*, pp. 84–89. Amsterdam: Amsterdam University Press.
- White House. 2019. *Artificial Intelligence for the American People*. Retrieved from: <https://www.whitehouse.gov/ai>
- Wright, E.O. 2010. *Envisioning Real Utopias*. London and New York: Verso.
- Wright, E.O. 2019. *How to be an Anti-Capitalist in the 21st Century?* London: Verso.
- Zuiderveen Borgesius, F. 2018. *Discrimination, Artificial Intelligence, and Algorithmic Decision-Making*. Strasbourg: Council of Europe, Directorate General of Democracy. Retrieved from: <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>