CATÓLICA
INSTITUTO DE
ESTUDOS POLÍTICOS

LISBOA

The Future is Now:

Liberal Democracies and the Challenge of Artificial Intelligence

Master Thesis

Maria Campos de Carvalho Cortesão Monteiro

MA in Governance, Leadership and Democracy Studies

Supervisor: Professor William Hasselberger

Instituto de Estudos Políticos

Universidade Católica Portuguesa

Lisbon, September 2021

Acknowledgments

Even though this is the first page of my thesis, it was the last one to be written. And so, after reading, thinking and writing so much in English (which was part of the challenge and fun of all this), I would like to express my gratitude as I feel it: in Portuguese.

E assim:

Muitos agradecimentos são devidos, pois esta pequena tese foi fruto de um longo e inesperado - e inesperadamente longo - processo.

Ao meu orientador, Professor William Hasselberger, agradeço muito a persistência e insistência, bem como os novos caminhos que me ajudou a descobrir e traçar.

Ao Instituto de Estudos Políticos, enquanto casa e na pessoa do Professor João Carlos Espada bem como de todos os que dele fazem parte, por me ter recebido tanto e tão bem ao longo destes muitos anos. Anos de muita aprendizagem, muitas ideias e muito crescimento. Em especial, agradeço muito ao Professor Orlando, por todas as oportunidades, conselhos e confiança e (tanta, tanta) paciência.

Aos meus colegas de investigação no projecto Rising Authoritarian Powers, que pelo trabalho de grupo e vontade de fazer bem feito deram (até sem saber, talvez) um empurrão valente e necessário a este processo.

Agradeço a todos os que (inexplicavelmente) continuaram sempre a perguntar o que era feito e quando ficava feito, sem duvidar que assim seria e mesmo quando já nem eu tinha assim tanta certeza. Em particular agradeço muito à minha study buddy e companheira de sempre, Lili. Ao Hugo Chelo e à Francisca, ao Luís, à Tanica, ao Pedro, por uma vida tão mais bonita. É que sem vida não há tese nem há nada.

À minha família, por tudo e por tudo o resto, o que veio e virá. Avós e avô, tias e tios e primos e primas.

E por fim, mas sempre no início, aos meus pais e manas. Por aturarem a minha teimosia (não é fácil) e serem ainda mais teimosos que eu (ainda menos). Por cismarem que eu consigo, na tese e na vida no geral. Talvez isso faça parte de serem meus pais e minhas manas mas é precisamente por isso que (lhes) estou muito agradecida: por serem os meus pais e as minhas manas. Obrigada.

E pronto. Finalmente aqui está.

MA in Governance, Leadership and Democracy Studies

The Future is Now: Liberal Democracies and the Challenge of Artificial Intelligence

Maria Cortesão Monteiro

Abstract

Current technological developments, such as Artificial Intelligence, Big Data, and Machine Learning, have a significant impact on multiple – if not all – dimensions of our lives. They are deployed, used, and fed by people in a social, cultural, and political context. They can also be used to change and influence that same context, namely the governmental system, the social dimension, and individuals' rights and liberties in a political regime. This dissertation will entail analyze how these technologies are being deployed in different dimensions of liberal democratic societies This analysis will delve into the dynamic of Big Tech data collection, the challenges posed by algorithmic decision-making tools on democratic institutions and the profound and tacit impact of surveillance technology. It will focus on the major challenges these technological deployments may represent, separately and as a whole, on a human rights and ethical perspective as well as an institutional one. The dissertation aims to contribute to a better understanding of the impact of diverse uses of AI and policy options in the socio-political dimension, instigating awareness and reflection on the paths being built. This reassessment is particularly focused on the possibility of AI as a strengthener and enhancer of liberal democracies, their values, and institutional structures.

Keywords: *Democracy; Artificial Intelligence; Big Data; Technology; Machine Learning European Union*.

Word Count Dissertation: 19005
Word Count Abstract: 203

Table of Contents

Introduction 5

Part 1: Who A(m) I – Understanding the concepts 8

Part 2: Artificial Intelligence and Liberal Democracies – Present Challenges . . . 12

2.1 Black Mirror on the Wall: The Challenges posed by Social Media and Big Tech companies 11

2.2 The Road to Hell is Paved with Good Intentions: The Challenges posed by Algorithmic-Decision Making 23

2.3. Every step you take, every move you make: The Challenges posed by Surveillance Technology 28

2.4. Smile you're being watched: the Chinese Updated Dictatorship 32

Part 3: The Future is Now 38

Conclusion 56

The Future is Now: Liberal Democracies and the Challenge of Artificial Intelligence

Introduction

> Success in creating AI would be the biggest event in human history.
> Unfortunately, it might also be the last, unless we learn how to avoid the risks.
> Hawking et. al, "Transcending Complacency on Superintelligent Machines", 2014

Liberal democracies are going through a very interesting time. Interesting precisely in all the concept's ambiguity: challenging and dangerous, exciting and uncertain, intriguing and full of possibilities. Those exact same words are ideal to describe Artificial Intelligence, the main focus of the following dissertation. Artificial Intelligence is, indeed, an excellent illustration and practical case to understand the type of crossroad that democracies presently face.

Artificial Intelligence brings tremendous possibilities. Its impact range encompasses in-depth, political, military, economic, and social dimensions. Its opportunities are as varied as advances in medicine and health, education, science, transportation, communication, economic growth, institutional strengthening, environmental sustainability. However, as with every major technological innovation, AI brings with it many challenges and risks. From the suspension of fundamental human rights to the erosion of institutions; the disruption of social relations; the emergence of unaccountable oligopolistic power structures; the possibility of more effective authoritarianism; the sovereignty crises; international tensions: nothing is left untouched. AI poses complex challenges in almost every dimension of human life: ethical and psychological, of privacy and liberty, issues of transparency and accountability, in work and leisure, of skill and knowledge development, inequality and inclusiveness, law,

policy and regulation, governance, and democracy, warfare and safety. The magnitude of the risks is also very illustrative of the potential of AI.

Artificial Intelligence has a 'dual-use nature'[1]. All its technical applications might be, in practice, used for good or for ill[2]. AI is an indispensable and inevitable element in our present and future societies and the more it becomes so, the more it is essential to reflect and ponder how we use the power it might provide us. The fundamental debate now is, as the researcher Brian Christian argues, how to ensure that this technologies and new technological models capture and mirror our norms and values, and understand what we intend for them to do and what we want.[3] This is what the author calls "the alignment problem"[4].

This dissertation will analyze AI deployment in liberal democracies from three different societal layers: individual, institutional, statewide. As we will see, possibilities and risks in these different areas are intertwined, overlapped and interdependent. These different layers are a simplification for the sake of a clearer analysis, but in fact AI must be taken into account and dealt with from a holistic perspective. From the analysis of the present AI 'path', we'll approach the challenges for the future: what are liberal democracies doing and what should they focus on if they want to build a robust AI democratic framework for the future. The truth is that, even though Artificial Intelligence and machine learning technologies are "ostensibly technical field[s]"[5], they are now

---

[1] Darrell M. West and John R. Allen, *Turning Point: Policymaking in the Era of Artificial Intelligence* (Washington: Brookings Institution Press, 2020).

[2] The same technology, for example image recognition, can be used in a benefic way – assist and improve cancer diagnosis – or in a nefarious way – identify and track citizens.

[3] Brian Christian, *The Alignment Problem: Machine Learning and Human Values*, First edition (New York, NY: W.W. Norton & Company, 2020), 27.

[4] Christian, The Alignment Problem, 27.

[5] Christian, 28.

fundamentally present and embedded in human questions. “Our human, social, and civic dilemmas are becoming technical. And our technical dilemmas are becoming human, social, and civic.”[6]

Throughout this dissertation, there will be moments of apparent alarming techno-skepticism. It is essential to disclaim that skepticism does not come from a luddite[7], anti-technology perspective. On the contrary: it is precisely because we acknowledge and understand the power and potential of AI that we aim at contributing to a joint reflection about the current path AI is taking. This dissertation aims, most of all, to contribute to the development, deployment, and use of AI that is reliable, trustable, and an ally not of authoritarianism but of liberty and democracy.

10/22/21 11:52:00 AM[6] Christian, 28.

[7] “Luddite” is now a blanket term used to describe people who dislike new technology, but its origins date back to an early 19th-century labor movement that railed against the ways that mechanized manufactures and their unskilled laborers undermined the skilled craftsmen of the day”. *In* Evan Andrews, “Who Were the Luddites?,” HISTORY, June 26, 2019, https://www.history.com/news/who-were-the-luddites.

Part 1: Who A(m) I – Understanding the concepts

Firstly, before trying to understand the impact of its deployment, we should start by understanding what it is. What do we mean by Artificial Intelligence?

Artificial Intelligence is a difficult concept to grasp. It is, in fact, more of a field than one specific and well-defined "thing". It is more a type of system than a tangible object. As a field, it embraces various subfields, and its systematic nature is precisely one reason it has so many usages and dimensions. As such, it is harder for it to have a precise and consensual definition.

John McCarthy, one of AI's founding scholars and the first scientist to coin the term "artificial intelligence", in 1956, defined AI as "getting a computer to do things which, done by people, are said to involve intelligence."[8] Marvin Minsky, another of AI's foundational scientists, defined it "the science of making machines do things that would require intelligence if done by men"[9]. A recent Stanford University report defines AI as "a science and a set of computational technologies that are inspired by – but typically operate quite differently from – the ways people use their nervous systems and bodies to sense, learn, reason, and take action."[10]

In a document published in 2019, an independent expert panel convened by the European Commission[11] defines AI as systems that display intelligent behavior by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals. This intelligence is displayed and achieved through algorithms

[8] Lindsey Andersen, "Human Rights in the Age of Artificial Intelligence" (Access Now, November 2018), https://www.accessnow.org.

[9] Andersen.

[10] Andersen, 18.

[11] AI HLEG, "Ethics Guidelines for Trustworthy AI," ebook (Bruxelles: European Comission, April 8, 2019), https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

translated into computer code and embedded into a machine (computer, smartphone, etc.). An algorithm resembles a recipe, a series of instructions or guidelines that describe how to perform and accomplish a task, how to achieve the final specific goal. It is composed of a set of instructions, and from the data inputted, it provides outputs. An algorithm can be developed by a programmer, hand-made, or automatically generated from data.

This type of AI, which aims at fulfilling a specific task or goal, is what is currently in use. It is called 'narrow AI' or 'weak AI'. It is goal-oriented and simulates human behavior according to the parameters and data which are fed into it. It leads to a particular output based on the information available (input). It is applied to solve a specific problem, and is called 'narrow' or 'weak' because it does only what it is designed to do.

As the EU Expert Panel explains[12], AI researchers use mostly the notion of rationality. AI functions with rationality in the sense that it has the ability to choose the best action to take in order to achieve a specific goal, given certain criteria to be optimized and the available resources. By 'the best action' is meant the most efficient way of accomplishing the goal it was created for, according to the data it was fed and the environment in which the system is immersed. It does not involve creativity or wisdom, adaptability, or common sense.

Narrow AI comes in contrast with 'Strong AI', or 'General AI'. The latter exhibits "human" intelligence: it will allegedly be able to learn and apply its intelligence to any situation or problem, being, therefore, capable of performing and mimicking human Intelligence or behaviors. This type of AI is not a primary concern. If it ever arrives, it will raise a new range of very significant and vital questions. But until the time comes, let's leave that to Hollywood.

[12] AI HLEG.

One of the components of the AI's field is machine learning. A report of the European Parliament defines it as "an AI component that provides systems with the ability to automatically learn over time, generally from large quantities of data."[13] The learning process occurs because, from the data that it is fed, the algorithm can generate another algorithm – the model. For example, the Amazon recommendation algorithm uses customers' profiles to learn which products are likely to be of interest to them. The recommendation model can produce personalized recommendations from the data fed from your profile[14]. Machine learning is valuable due to its processing power: it can quickly find or highlight data patterns that would otherwise be missed. It learns to predict outputs based on previous examples of relationships between input data and outcomes[15]. Therefore, it contributes to enhanced problem solving on a wide range of areas and issues, from image analysis systems that help in medical diagnoses, fraud detection systems, or administrative procedures. The vast majority of Artificial Intelligence in the world today is powered by machine learning.

Another fundamental component of AI deployment in democratic societies are algorithmic decision-making systems, or algorithmic decision systems (ADS). ADS is a specific type of algorithm used to support decision-making. Castelluccia and Daniel Le Métayer use the expression 'system' to stress that these algorithms should be analyzed in a general setting that includes their parameters, context of deployment, and the use, or

---

[13] Claude Castelluccia and Daniel Le Métayer, "Understanding Algorithmic Decision-Making: Opportunities and Challenges" (European Parliamentary Research Service, March 2019), 4.

[14] Castelluccia and Le Métayer, 4.

[15] "Machine learning is sometimes considered a form of "self-programming," since it's primarily the data that determines the detailed form of the learned model." Michael Kearns and Aaron Roth, *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* (New York: Oxford University Press, 2020), 6.

not, of machine learning training data[16]. Their ability to process a significant amount of data supports decision-making and allows tailored approaches and applications.

AI is also a complex concept to grasp because it is more than a technology. Or, better said, it represents more than a technology. It raises a technical scenario that will fundamentally influence and shape the future of the political and social landscape. Politically and socially, AI must be approached holistically, both in its development as in its deployment. However, for the sake of clarity, in this dissertation we will subdivide our analysis of the impact of its deployment into three parts: individuals, institutions, society.

[16] Castelluccia and Le Métayer, "Understanding Algorithmic Decision-Making: Opportunities and Challenges," 1.

Part 2: Artificial Intelligence and Liberal Democracies – Present Challenges

2.1. Black Mirror on the Wall: The Challenges posed by Social Media and Big Tech companies

> Leading tech companies were respected and treated as emissaries of the future. Nothing in past experience prepared people for these new practices, and so there were few defensive barriers for protection. Individuals quickly came to depend upon the new information and communication tools as necessary resources in the increasingly stressful, competitive, and stratified struggle for effective life. The new tools, networks, apps, platforms, and media thus became requirements for social participation. Finally, the rapid buildup of institutionalized facts – data brokerage, data analytics, data mining, professional specializations, unimaginable cash flows, powerful network effects, state collaboration, hyperscale material assets, and unprecedented concentrations of information power – produced an overwhelming sense of inevitability.
>
> S. Zuboff, "Big Other: Surveillance capitalism and the prospects of an information civilization", 2015, (p. 85)

Little more than a decade ago, the Internet was promising a new era of transparency, in which open access to information would, many predicted, result in extraordinary liberty[17]. The world would be more connected and more open, and freedom would undoubtedly prevail over any dictatorship[18].

According to the media researcher Taina Bucher[19], life increasingly takes place in and through an algorithmic media landscape. In this landscape, people interact with specific media companies and platforms. As we will see, platforms interact with people

---

[17] Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge: Harvard University Press, 2015), 14.

[18] For instance, "[o]n March 8, 2000, U.S. President Bill Clinton hailed the arrival of a new era, one in which the internet would mean the triumph of liberty around the world. He dismissed China's fledgling efforts to restrain online speech. "Good luck," quipped Clinton. "That's sort of like trying to nail Jello to the wall." *In* Bethany Allen-Ebrahimian, "The Man Who Nailed Jello to the Wall," Foreign Policy, June 29, 2016, https://foreignpolicy.com/2016/06/29/the-man-who-nailed-jello-to-the-wall-lu-wei-china-internet-czar-learns-how-to-tame-the-web/.

[19] Taina Bucher, *If...Then: Algorithmic Power and Politics* (New York: Oxford University Press, 2018).

as well. This interaction has been subtle for a long time: the benefits have been clear – and are significant – from the beginning, but the costs or risks took longer to arise and be perceived in the public arena.

Individuals constitute the basis of this growing landscape. Data, all of it, although generated in different ways and with different origins, is provided by people: it is socially constructed. And, when it comes to AI, "there is no data as more data", as Robert Mercer, founder of Cambridge Analytica, reportedly and revealingly claimed[20]. More data means more efficiency, more accuracy, better results. And our ever-expanding internet-use culminated in the appearance of Big Data, massive data sets containing gigabytes or terabytes (or more) of data that didn't exist until a few years ago. Big Data marks a fundamental change in multiple dimensions of our life.

Steel, oil, *data*

One of the fundamental transformative dynamics of this new Data Age is that individuals become, for the first time, direct providers of the most valuable matter[21]. We become 'raw material' providers. This material is the foundational element of an impactful and dynamic business model. Data is a very efficient raw material: every single particle is, in principle, usable and useful for some task, it is easily processed and transformed. Its processed result is not only valuable *per se*, but it can also enlarge and restart the cycle. Data-mining is a never-ending, highly profitable process.

Moreover, data is a resource that we humans can create ourselves. With AI and technological development, we are indeed producing it in a significant amount: the last

[20] Kai-Fu Lee, *AI Superpowers: China, Silicon Valley, and the New World Order* (Boston: Houghton Mifflin Harcourt, 2018), 108.

[21] "[A]s as individuals, we aren't just the recipients of the fruits of this data analysis: we *are* the data, and it is being used to make decisions *about us*—sometimes very consequential decisions." Kearns and Roth, *The Ethical Algorithm*, 2.

decade allegedly ended with 25 times more digital data than when it started[22]. In terms of profit, for big tech companies, this increment is phenomenal. Their business model is fundamentally centered on quantity – and the algorithmic processing ability allows precisely that. Quantity: of time spent, of clicks and comments, of questions, searches, reads, likes, shares, messages, unlocks, touches[23]. Individuals produce one of the most efficient raw materials that ever existed.

In a mine-like procedure, this material must be "extracted". As Shoshana Zuboff points out[24], the use of the concept of data "extraction"[25] is very revealing of the business dynamics that sustain this model. Extraction, she says, is a one-way process, not a relationship: "connotes a 'taking from' rather than either a 'giving to', or a reciprocity of 'give and take'." In fact, "[t]he extractive processes that make big data possible typically occur in the absence of dialogue or consent, despite the fact that they signal both facts and subjectivities of individual lives."[26]. The truth is that data is provided with only tacit consent, sometimes without the awareness of the data 'producer', and that the "[t]he current widespread culture of data extraction continues to grow despite concerns about

---

[22] Brad Smith and Carol Ann Browne, *Tools and Weapons: The Promise and the Peril of the Digital Age* (New York: Penguin Press, 2019), xiv.

[23] As Shoshana Zuboff puts it, "It is a ubiquitous networked institutional regime that records, modifies, and commodifies everyday experience from toasters to bodies, communication to thought, all with a view to establishing new pathways to monetization and profit" *in* Shoshana Zuboff, "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization," *Journal of Information Technology*, 2015, 81.

[24] Zuboff, "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization."

[25] Kate Crawford writes that "Terms like "data mining" and phrases like "data is the new oil" were part of a rhetorical move that shifted the notion of data away from something personal, intimate, or subject to individual ownership and control toward something more inert and nonhuman. Data began to be described as a resource to be consumed, a flow to be controlled, or an investment to be harnessed. The expression "data as oil" became commonplace, and although it suggested a picture of data as a crude material for extraction, it was rarely used to emphasize the costs of the oil and mining industries: indentured labor, geopolitical conflicts, depletion of resources, and consequences stretching beyond human timescales. Ultimately, "data" has become a bloodless word; it disguises both its material origins and its ends. And if data is seen as abstract and immaterial, then it more easily falls outside of traditional understandings and responsibilities of care, consent, or risk." Kate Crawford, *Atlas of Ai: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven: Yale University Press, 2021), 113.

[26] Zuboff, "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization," 79.

privacy, ethics, and safety."[27] As Kate Crawford points out, "the new AI gold rush consists of enclosing different fields of human knowing, feeling, and action – every type of available data – all caught in an expansionist logic of never-ending collection."[28]

Users typically lack knowledge and information about what is extracted, why it is valuable, and why it is used – and this obscurity is part of the business model and the corporate attitude of these companies. In other words, it is beneficial to companies to keep the nature and purposes of data collection opaque[29]. That attitude mirrors an indifference regarding the populations and societies in which it functions contrary to the free market's traditional market approach. Transparency and closeness are, usually, effective trust-building competition methods in a free market. There appears to be a certain immunity to the conventional reciprocities in which populations and capitalists needed one another[30].

"Senator, we run ads"[31]

However, this indifference might arise precisely due to a fundamental characteristic of this business model that we are acknowledging as part of a deceptive process[32]. Since

---

[27] Crawford, *Atlas of Ai*, 214.

[28] Crawford, 217.

[29] Are violations of privacy and fairness the result of incompetent software developers or, worse yet, the work of evil programmers deliberately coding racism and back doors into their programs? The answer is a resounding no. The real reasons for algorithmic misbehavior are perhaps even more disturbing than human incompetence or malfeasance (which we are at least more familiar with and have some mechanisms for addressing). Society's most influential algorithms — from Google search and Facebook's News Feed to credit scoring and health risk assessment algorithms — are generally developed by highly trained scientists and engineers who are carefully applying well-understood design principles. The problems actually lie within those very principles, most specifically those of machine learning. Kearns and Roth, *The Ethical Algorithm*, 7–8.

[30] Zuboff, "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization," 80.

[31] In April 2018, on his Congressional hearing, when asked how Facebook operates without making people pay Mark Zuckerberg simply replied "Senator, we run ads".

[32] "For many, Facebook and Google are part of the mythical "open web" – one where hyperlinks can be freely posted on any website. That makes sense on, say, Wikipedia, but it does not make sense on Facebook and Google which, despite their scale and mission statements, are essentially closed advertising ecosystems. Sure, news organisations can choose to post articles and videos on Facebook. And in return, people on Facebook may play those videos and click on those links. Rarely, they might click on an advert on the news organisation's website or pay to subscribe to that publication. (...) On Google, the value exchange is

the Cambridge Analytica scandal, disclosed in 2018, – and multiple cases ever since - it started to become clear that the users are not the clients[33]. Big Tech companies' business model is sustained on data. The services are 'offered' (free of charge) in exchange for data collection. Users are, however, more than 'raw material' providers. The business model is, in a sense, circular. After the extraction, the data is processed and analyzed by the algorithm. Then, it allows for what is denominated as 'profiling': "Profiling, in itself, means extrapolation of data available on the internet through processes of automated information gathering and subsequent construction and application of profiles"[34]. That is, the data analysis allows the creation of each user's present and future profiles and the construction of custom audiences, groups of users defined by specific characteristics, tastes, and behavior. And here is where much of the value of data comes from[35]: this in-

---

different but similarly broken. Google trawls the internet and indexes almost everything on it by default. A news organisation can opt out by inserting a short piece of code on their website's backend, but doing so would be commercially ruinous. Sure, publications don't *choose* to be on Google in the same way that they choose to be on Facebook, but the sheer scale of both platforms makes it an inevitability." James Temperton, "Facebook's Australia News Ban Is the Best Decision It's Ever Made," *Wired UK*, accessed June 29, 2021, https://www.wired.co.uk/article/facebook-australia-rupert-murdoch.

[33] "In early 2018, The New York Times reported that Cambridge Analytica, a political research firm, had inappropriately harvested information from millions of Facebook profiles, largely because of Facebook's improper stewardship of user data." Noam Scheiber and Mike Isaac, "Facebook Halts Ad Targeting Cited in Bias Complaints," *The New York Times*, March 19, 2019, https://www.nytimes.com/2019/03/19/technology/facebook-discrimination-ads.html. Scheiber and Isaac.

According to The Guardian, "the data analytics firm that worked with Donald Trump's election team and the winning Brexit campaign harvested millions of Facebook profiles of US voters, in one of the tech giant's biggest ever data breaches, and used them to build a powerful software program to predict and influence choices at the ballot box", Carole Cadwalladr and Emma Graham-Harrison, "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach," *The Guardian*, March 17, 2018, sec. News, http://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

[34] Committee of Experts on Internet Intermediaries (MSI-NET), "Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications" (Council of Europe, March 2018), 15.

[35] "Today, Google monitors every signal about us it can get its hands on. The power of this data can't be underestimated: If Google sees that I log on first from New York, then from San Francisco, then from New York again, it knows I'm a bicoastal traveler and can adjust its results accordingly. By looking at what browser I use, it can make some guesses about my age and even perhaps my politics.

How much time you take between the moment you enter your query and the moment you click on a result sheds light on your personality. And of course, the terms you search for reveal a tremendous amount about your interests." Eli Pariser, *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think* (New York: Penguin Books, 2014), chap. 1.

depth analysis of each users' social, emotional, biological, market behaviors. The business model is circular in the sense that individuals are at both ends of the process: as providers of the material and as targets of the processed information they shared themselves[36]. The raw material is harvested and processed, and its final product is information that can be used for very different purposes, from commercial to social, from political to military.

Advertisers are the biggest clients: data-intensive advertising helps generate over $150 billion a year[37]. Data analysis sustains detailed profiling, which allows micro-targeting and an ever-more efficient advertising platform[38]. This profiling also results in the "filter bubble"[39] – the users are mostly exposed to content and information that meets their preexisting beliefs and preferences and follows them up. Essentially, through our use of these platforms, we reveal information that leads to a constant reshaping of those same tools according to our preferences. The filter bubble facilitates what is called 'nudging'[40], a behavioral economics concept which entails an indirect and discreet

---

[36] "Popularity is not only a quantifiable measure that helps companies such as Facebook and Netflix to determine relevant content. User input and the patterns emerging from it are turned into a means of production. What we see is no longer what we get. What we get is what we did and that is what we see." Bucher, *If...Then*, 2.

[37] 2015 data, *in* Pasquale, *The Black Box Society*, 19. This value – even though we don't have the specific data – has most certainly increased since then (and keeps on doing so).

[38] "From the point of view of the online advertiser, the question is simple. Which company can deliver the most return on a dollar spent? And this is where relevance comes back into the equation. The masses of data Facebook and Google accumulate have two uses. For users, the data provides a key to providing personally relevant news and results. For advertisers, the data is the key to finding likely buyers. The company that has the most data and can put it to the best use gets the advertising dollars." Pariser, *The Filter Bubble*, chap. 1.

[39] Term coined by the Internet activist Eli Pariser in his book "The Filter Bubble: What the Internet is Hiding From You", The Penguin Press, 2011. : "The basic code at the heart of the new Internet is pretty simple. The new generation of Internet filters looks at the things you seem to like—the actual things you've done, or the things people like you like—and tries to extrapolate. They are prediction engines, constantly creating and refining a theory of who you are and what you'll do and want next. Together, these engines create a unique universe of information for each of us—what I've come to call a filter bubble—which fundamentally alters the way we encounter ideas and information." Pariser, sec. Introduction.

[40]"Nudging is a behavioral concept imported from sociology, where indirect suggestions and hints impact the decisions of institutions and individuals without a direct threat or persuasion. A robust nudge is not a threat, or coercion, as it does not appear binding and most of the time guides the target into the desired set of behavior in a way that such decision looks like the target's own choice, among the alternatives. Due to their ability to harvest large amounts of user decision and choice, algorithmic architectures are ideally-

suggestion that steers someone's behavior, desired choices, and actions without them realizing it, and often while they assume they have full agency in the decision required. This influence and consequent societal impact turn algorithms into more than merely "coded instructions telling the machine what to do, but rather "emergent accomplishments or socio-material practices."[41]

It is important and fair to state that we all have benefited from this model. There has never been faster and cheaper access to information and communication; we have never had such easy access to information and access to such developed technology.[42] But we must be aware that this was not a gift (even if the services are apparently "free"); it is an exchange. And even though we all benefited from these exchanges so far, the less we analyze carefully and critically, the more it poses a risk for our own development – as social beings and as democratic citizens – and for the development and sustainability of a healthy liberal democracy.

Given the increase of data provided, the evolution of algorithmic development, and the improvement of computing power, we are in a phase of nearly optimal personalization. The engagement is ever-increasing and feeds a virtuous (or vicious?) cycle. These companies are gradually more capable of reporting, predicting, and modifying human behavior and, therefore, increasing engagement. We start perceiving

---

designed for automated nudges because they can automatically learn from user data to make successive choices conducive for a nudge. With a large live dataset of user behavior, A.I.-based systems will have an extensive repertoire of possible nudges that can steer individuals into the desired choices and behavior without them realizing it, and worse, appear as if they have agency in the decision required." H. Akın Ünver, "Artificial Intelligence, Authoritarianism and the Future of Political Systems," Cyber Governance and Digital Democracy 2018/9 (Centre for Economics and Foreign Policy Studies, July 2018), 8.

[41] Bucher, *If...Then*, 152.

[42] Hannah Fry, *Hello World* (New York: W.W. Norton & Company, 2018), chap. Data. : "It's important to remember that we've all benefited from this model of the internet. All around the world, people have free and easy access to instant global communication networks, the wealth of human knowledge at their fingertips, up-to-the-minute information from across the earth, and unlimited usage of the most remarkable software and technology, built by private companies, paid for by adverts."

ourselves through our engagement: our self-expression, our empowerment, our connections, our voice is now based on and expressed primarily in and through engagement with these platforms[43].

"Don't be evil"[44]

Our relation with digital technology shapes us in a deep psychological sense. The tech writer Nicholas Carr claims that our constant disposition to reach to technology for aid, companionship, orientation, and research leads us to develop certain attitudes towards technology. The author claims we often fall victim to a pair of perspectival and cognitive ailments, namely "automation complacency and automation bias."[45]

Automation complacency consists essentially in the false sense of security we experience when dealing with a machine. We are so confident in the flawless functioning of the machine that we eventually become detached, disengaged from the activity we are performing[46]. Automation bias is the excess of faith we place in the accuracy of the algorithm's answer, as opposed to other sources or authorities. This trust is so absolute and unquestionable that it might lead us to dismiss other information sources, our own senses included[47]. Both complacency and bias are "symptoms of a mind that is not being

---

[43] As Shoshana Zuboff puts it "Individual needs for self-expression, voice, influence, information, learning, empowerment, and connection summoned all sorts of new capabilities into existence in just a few years: Google's searches, iPod's music, Facebook's pages, YouTube's videos, blogs, networks, communities of friends, strangers, and colleagues, all reaching out beyond the old institutional and geographical boundaries in a kind of exultation of hunting and gathering and sharing information for every purpose or none at all." Zuboff, "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization," 79.

[44] Google's unofficial motto, 'Don't be evil' has been part of the corporate code of conduct since 2000. It was removed in 2018, when Google was reorganized under a new parent company, Alphabet. It then assumed a slightly adjusted and 'nicer' version of the motto: "Do the right thing". Katy Conger, "Google Removes 'Don't Be Evil' Clause From Its Code of Conduct," Gizmodo, May 18, 2018, https://gizmodo.com/google-removes-nearly-all-mentions-of-dont-be-evil-from-1826153393.

[45] Nicholas G Carr, *The Glass Cage: Automation and Us* (New York: W. W. Norton & Company, Inc., 2014), chap. 4.

[46] Carr, chap. 4.

[47] Carr, chap. 4.

challenged, that is not fully engaged in the kind of real-world practice that generates knowledge, enriches memory, and builds skill."[48] Carr claims we turn from actors into observers. We become creatures of the screen – and the screen guides us through who we are. More and more, as George Dyson writes, "Facebook defines who we are, Amazon defines what we want, and Google defines what we think."[49]

This ability to influence is a form of power, and with great power comes great responsibility. This cliché is precisely one of the pillars of a liberal democratic regime. In a democratic regime, this responsibility is sustained upon public accountability and transparency. The possible impact these companies have on human minds and democratic societies should entail increased corporate responsibility. That does not seem to happen. In fact, these companies encourage a 'black box' model: even though they have access to every detail of our digital life, *we* cannot access fundamental features of the functioning of *their* decision-making process[50]. We know little to nothing about how they use their data and analytics. We know little to nothing about how they use their power. This imbalance of information is threatening to liberal democracies.

---

[48] Carr, chap. 4.

[49] George Dyson and Inc, *Turing's Cathedral* (New York: Pantheon Books, 2012), chap. 17.

"Adam J. White glosses these statements: "To say that the perfect search engine is one that minimizes the user's effort is effectively to say that it minimizes the user's active input. Google's aim is to provide perfect search results for what users 'truly' want—even if the users themselves don't yet realize what that is. Put another way, the ultimate aspiration is not to answer the user's questions but the question Google believes she should have asked." As Eric Schmidt [former Google CEO] told the Wall Street Journal, "[O]ne idea is that more and more searches are done on your behalf without you having to type. . . . I actually think most people don't want Google to answer their questions. They want Google to tell them what they should be doing next." *In* Matthew Crawford, "Algorithmic Governance and Political Legitimacy," *American Affairs* III, no. 2 (Summer 2019), https://americanaffairsjournal.org/2019/05/algorithmic-governance-and-political-legitimacy/.

[50] For instance, "Google's search algorithm is one of its best-kept trade secrets. The company developed the algorithm in 1997 and continues to refine and update it (…). While the company announces some changes, most of them are unannounced, so that people and businesses are kept in the dark." "Everything You Need to Know about Trade Secrets | IPTSE," October 19, 2020, https://iptse.com/understanding-trade-secrets/.

Communication and information are fundamental pillars of a liberal democracy. They lie at the core of freedom of expression, and empower citizens to think critically and make informed decisions. In fact, in all its elements, a liberal democracy is fundamentally sustained in its capability of embracing antagonisms and tensions within its social body in a frequently unstable equilibrium: and that capability is one of the core elements of the possibility of living in freedom[51]. This requires, from every institution and every citizen, the ability to deal with ambiguity and discussion, difference and disagreement. One of the threats these corporations pose to democracies is an oligopolistic detention of communication and information means in a profit-driven mindset[52]. Big Tech is currently sustained on an idea of accumulation, manipulation, and influence – all inside 'the box'. These digital platforms, wielded on a brutal economic and influential unchecked power, are communication bottlenecks. They act, and are in fact, "crucial gatekeepers[53] for human beings who wish to seek, receive or impart information."[54] They dominate the sharing and dissemination of information – capable of

[51] "The growing use of AI in more and more areas represents a retreat from facing conflict, an attempt to remove it from the social field." Marie David, "AI and the Illusion of Human-Algorithm Complementarity," *Social Research: An International Quarterly* 86, no. 4 (2019): 903.

[52] "As long as AIs drive an automated public sphere optimized for profits, expect the same pattern: flurries of concern spurred by media shaming, followed once again by irresponsibility. Although such backsliding is not inevitable, it is a clear and present danger of AI driven primarily by the demands of shareholders of massive firms." Frank Pasquale, *New Laws of Robotics* (Harvard University Press, 2020), 95.

[53] "While there is no settled definition of what constitutes a digital gatekeeper, this term commonly refers to platforms providing online services (e.g. online marketplaces) or controlling and influencing access to online services (e.g. operating systems, app stores and voice assistants) and thereby exercising control over entire ecosystems, with a strong impact on competition and innovation in the digital field." European Parliament, "Regulating Digital Gatekeepers - Background on the Future Digital Markets Act," December 8, 2020, 2.

[54] Committee of Experts on Internet Intermediaries (MSI-NET), "Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications," 17.

filtering it and manipulating it[55], or even shutting it down, as we've seen in Australia's 2021 Facebook news ban[56] (since revoked but still a powerful illustration).

This poses a severe threat to democracy. 'Filter bubbles' cluster individuals[57] into groups of similar ideas and people of similar dispositions, which "reinforces a positive view of their own tastes, and further distances them from other tastes which become more and more negative to them, especially with regard to political views" [58]. This reinforcement contributes "political polarization and ever more extreme views"[59]. AS Zuboff notes, this *modus operandi's* "effects are felt in the real world, where they splinter shared reality, poison social discourse, paralyze democratic politics, and sometimes instigate violence and death."[60]. Ultimately, Google's former unofficial conduct motto appears to be, disconcertingly, our last redoubt of hope: 'Don't be evil'. But no liberal democracy can be content to entrust concentrated political and economic power based on

[55] "The real problem centers around the platforms' ability to either amplify or silence certain messages, and to do so at a scale that can alter major political outcomes" Francis Fukuyama, "Making the Internet Safe for Democracy," *Journal of Democracy* 32, no. 2 (April 2021): 39.

[56] Facebook blocked news to Australians on its platform amid a dispute over a proposed law which would force it and Google to pay news publishers for content. After conversations with the Australian government, it has reversed the ban.

[57] Where " the values reflected back on us are our own, some may be in the form of news articles, posts by friends, and some through advertising which aims to target us through what people of similar group demographics and tastes liked" *in* Lauren Toulson, "Why Algorithms Erode Democracy. The Way You Navigate the Internet Is… | by Lauren Toulson | CARRE4 | Medium," February 23, 2021, https://medium.com/carre4/why-algorithms-erode-democracy-d1fcdeab63a5.

[58] Toulson.

[59] Without any human editors or managers to take direct responsibility for algorithmic choices, online platforms can help the worst elements in societies appear credible and authoritative. (…) In the fever swamp of automated search results, support for climate denialists, misogynists, ethno-nationalists, and terrorists easily grows and spreads."Frank Pasquale, *New Laws of Robotics*, 98.

[60] Shoshana Zuboff, "Opinion | Facebook and the Surveillance Society: The Other Coup - The New York Times," New York Times, January 29, 2021, https://www.nytimes.com/2021/01/29/opinion/sunday/facebook-surveillance-society-technology.html.

assumptions and hopes of good intentions[61]. There are fundamental conditions which must be positively and formally protected – among them, fundamental human rights.

The Article 12 of Universal Declaration of Human Rights (UDHR) affirms the right to *privacy*[62]. The right to privacy is inherently associated with data protection. The access to private information that bulk data collection and the ability of AI tools to track and analyze our online and offline lives represents a direct threat to this right[63]. Moreover, freedom of expression, thought, religion, assembly, and association (articles 18, 19, 20 of the UDHR[64]) are rights that only truly flourish in a landscape of freedom. They thrive when the right to privacy is guaranteed. These technologies can identify and track people, their movements, relationships, thoughts, and beliefs. The fear that might be instigated by the possibility of being watched or lacking anonymity unveils effective limitations to

[61] Francis Fukuyama, Barak Richman, and Ashish Goel, "How to Save Democracy From Technology | Foreign Affairs," *Foreign Affairs*, February 2021, https://www.foreignaffairs.com/articles/united-states/2020-11-24/fukuyama-how-save-democracy-technology.

[62] **Article 12** claims: "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks." United Nations, "Universal Declaration of Human Rights," 1948, https://www.un.org/en/about-us/universal-declaration-of-human-rights.

[63] "The analysis of data using AI systems may reveal private information about individuals, information that qualifies as protected information and should be treated as sensitive even if derived from big data sets fed from publicly available information." Andersen, "Human Rights in the Age of Artificial Intelligence."

[64] **Article 18**: Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief, and freedom, either alone or in community with others and in public or private, to manifest his religion or belief in teaching, practice, worship and observance.

Article 19: Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.

Article 20:

1. Everyone has the right to freedom of peaceful assembly and association.
2. No one may be compelled to belong to an association.

United Nations, "Universal Declaration of Human Rights."

the constrained exercise of fundamental human rights[65]. These same risks are deepened by the analysis we will entail in the next part: the institutional use of data and algorithms.

2.2. The Road to Hell is Paved with Good Intentions: The Challenges posed by Algorithmic-Decision Making

On an institutional level, Artificial Intelligence has also been the source of multiple advances. There is a growing socio-technological approach to various dimensions of institutional processes in terms of improving systematic decision-making. Algorithmic decision systems have been increasingly applied in areas such as credit access, employment, information, e-commerce, recommendation systems, health, justice, policing, banking and insurance. AI and ADS might contribute to solving multiple public and private institutional dysfunctions, namely in terms of inefficiency, corruption, or bias. In this part, we'll be more focused on public institutions' deployment of AI, but these risks we present are constant in many private sector dynamics of algorithmic use.

This analysis is increasingly relevant because "once AI moved out of the laboratory contexts of the 1980s and 1990s and into real-world situations—such as attempting to predict which criminals will reoffend or who should receive welfare benefits—the potential harms expanded."[66] Multiple cases point out problematic dimensions of ADS deployment in many instances. ADS have been proven to perpetuate particular bias and foster discrimination in fundamental areas of liberal democratic institutions. ADS models used in institutional procedures are designed to sort and filter, whether by ranking search results or categorizing people into buckets. Anonymity is a fundamental principle of the

[65] As we will analyse more thoroughly ahead, "Violations of the right to privacy have a chilling effect on free expression. When people feel that they are being watched, or lack anonymity, they have been shown to self-censor and alter their behavior". Andersen, "Human Rights in the Age of Artificial Intelligence."

[66] Crawford, *Atlas of Ai*, 208.

idea of the rule of law: everyone is completely equal before the law and entitled to equal protection against any discrimination (articles 7 and 8 of the UDHR [67]). This discrimination can interfere with human rights when treating different groups of people differently[68].

In criminal justice, for instance, some risk assessment algorithms are used to predict a defendant's future risk for misconduct. One infamous case is the one of COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), an algorithm which was proved to be twice as likely to mistakenly label you as high risk if you were black as it was if you were white[69]. In law enforcement, algorithmic predictive policing can also unduly target a 'type' of person/community (e.g., of particular ethnic belonging, race, immigrants, etc.) as more crime-prone, bypassing or ignoring important socio-historic dynamics or biases at play[70].

AI can be biased both at the system and at the input level. Bias at the system level involves developers building their own bias into the parameters or labels defined. At the input level, ADS may be biased due to the data that it is fed with. Algorithms are provided data that contains pre-existent social and cultural biases. And "every dataset used to train machine learning systems, whether in the context of supervised or unsupervised machine learning, whether seen to be technically biased or not, contains a worldview."[71] The

---

[67] **Article 7**: All are equal before the law and are entitled without any discrimination to equal protection of the law. All are entitled to equal protection against any discrimination in violation of this Declaration and against any incitement to such discrimination.

Article 8: Everyone has the right to an effective remedy by the competent national tribunals for acts violating the fundamental rights granted him by the constitution or by law.

United Nations, "Universal Declaration of Human Rights."

[68] Andersen, "Human Rights in the Age of Artificial Intelligence," 24.

[69] Fry, *Hello World*, chap. Justice.

[70] Marie-Valentine Florin and Kujtese Bejtullahu-Michalopoulos, "The Governance of Decision-Making Algorithms" (EPFL International Risk Governance Center, November 2018).

[71] Crawford, *Atlas of Ai*, 135.

outputs that arise from the analysis of that same data comprehensively reflect that worldview, mirror those biases, contributing to their perpetuation. The past actively conditions the future. The fact is that "[h]umans have always deployed technology with the hope of improving the systems that operate around them"[72] and so, when "the data reveals racism and bias in the system, risk assessment tools must account for that bias"[73].

In all these areas, however, human analysis and human decisions would also have their own biases[74] and perpetuate certain discriminations or inequalities[75]. It is fundamental to compare the risk of bias with and without ADS. However, algorithmic biases might be more problematic than human biases because they are shielded by a precedent partiality regarding the algorithm's effectiveness and fairness *per se.* In fact, "the rationale offered is that automated decision-making will be more reliable"[76] — and in a liberal democracy we welcome any tool that might help increase procedural fairness and impartial judgement[77].

---

[72] Vincent Southerland, "With AI and Criminal Justice, the Devil Is in the Data," American Civil Liberties Union, April 9, 2018, https://www.aclu.org/issues/privacy-technology/surveillance-technologies/ai-and-criminal-justice-devil-data.

[73] Southerland.

[74] For instance, "Decades of criminological research, dating to at least the nineteenth century, have shown that police databases are not a complete census of all criminal offences, nor do they constitute a representative random sample. Empirical evidence suggests that police officers – either implicitly or explicitly – consider race and ethnicity in their determination of which persons to detain and search and which neighborhoods to patrol." Kristian Lum and William Isaac, "To Predict and Serve?," *Significance* 13, no. 5 (October 2016): 16, https://doi.org/10.1111/j.1740-9713.2016.00960.x.

[75] "Understanding the relation between bias and classification requires going beyond an analysis of the production of knowledge (…) To see that requires observing how patterns of inequality across history shape access to resources and opportunities, which in turn shape data. That data is then extracted for use in technical systems for classification and pattern recognition, which produces results that are perceived to be somehow objective. The result is a statistical ouroboros: a self-reinforcing discrimination machine that amplifies social inequalities under the guise of technical neutrality." Crawford, *Atlas of Ai*, 130.

[76] Matthew Crawford, "Algorithmic Governance and Political Legitimacy", American Affairs Journal, Volume III, no. 2, (Summer 2019): 1.

[77] As Marie David puts it "By appearing to render certainty, algorithms create the illusion of control. (...)With algorithms, we once again seek to remove this essential experience of uncertainty from human life" *in* David, "AI and the Illusion of Human-Algorithm Complementarity," 893.

However, ADS' decisions are frequently perceived through the cognitive ailments of bias and complacency, as Carr notes. Moreover, in this context, these attitudes gain an institutional force – and that transforms them into immediate risks to democratic structural functioning. This institutional algorithmisation is linked to a deeper attitude ascending in tech community, to which Meredith Broussard (2018) calls "technochauvinism": the belief that tech is always the solution (Broussard, 2018, p. 95). AlgorithmWatch, a non-profit research and advocacy organization commited to evaluating and shedding light on ADM processes, also warns about a tendency of "technological solutionism", an ideology that conceives of every social problem as a "bug in need of a fix" through technology, rhetoric widely adopted – both in the media and in policy circles – to justify the uncritical adoption of automated technologies in public life[78].

We increasingly rely on the objectivity of the algorithm. All these conceptions are bolstered by the idea of neutrality or objectivity – a confident assurance that we take from a "mathematical formulation". This idea of total neutrality is, as we now see, a myth. Moreover, it is difficult to demand accountability or analyze the alleged neutrality: ADS contributes significantly to the black box society. And that goes directly against a democratic institutional structuring – based on transparency and accountability[79]. As Freedom House (Shahbaz & Funk, 2020) points out, algorithms are quickly replacing human judgment in vital areas of human life. The results are likely to create new

[78] Algorithm Watch, "Automating Society Report 2020" (Algorithm Watch, October 2020), 10, https://automatingsociety.algorithmwatch.org.

[79] As Marie David puts it "Since algorithms have become extremely complex and process huge data volumes, it is difficult, if not impossible, to understand how they arrive at particular decisions. In fact, the problem of bias will make coexistence between human and artificial rationality completely unrealistic. If a human expert and an algorithm disagree, then we are very likely to deem the machine more reliable. Its calculations appear unequivocal. Moreover, algorithms cannot "justify" their decisions, paradoxically making them seem more reliable than people who can. Human decisions can always be reviewed, debated, and rejected, whereas algorithmic decisions will remain impenetrable, always the result of byzantine calculations." David, "AI and the Illusion of Human-Algorithm Complementarity," 896–97.

inequalities and further disadvantage those who were already vulnerable to discrimination[80].

Opacity is once again a key issue. If the algorithm behind institutional decisions is not accessible to every citizen[81], this contributes to a de-legitimation of institutions. The previously mentioned COMPAS, for instance, "is a proprietary, closed-source tool, so neither attorneys, defendants, nor judges know exactly how its model works."[82] Not only is it sometimes not accessible, but it can also be *incomprehensible*: millions of lines of code make it difficult to understand or to make sure what principles were inculcated in the software and how algorithms make decisions. In liberal democracies, it is essential to understand the process preceding the decision to be able to agree or disagree with it: the procedural part of the decisions is, in a liberal democracy, often more important than its result. Institutional decisions – political, judicial, institutional – are grounded on law, principles, ideals, precedents: they are articulate and defended (or defendable). There is an accountability to be given, and that is the basis of political and institutional legitimacy in liberal democracies as Matthew Crawford writes,– "political accountability is the very essence of representative government."[83]. A court decision, for instance, is explained and its reasoning is presented: "[h]e grounds the decision in law, precedent, common sense, and principles that he feels obliged to articulate and defend. This is what transforms the

[80] "Machine learning systems are, in a very real way, constructing race and gender: they are defining the world within the terms they have set, and this has long-lasting ramifications for the people who are classified. When such systems are hailed as scientific innovations for predicting identities and future actions, this erases the technical frailties of how the systems were built, the priorities of why they were designed, and the many political processes of categorization that shape them. " Crawford, *Atlas of Ai*, 146.

[81] "Such algorithms tend to be shrouded in secrecy, closed off to external auditors who might be able to test them for bias and make needed corrections. Because of that secrecy, often imposed by private companies that refuse to reveal their source code, people can't effectively contest the decisions made by these tools. The data and algorithms used to make fateful decisions about people's lives are simply out of public reach. In "Will Artificial Intelligence Make Us Less Free?," American Civil Liberties Union, accessed June 14, 2021, https://www.aclu.org/issues/privacy-technology/will-artificial-intelligence-make-us-less-free.

[82] Christian, *The Alignment Problem*, 18.

[83] Crawford, "Algorithmic Governance and Political Legitimacy."

decision from mere fiat into something that is politically legitimate, capable of securing the assent of a free people. It constitutes the difference between simple power and authority."[84]

This opacity makes it difficult to challenge a decision based on algorithmic results and might open space for arbitrariness. Institutional deployment of ADS can contribute, therefore, to another focus of non-scrutinized authority. This type of authority reaches a climax with the technological tools we will analyse in the following part.

2.3. Every step you take, every move you make: The Challenges posed by Surveillance Technology

One of the most impressive aspects of Artificial Intelligence is its ubiquity. Omnipresence and power are a combination that does not go well with liberal democratic conceptions of the rule of law and liberty. This third 'layer' is different from what we've previously analyzed because it is not applied to specific, contingent societal cases: from the moment these technologies are deployed, they embrace every citizen and every citizen lives within that network from then on.

New technologies bring unprecedented monitoring and surveillance capabilities; technology has brought new possibilities that change the nature of how governments carry out surveillance and what they choose to monitor[85]. Balancing technological deployment and security interests with civil liberties and human rights protection becomes a strenuous and vital endeavor.

---

[84] Crawford.

[85] Steven Feldstein, "The Global Expansion of AI Surveillance," Carnegie Endowment for International Peace, September 2019, https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847.

According to the "Artificial Intelligence Global Surveillance Index"[86] (AIGS), at least 75 out of 176 countries globally are actively using AI technologies for surveillance purposes. This includes smart city/safe city platforms (56 countries), facial recognition systems (64 countries), and smart policing (52 countries)[87]. Liberal democracies are major users of AI surveillance[88]. The index shows that 51 per cent of advanced democracies deploy AI surveillance systems. In contrast, 37 per cent of closed autocratic states, 41 per cent of electoral autocratic/competitive autocratic states, and 41 per cent of electoral democracies/illiberal democracies deploy AI surveillance technology[89]. As the AIGS report points out, it is essential to note that state surveillance technology is not inherently unlawful or undemocratic. In the report, the AI surveillance technology is perceived as "value-neutral"[90]. The reasons behind this technology's deployment need not be rooted in a desire to increase political repression and control or limit individual freedoms. Their presence does not mean that a government is using them for anti-democratic purposes[91]. But the line is very thin, and the boundaries are subtle. These boundaries are also easily 'crossed': once the technical surveillance capacities are in place, it is a small step – or, at least, easier – to their illegitimate or unethical deployment.

According to the previously mentioned Report, AI Surveillance technologies can be divided into three distinct areas: smart cities or safe cities, facial recognition systems, and smart policing[92]. Smart cities aim at improving city management and service delivery.

---

[86] Feldstein.

[87] Feldstein, 1.

[88] Feldstein, 2.

[89] Feldstein, 2.

[90] Feldstein, 13.

[91] The purpose of the report is 'merely' to "to identify which countries possess sufficiently advanced tools that allow them to pursue a range of surveillance objectives". – which is actually concerning enough. Feldstein, 13.

[92] Feldstein, "The Global Expansion of AI Surveillance."

This technology may help municipal authorities managing traffic congestion, direct emergency vehicles, improve sustainability and increase administrative processes. One of the main goals of smart cities is also public safety. IBM, the American technology company and one of the firsts to coin the term, designed "a brain like municipal model"[93] that centralizes the processing and analysis of information relevant to city safety and aims to increase police and security forces capabilities. The data is harvested by sensors, tracking devices, and surveillance technology. In our view, an obvious concern is the centralization of so much information about citizen movements, especially when the line that defines security and what it allows is not a very fixed or *linear line*.

Facial recognition is a biometric technology that uses cameras to match stored or live footage of individuals with images from a database. It is different from CCTV because it can create detailed biometric maps of individuals without consent or even awareness. This technology, whose deployment has steadily increased over the past few years, is particularly intrusive and raises very sensitive issues. One of the reasons for that it has mostly been deployed in a vacuum of regulations that resulted in a "surveillance-first, ask permission-later system"[94]". Facial recognition has given proof of higher error rates when used on women and people of color, for instance, producing biased results, which can ultimately result in discrimination[95].

Smart Policing is an algorithmic tool to which is fed a significant amount of data – data such as "geographic location, historic arrest levels, types of committed crimes, biometric data, social media feeds" to "prevent crime, respond to criminal acts, or even

[93] Feldstein, 17.

[94] Feldstein, 19. Citing *Washington Post*, Drew Harrel.

[95] FRA, "Facial Recognition Technology: Fundamental Rights Considerations in the Context of Law Enforcement" (Vienna – Austria: FRA – EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 2019).

to make predictions about future criminal activity"[96]. Smart policing raises severe concerns in terms of civil liberties and discrimination, especially because, similarly to ADS, it is directly shaped by past data, possibly contributing therefore to the perpetuation of biases.

Besides their particular impact, what is of serious concern is the system enabled by these modern surveillance structures. This type of system[97] has often been compared with Jeremy Bentham's idea of the *panopticon*[98]. An all-seeing non-visible power. In an interesting twist, Zuboff rejects that comparison. Compared to the surveillance system that AI enables, the *panopticon* architecture "is prosaic"[99] – it had a single point of observation. With AI, power can no longer be summarized by the "totalitarian symbol of centralized command and control": this structure is as vertical as horizontal and records, modifies, and commodifies every aspect of everyday experience. However, one element of the panopticon remains present. The perception of constant surveillance slowly erodes society, changes behaviors, and reshapes minds – and that might be one of the more

---

[96] Feldstein, "The Global Expansion of AI Surveillance."

According to the RAND Corporation, predictive policing is defined as "the application of analytical techniques – particularly quantitative techniques – to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions".13 Much like how Amazon and Facebook use consumer data to serve up relevant ads or products to consumers, police departments across the United States and Europe increasingly utilise software from technology companies, such as PredPol, Palantir, HunchLabs, and IBM to identify future offenders, highlight trends in criminal activity, and even forecast the locations of future crimes. *In* Lum and Isaac, "To Predict and Serve?," 16.

[97] And, mostly, the Chinese system – which we will approach in the next subchapter.

[98] "The panopticon is a disciplinary concept brought to life in the form of a central observation tower placed within a circle of prison cells. From the tower, a guard can see every cell and inmate but the inmates can't see into the tower. Prisoners will never know whether or not they are being watched.

This was introduced by English philosopher Jeremy Bentham. It was a manifestation of his belief that power should be visible and unverifiable. Through this seemingly constant surveillance, Bentham believed all groups of society could be altered. Morals would be reformed, health preserved, industry invigorated, and so on – they were all subject to observation." *In* "Ethics Explainer: The Panopticon - What Is the Panopticon Effect?," *THE ETHICS CENTRE* (blog), July 18, 2017, https://ethics.org.au/ethics-explainer-panopticon-what-is-the-panopticon-effect/.

[99] Zuboff, "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization," 82.

significant dangers. It can also be one of the greatest opportunities for authoritarian regimes, such as the one of the Chinese Communist Party.

2.4. Smile you're being watched: the Chinese Updated Dictatorship

China has enthusiastically embraced and invested in the technologies we have mentioned above. Digital technologies and technological possibilities such as bulk data collection and AI processing of data enable both economic growth and the deepening and widening of multiple repressive efforts and activities.

China's dictatorship is updating itself with the tools of the 21st century and developing a cutting-edge technological authoritarian regime. The Chinese party-state is investing in data collection on a massive scale as a means of generating information and enhancing state security, and crucially the political security of the Chinese Communist Party. [100] The Chinese State Council clearly states the importance of this technology in its 2017 "New Generation of Artificial Intelligence Development Plan":

> Artificial intelligence technology can accurately perceive, forecast, early warn the major trends of infrastructure and social security operation, timely grasp the change of group awareness and psychology, respond actively decision-making, significantly improve the ability and level of social governance, and it is indispensable for the effective maintenance of social stability.

The "effective maintenance of social stability" is an ambiguous expression that might have various meanings and practical outcomes. In the CCP, this stability is fundamental to maintaining 'harmony', one of the party's favorite words in the past

[100] Samantha Hoffman, "Engineering Global Consent: The Chinese Communist Party's Data-Driven Power Expansion," Policy Brief (Australian Strategic Policy Institute, October 2019), 4.

decade[101]. The CCP is establishing a holistic structure of control that aims at what Samantha Hoffman calls "social management"[102]. Social management aims at maintaining the constant stability of the current social and political order and is specifically directed at preemptively ensuring state security – and this security goes beyond the prevention of civil unrest. The 'harmony' must exist in every aspect of everyday life. The structure behind the goal of social management relies on technology that is used in everyday life and services of normal economic and social activity, such as online activity and electronic purchases, in order to exert and expand political control. It is very interesting how "the system's ability to solve and manage problems does not diminish its political or coercive capacity"[103], on the contrary: it is precisely this ability to solve and manage everyday problems that give the CCP's system the ability to control and coerce.

The ubiquity of technology allows what Hoffman calls "control through convenience"[104]. The control is garnered by the creation of a desire in the individual herself to use or be a part of the system. The permanent connectivity, ease of use, and convenience are a big part of the growing use of technologies in our daily lives. In China, the state obtaining data and asserting its interests is 'part of the package' of this system of control-through-convenience.

---

[101] "The autocrat who wants to create his own truth needs to conquer the word. In China there is no repression; there is simply 'maintaining stability' (weiwen) and a 'harmonious society' (hexie shehui). In the past decade, harmony has been one of the Party's favourite words: the harmony between orders and obedience. Harmony is when ordinary people don't make a fuss. *In* Kai Strittmatter, *We Have Been Harmonised: Life in China's Surveillance State* (United Kingdom: Old Street Publishing, 2019), chap. The Word.

[102] Samantha Hoffman, "Programming China: The Communist Party's Autonomic Approach to Managing State Security." (PhD Thesis, University of Nottingham, 2017).

[103] Samantha Hoffman, "Social Credit: Technology-Enhanced Authoritarian Control with Global Consequences," Policy Brief (Australian Strategy Policy Institute, June 28, 2018), 5, https://www.aspi.org.au/report/social-credit.

[104] Hoffman, "Engineering Global Consent: The Chinese Communist Party's Data-Driven Power Expansion."

A central element of all this strategy is the Social Credit System (SCS). The SCS is, essentially, Big Data collection and analysis to monitor, shape, and rate behavior through economic and social processes. In this 'citizens' score system', good behavior and loyalty are rewarded with literal points, and the breaking of trust or disobedience are likewise sanctioned. Loyalty facilitates your life; breaking the trust makes it (sometimes a lot) harder[105]. The Chinese State Council made that evident in their "Planning Outline for the Construction of a Social Credit System"[106]: "it uses encouragement to keep trust and constraints against breaking trust as incentive mechanisms, and its objective is raising the honest mentality and credit levels of the entire society."

The SCS aims at building and sustaining this "trust" and "social harmony", controlling individuals', companies', and other entities' behavior – economic, social, and moral – and their conformity with the CCP's policies, directions, and will. The SCS "is intended to record every action and transaction by each Chinese citizen in real-time and to respond to the sum of an individual's economic, social and moral behavior with rewards and penalties"[107]. This is only possible thanks to AI technology: it enhances the states' ability to monitor, reward and repress – and, more impressively, to shape new people: citizens with an "honest mentality".

---

[105] "Measures of this kind have already been put into effect: By mid- 2018, due to poor social-credit scores, more than 11 million people had reportedly been placed under limits on the purchase of airline tickets, and 4.25 million people were restricted in buying high-speed rail tickets. For those whose political activity draws negative official attention, this system is likely to have serious repercussions. The experiences of individuals currently on social-credit blacklists suggest that those subject to restrictions may not be notified when listed, and will not have easy access to appeal procedures." Qiang Xiao, "The Road to Digital Unfreedom: President Xi's Surveillance State," *Journal of Democracy* 30, no. 1 (January 2019): 60.

[106] Chinese State Council, "Planning Outline for the Construction of a Social Credit System (2014-2020)," *China Copyright and Media* (blog), June 14, 2014, https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/.

[107] Strittmatter, *We Have Been Harmonised*, chap. New China, New World.

Only with AI can a system like this exist. AI technology not only makes this possible, but it also assures a more efficient and less costly oppression system. AI technology allows the automation of many functions that were performed by dense security force infrastructures, more expensive, and more prone to human error. Wide-spread surveillance, intimidation, harassment, and spreading of fear are not only easier[108] but also cover a much wider net. The omnipresent algorithms eventually start fulfilling the goal of harmony and prosperity: they create "economically productive, socially harmonized and politically compliant subjects, which will ultimately censor and sanction themselves at every turn"[109]. The chilling effect of a constant vigilance creates the habit of 'good behavior' – the pre-emptive dimension of social management establishes a society that almost does not need to be managed.

Bao Pu, a Chinese writer and activist[110], puts into words what is hard not to feel when understanding the authoritarian possibilities of AI: "Technology always benefits the side with the greater resources. So the internet will always serve the CCP more than its opponents."[111]. As we've seen, there's no data like more data (i.e., large datasets lead do better AI), and one of China's natural resources is precisely an overabundance of data with few privacy protections.

---

[108] "Instead of relying on a dense security-force infrastructure to enable wide-spread surveillance, harassment, and intimidation of opponents across the state's territory, authoritarian leaders can use AI to cultivate a digital repression capability at a lower cost—and reduce principal-agent concerns. In fact, the most advanced surveillance operations rely on relatively few human agents: Many functions are instead automated through AI. Moreover, in comparison to human operatives with limited reserves of time and attention, AI systems can cast a much wider net. (…) Such is the elegant simplicity of AI repression: It requires considerably fewer human actors than conventional repression, entails less physical harassment, and comes at a lower cost. (…). In comparison, for example, East Germany's Stasi security service relied upon an informant network equivalent to 1 percent of the country's total population leading to sizeable and persistent economic costs." Steven Feldstein, "The Road to Digital Unfreedom: How Artificial Intelligence Is Reshaping Repression," *Journal of Democracy*, January 2019.

[109] Strittmatter, *We Have Been Harmonised*, chap. New China, New World.

[110] Bao Pu is a political commentator and veteran human rights activist, is a publisher and editor of New Century Press in Hong Kong. "Bao Pu," Simon & Schuster, accessed October 11, 2021, https://www.simonandschuster.com/authors/Bao-Pu/66810478.

[111] Strittmatter, *We Have Been Harmonised*, chap. The Net.

Not only that: “That data is not just impressive in quantity, but thanks to China's unique technology ecosystem – an alternate universe of products and functions not seen anywhere else – that data is tailor-made for building profitable AI companies.”[112]. The CCP strategy is to build a domestic AI industry worth nearly US$150 billion in the next few years and to become the leading AI power by 2030. But, as we’ve seen, Xi Jinping’s ambitions go beyond the domestic realm: Speaking at the CCP Congress in October 2017, President Xi Jinping publicly outlined his plan to transform China into a “cyber superpower”, offering the country’s model of governance – including its management of the Internet – as “a new option for other countries”[113]. If no alternative model is presented, AI will globally develop increasingly through the guidelines and technological means of the “world's worst abuser of internet freedom for the sixth consecutive year”[114].

[112] Lee, *AI Superpowers*, 16.

[113] Sarah Cook, “China’s Cyber Superpower Strategy: Implementation, Internet Freedom Implications, and U.S. Responses,” Freedom House, September 28, 2018, https://freedomhouse.org/article/chinas-cyber-superpower-strategy-implementation-internet-freedom-implications-and-us.

[114] Freedom House, “Freedom on the Net 2020 - The Pandemic’s Digital Shadow” (Freedom House, 2020), https://freedomhouse.org/sites/default/files/2020-10/10122020_FOTN2020_Complete_Report_FINAL.pdf.

Part 3: The Future is Now

> All that is necessary for the triumph of an AI-driven, automation-based dystopia is that liberal democracy accept it as inevitable.
>
> Lanier & Weyl," AI is an Ideology, Not a Technology", 2020

Life is becoming increasingly digital. And that is a tendency that will prevail and increase. With the Internet of Things[115] and 5G, we will be ever-more permanently connected in ever-more dimensions. According to the Digital 2021: Global Overview Report*,* between January 2020 and January 2021, the number of social media users has grown by 490 million (13 per cent increase) and there were more 316 million internet users (7.3 per cent), which led to the increase of global internet penetration to 59,5 per cent. The amount of time spent on mobile devices and on the internet has also been steadily increasing over the past few years[116].

The coronavirus pandemic has accelerated this digitalization in multiple areas of our lives. Technology has significantly facilitated this forced transition and urgent adaptation. Almost all human activities, from work to school to social life, were transferred to the digital world dimension and that was vital to mitigate an already dramatic crisis. However, recently we have also seen the confirmation or exemplification of many concerns discussed throughout this dissertation. We have seen the spread of disinformation and conspiracy theories that deepen relativism, uncertainty, and fear – in

---

[115] "In the broadest sense, the term IoT encompasses everything connected to the internet, but it is increasingly being used to define objects that "talk" to each other. (…) By combining these connected devices with automated systems, it is possible to "gather information, analyse it and create an action" to help someone with a particular task, or learn from a process. (...) *In* Andersen, "Human Rights in the Age of Artificial Intelligence," 8.

The Internet of Things will turn many aspects of our environment into data-collection sensors and services.

[116] Data Reportal, "Digital 2021: Global Overview Report," DataReportal – Global Digital Insights, January 2021, https://datareportal.com/reports/digital-2021-global-overview-report.

what was denominated an "infodemic" by the World Health Organisation[117]. Multiple Artificial Intelligence surveillance tools have been developed and deployed in the name of 'public health' or safety[118]. Besides, the urgent need to combat COVID-19 did not allow for the pondered public debate that the deployed technologies required. According to Freedom House (Data Reportal, 2021), the pandemic hastened urged the expansion of biometric surveillance and algorithmic decision-making in healthcare, policing, education, finance, immigration, and commerce. The pandemic also correlated with a dramatic decline in global internet freedom, deteriorating for ten consecutive years[119]. It also contributed significantly to global democratic decline, as presented by Freedom House's 2021 report[120].

In regards to Artificial Intelligence and democratic sustainability, not acting is one of the most impactful actions. It gives space to the development of self-regulating systems, which pose serious threats to democratic health whenever the tech industry's

---

[117] "The Coronavirus disease (COVID-19) is the first pandemic in history in which technology and social media are being used on a massive scale to keep people safe, informed, productive and connected. At the same time, the technology we rely on to keep connected and informed is enabling and amplifying an infodemic that continues to undermine the global response and jeopardizes measures to control the pandemic.

An infodemic is an overabundance of information, both online and offline. It includes deliberate attempts to disseminate wrong information to undermine the public health response and advance alternative agendas of groups or individuals. Mis- and disinformation can be harmful to people's physical and mental health; increase stigmatization; threaten precious health gains; and lead to poor observance of public health measures, thus reducing their effectiveness and endangering countries' ability to stop the pandemic." *In* World Health Organization, "Managing the COVID-19 Infodemic: Promoting Healthy Behaviours and Mitigating the Harm from Misinformation and Disinformation," World Health Organization, September 23, 2020, https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation.

[118] "Facing such a stark health emergency, Covid-19 has prompted a sharp acceleration in efforts by governments to introduce more automated forms of decision making. It has provided an impetus and a rationale for authorities to try out new systems often without adequate debate, while offering opportunities for surveillance technology companies to pitch their products as tools for the common good." In Siddharth Venkataramakrishnan, "Algorithms and the Coronavirus Pandemic," *Financial Times*, January 10, 2021, https://www.ft.com/content/16f4ded0-e86b-4f77-8b05-67d555838941.

[119] Freedom House, "Freedom on the Net 2020 - The Pandemic's Digital Shadow."

[120] Freedom House, "Freedom in the World: Democracy under Siege," Freedom House, 2021, https://freedomhouse.org/report/freedom-world/2021/democracy-under-siege.

self-interest is separate from public interest. This contributes to the erosion of liberal principles, and the AI dominant horizon gradually looks more like the Chinese dystopian model. According to Jaron Lanier and Glen Weyl[121], it is surprising that leaders of the Western tech companies and governments have been so quick to accept this ideological model of *quasi* authoritarian quantitative power. This ideological model appears to perceive regulations and human rights as constraints on data collection and, therefore, constraints on competitiveness[122]. The authors claim this echoes a possible loss of faith in the philosophy and institutions of liberal democratic capitalism. This encourages wrong perception of why and how technologies can matter, how their development should be fostered, and how institutions can instill adaptation and compromise[123]. Even though this paradigm is not the only possibility, it can, in fact, be the prevailing one. If liberal democracies fail in defending and advancing their own principles, interests, and values in the digital era, digital authoritarianism will increasingly prevail, even if by default[124].

Liberal democracies must, therefore, rethink and reimagine the role of technology in human affairs, and the best way for democracies to block the rise of the digital authoritarianism paradigm is to present and develop a better model for deploying and managing AI.

---

[121] Jaron Lanier and Glen Weyl, "AI Is an Ideology, Not a Technology," *Wired*, March 15, 2020, https://www.wired.com/story/opinion-ai-is-an-ideology-not-a-technology/.

[122] "The usual narrative goes like this: Without the constraints on data collection that liberal democracies impose and with the capacity to centrally direct greater resource allocation, the Chinese will outstrip the West. AI is hungry for more and more data, but the West insists on privacy. This is a luxury we cannot afford, it is said, as whichever world power achieves superhuman intelligence via AI first is likely to become dominant. If you accept this narrative, the logic of the Chinese advantage is powerful. What if it's wrong? Perhaps the West's vulnerability stems not from our ideas about privacy, but from the idea of AI itself.", Lanier and Weyl.

[123] "Pluralistic visions of liberal democratic market societies will lose out to AI-driven ones unless we reimagine the role of technology in human affairs." Lanier and Weyl.

[124] Freedom House, "Freedom on the Net 2018 - The Rise of Digital Authoritarianism" (Freedom House, October 2018), https://freedomhouse.org/report/freedom-net/2018/rise-digital-authoritarianism.

Step Outside The Box

Authoritarian powers are stepping forward with cyber sovereignty agendas, displaying their own rules for governing their 'own' alternative Internet order. China's Great Firewall[125] is an extreme example of this tendency that entitles countries to define their internet rules, even if those rules infringe international norms or fundamental rights. Moreover, the digital realm is increasingly an 'arena' in the quest for economic and geopolitical influence, as the Chinese 'Digital Silk Road'[126] billion dollar investment mirrors.

Democracies themselves are subject to this gradual erosion in freedom of information. The digital realm, due to its own opaque nature, contributes to a normative blur, in which it is harder to separate a democratic environment from a dictatorial one[127]. What used to be clear democratic landmarks, such as freedom of expression, now might be exploited to further undemocratic dynamics.[128] And infringements of liberal freedoms can occur silently and opaquely, behind firewalls and hidden algorithms.

---

[125] The Golden Shield Project, often called the "great firewall of China", is an initiative managed by the Ministry of Public Security division of the Chinese government. As the nickname implies, the focus of this project is to monitor and censor what can and cannot be seen through an online network in China. This project started in 1998 and is still continually improving in restriction techniques through multiple methods. See "China's Great Firewall," Free speech vs Maintaining Social Cohesion: A Closer Look at Different Policies, accessed June 18, 2021, https://cs.stanford.edu/people/eroberts/cs181/projects/2010-11/FreeExpressionVsSocialCohesion/china_policy.html.

[126] "China's Digital Silk Road (DSR) was launched in 2015 as a component of Beijing's vast vision for global connectivity, the Belt and Road Initiative (BRI). Like the BRI, the DSR is not monolithic and involves many actors at all levels across the Chinese public and private sectors. (...) According to one estimate, by 2018, DSR-related investments in digital infrastructure projects outside of China had reached $79 billion. The DSR aims to improve digital connectivity in participating countries, with China as the main driver of the process." *In* Richard Ghiasy and Rajeshwari Krishnamurthy, "China's Digital Silk Road and the Global Digital Order," *The Diplomat*, April 23, 2021, https://thediplomat.com/2021/04/chinas-digital-silk-road-and-the-global-digital-order/.

[127] "It used to be pretty easy to define the difference between a democratic information environment and a dictatorial one. They had censorship and state-controlled media; we had freedom of speech, pluralism, and the "marketplace of ideas." Steven Feldstein and Peter Pomerantsev, "Democracy Dies in Disinformation," American Purpose, February 10, 2021, https://www.americanpurpose.com/articles/democracy-dies-in-disinformation/.

[128] As Tilly Kenyon writes, "dictators don't just censor by constraining the amount of media content: They exploit freedom of expression to flood the environment with so much disinformation that the truth is

If democracies cannot establish a common ground for an alternative framework, there is a risk that the democratic landscape, too, becomes part of this dynamic and becomes even more fractured. An inability to coordinate standards and values might lead to "splinternets"[129], a digital world divided between different approaches and increasingly protectionist tendencies.

Democracies must look inward to understand what to preserve and build upon. As Feldstein and Pomerantsev[130] point out, it is fundamental that the new framework is built with a broader mandate than simply countering China and Russia. Democracies must recognize the inner-sources of current vulnerabilities and risks of their digital realms. Protecting personal and human rights and striving for transparency and accountability in the digital sphere will ultimately strengthen democratic public life, distinguishing democracies from the authoritarian model, and more effectively countering that rival model. The absence of a framework for protecting fundamental norms and values is, as we have seen, contributing to the increasing vulnerability of liberal democracies. In fact, the "void" is filled by private companies and state bureaucracy which, as we have seen, may not align with the public interest.

Focus on *transparency* is an essential first step, not as an end in itself but as the basis for a digital landscape based on human rights. Transparency is necessary in what regards purpose, use, and functioning in developing and deploying digital tools and

---

swamped, a sort of censorship through noise. (…) Even in Russia and China, leaders don't just restrict communication but overload or flood information channels. Pluralism, meanwhile, displays a polarization so extreme that it destroys the possibility of a shared reality, let alone a "marketplace of ideas." Feldstein and Pomerantsev.

[129] "The splinternet is often referred to as 'internet balkanization', meaning the fragmentation of the worldwide web into smaller nationally administered internets." *In* Tilly Kenyon, "How Is the Splinternet Changing the World Wide Web? | Digital Transformation," *Technology*, accessed June 18, 2021, https://technologymagazine.com/digital-transformation/how-splinternet-changing-world-wide-web.

[130] Feldstein and Pomerantsev, "Democracy Dies in Disinformation."

algorithmic procedures[131]. This allows for accountability and explainability, contributing to the mitigation of the 'black box' dynamic's nefarious effects.

Interesting proposals of open and transparent but secure information systems are those of Taiwan and Estonia. In Taiwan's Digital Democracy, a civic tech community named *g0v* (pronounced 'Gov Zero'), which played an important role in the Sunflower protests[132], built the *vTawain* platform. *vTaiwan* is a digital platform with open-source software tools that enables policies to be discussed between citizens, civil society organizations, experts, and elected representatives via the website or even face-to-face meetings and hackathons. It fosters and encourages the use of what the Digital Minister, Audrey Tang, calls "collective intelligence of civil society". The debate produces consensus that the government can turn into laws and helps policymakers make decisions that gain legitimacy precisely through consultation and plural participation.

Estonia, a country that tied with Iceland for the best internet freedom score in 2018[133], developed *X-Road* platform, an open-source software and ecosystem solution that provides unified and secure data exchange between organizations and producing or

---

[131] EU's General Data Protection presents a clear definition of transparency in its **Recital 58** :

"The principle of transparency requires that any information addressed to the public or to the data subject be concise, easily accessible and easy to understand, and that clear and plain language and, additionally, where appropriate, visualisation be used. Such information could be provided in electronic form, for example, when addressed to the public, through a website. This is of particular relevance in situations where the proliferation of actors and the technological complexity of practice make it difficult for the data subject to know and understand whether, by whom and for what purpose personal data relating to him or her are being collected, such as in the case of online advertising" "Recital 58 - The Principle of Transparency," accessed June 29, 2021, https://gdpr-info.eu/recitals/no-58/.

[132] The Sunflower Movement, led by students and activists, arose when President Ma Ying-jeou's government attempted to establish a trade agreement with China - which claims Taiwan as its territory. For more than three weeks the Movement held and organized protests occupying government buildings, trying to avoid an agreement that they felt would give China too much leverage over Taiwanese economy. In the aftermath of this protests, the Government invited Sunflower activists to create a platform through which it might better communicate with Taiwan's youth. And this subsequently led to the creation of various platforms of civic participation and participatory data-governance and sharing.

[133] Freedom House, "Freedom on the Net 2018 - The Rise of Digital Authoritarianism."

consuming services. *X-Road* implements a set of standard features to support and facilitate data exchange and ensures confidentiality, integrity, and interoperability.[134]

Calls for *openness* and *dignity* are also part of Lanier and Weyl's "A Blueprint for a Better Digital Society"[135]. Their argument is that a "a market for data would restore dignity to data creators, who would become central to a dignified information economy"[136]: Essentially, they propose, "people will be paid for their data and will pay for services that require data from others"[137]. This might contribute to a positive cycle[138], where an "open market will become more aligned with an open society when the customer and the user are the same person."[139] According to the authors, an additional layer of organizations of intermediate size is required to bridge the gap between individuals and big tech platforms. They call them "mediators of individual data or MIDs"[140], which can not only "bring the power of collective bargaining to the people who are the sources of valuable data" but can also contribute to the healing of social structures that were broken and supplanted "by algorithms that target people for advertising"[141]: individuals might be

[134] Freedom House, "Freedom on the Net 2018 - The Rise of Digital Authoritarianism."

[135] Jaron Lanier and Glen Weyl, "A Blueprint for a Better Digital Society," *Harvard Business Review*, September 26, 2018, https://hbr.org/2018/09/a-blueprint-for-a-better-digital-society.

[136] Lanier and Weyl, 3.

[137] Lanier and Weyl, 4.

[138] "Individuals' attention will be guided by their self-defined interests rather than by manipulative platforms beholden to advertisers or other third parties. Platforms will receive higher-quality data with which to train their machine learning systems and thus will be able to earn greater revenue selling higher-quality services to businesses and individuals to boost their productivity. The quality of services will be judged and valued by users in a marketplace instead of by third parties who wish to influence users." Lanier and Weyl, 4.

[139] Lanier and Weyl, 4.

[140] Lanier and Weyl, 5.

[141] As we've seen in the part dedicated to the challenges of Big Tech business model, "these [algorithms] tend to corral individuals into divergent groups. Incentives to increase online "engagement" can then result in heightened social rifts as suspicions are raised about the "other." This tendency of the current network architecture is so prominent that it has become a favorite tool for information warfare; both rich and poor societies have been disrupted by malicious social media campaigns that emphasize and encourage societal divisions." Lanier and Weyl, 6.

part of multiple MIDs, developing complex digital identities and interests, "interests instead of managed, corralled identities that are ripe for targeting."[142] The authors claim that "[a]ny dignified future economy that relies heavily on information technology must value the people who add the data"[143], and so the argument for MIDs "flows from fundamental principles"[144]. According to them "[t]he influence of the internet on all aspects of human experience is so great that we must demand data dignity if we are to retain any dignity at all."[145] The digital environment must ensure and upheld this dignity, fundamental for a truly liberal order.

Francis Fukuyama presents a similar possible remedy for a more accountable and fair digital realm. He evokes James Madison's famous *Federalist 51* and its defense of the division of powers, a system of checks and balances. Fukuyama claims that these concerns, at the time mostly focused on state power, should apply "doubly to concentrations of private power"[146]. Like Lanier and Weyl, the author is also trying to find ways to "reduce the underlying power of today's internet platforms"[147]. He argues in favour of "middleware": "software that rides on top of a platform and affects the way in which users interact with the data that the platform carries"[148]. According to Fukuyama, "middleware could reduce the platforms' power by taking away their ability to curate content, giving this function to a wide variety of competitive firms"[149] – for instance, Lanier and Weyl's MIDs. Essentially, Fukuyama claims, it is vital to start

[142] Lanier and Weyl, 6.

[143] Lanier and Weyl, 17.

[144] Lanier and Weyl, 18.

[145] Lanier and Weyl, 18.

[146] Fukuyama, "Making the Internet Safe for Democracy," 39.

[147] Fukuyama, 40.

[148] Fukuyama, 42.

[149] Fukuyama, 42.

understanding what remedies to concentrated are politically and technologically realistic[150].

These types of *openness* and *transparency*, *fairness*, and *accountability* have to be the basis of a robust alternative digital model. Such a complex endeavor requires differing skills and collaborative work, bringing together software and product designers, engineers, ethicists, social scientists, and policymakers to draw on their respective expertise and integrate their knowledge to solve pressing problems. It must be sustained on a multilateral collaboration that embraces governments, technological companies, and civil society perspectives. How can democracies work together to build such a landscape, a liberal democratic digital space?

Step by Step

Firstly, and this dissertation aims to contribute to this assumption, it is necessary to understand something should be done. Democracies face challenges due to new technologies. Secondly, as the philosopher of technology Mark Coeckelbergh points, in the development of AI policy proposals there are several elements and steps that should be taken into account[151]. Moreover, there are various possible measures: "policy can mean regulation by means of laws and directives, say, legal regulation, but there are also other strategies that may or may not be connected to legal regulation, such as technological measures, codes of ethics, and education."[152]

The first step defined by Coeckelbergh is to justify the measures proposed. Looking back into the particular challenges and risks that we've dwelled into throughout this

[150] Fukuyama, 44.

[151] Mark Coeckelbergh, *AI Ethics*, The MIT Press Essential Knowledge Series (Cambridge, MA: The MIT Press, 2020).

[152] Coeckelbergh, 146.

dissertation, we can find multiple justifications for the need of certain regulations. For instance, trying to reduce biased algorithmic decision making is important due to its possible harm of fundamental human rights' principles[153]. Timing is pivotal, and that is actually one of the main challenges liberal democracies face when dealing with regulating AI: "in response to technology development, policy often comes too late, when the technology is already embedded in society."[154] In this process, it is also fundamental to define *who* should take action: the truth is, "many hands are involved in any technological action"[155]. As Coeckelbergh points out, the multiplicity of actors involved raises the question of "how to distribute responsibility for policy and change: is it mainly up to governments to take action, or should, for example, businesses and industry develop their own course of action to ensure ethical AI? When it comes to business, should one address only large corporations or also small and medium-sized businesses? And what is the role of individual (computer) scientists and engineers? What is the role of citizens?"[156] This issue if fundamental. The undefinition contributes to reckless or unaccountable development and deployment of AI technology.

However, even though there are still many loose ends, there is a "widely shared intuition that there is an urgency and importance in dealing with the ethical and societal challenges raised by AI"[157]. This has led to innumerous initiatives and policy documents that aim to provide some framework and guidance for present and future AI deployment.

[153] Coeckelbergh also presents this justification to a policy that aims to solve that particular issue. Coeckelbergh, 146.

[154] Coeckelbergh, 130.

[155] Coeckelbergh, 147 .

[156] Coeckelbergh, 131.

[157] Coeckelbergh, 132.

Many of these initiatives identify and are focused on ethical problems of AI, aiming to provide normative guidance to the development of a democratic technology framework.

Technological Anthropocentrism

The European Union has been taking an active role in the regulatory vanguard. According to the European Parliamentary Research Service's (EPRS) Report on the "European framework on ethical aspects of artificial intelligence, robotics and related technologies"[158], even though multiple private and public actors are debating and producing ethical guidelines, there is currently no comprehensive legal framework. That being the case, "the EU can profit from the absence of a competing global governance model and gain full 'first mover' advantages"[159], and consequently becoming a possible "global standard-setter in the area of artificial intelligence (AI) ethics"[160]. Moreover, EU action would "facilitate the adoption of EU standards globally and ensure that the development, uptake and diffusion of AI is based on the values, principles and rights protected in the EU."[161]

The uptake of a framework sustained on values, principles and rights of the EU could be important to mitigate and deal with the risks mentioned previously in this dissertation. One of the reasons that might be so is due to European keystone documents such as the Charter of Fundamental Rights of European Union (CFR), and what they uphold. The CFR, albeit not addressing directly the topic of AI, enshrines social and political vital rights and "important, fundamental guidance and legal obligations, which

[158] Tatjana Evas, "European Framework on Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies: European Added Value Assessment" (Brussels: European Parliamentary Research Service, September 2020).

[159] Evas.

[160] Evas.

[161] Evas.

inform the main analyses on AI."[162] As the EPRS points out, the Charter "emphasizes the human-centric nature of EU activities" in its own preamble[163].

Moreover the "need to protect fundamental rights in the light of scientific and technological developments"[164], while defining fundamental "general" rights which are relevant to the AI debate. Among these (and in consensus with the UNHRD) we can count the protection of human dignity (Article 1), the right to privacy and family life (Article 6), freedom of thought (Article 10), freedom of expression and information (Article 11), equality before the law (Article 20) and right to non-discrimination (Article 21). Furthermore, in its Article 8, it claims the right to the "Protection of personal data"[165].

The EU's concern with data protection[166] is mirrored, for instance, in the General Data Protection Regulation (GDPR), "the toughest privacy and security law in the

---

[162] Evas, 5.

[163] 'Conscious of its spiritual and moral heritage, the Union is founded on the indivisible, universal values of human dignity, freedom, equality and solidarity; it is based on the principles of democracy and the rule of law. It places the individual at the heart of its activities, by establishing the citizenship of the Union and by creating an area of freedom, security and justice.' "Charter of Fundamental Rights of the European Union," accessed June 29, 2021, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT.

[164] Evas, "European Framework on Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies: European Added Value Assessment," 5.

And, as stated in the CFR: 'To this end, it is necessary to strengthen the protection of fundamental rights in the light of changes in society, social progress and scientific and technological developments by making those rights more visible in a Charter.' "Charter of Fundamental Rights of the European Union."

[165] **Article 8:**

Protection of personal data

1. Everyone has the right to the protection of personal data concerning him or her.

2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.

3. Compliance with these rules shall be subject to control by an independent authority.

"Charter of Fundamental Rights of the European Union."

[166] "Some scholars argue that GDPR must be read narrowly as a legal instrument regulating strictly personal data, while others suggest a broader reading of GDPR and claim that GDPR incorporates ethical values and thus provides a normative, value-driven framework, encompassing, among other things, fundamental rights and principles. Hielke Hijmans and Charles D. Raab, for example, argue there is a close relationship between data protection and ethics: The fundamental right to data protection gives an individual a claim

world"[167]. The GDPR signals its commitment and firm stance on data privacy and security, impacting organizations globally, so long as they target or collect data related to people in the EU. Ultimately, it aims to "protect and empower all EU citizens with regard to data privacy"[168]. It upholds principles such as the right to be forgotten[169], the right to access[170], and the right "not to be subject to a decision based solely on automated processing"[171], among others.

According to the EPRS this strong stance on issues of privacy and data protection, "makes the EU's strategic approach to AI substantially different from the US one, which focuses on private-sector initiatives and is self-regulation driven, and the Chinese strategy, which prioritizes a government-led approach, with substantial coordination of private and public sectors."[172]

Another fundamental element of the European Union's AI agenda is its "human-centric" approach. This is a particular path that the EU has the potential to explore and

---

that her data is being processed in a fair manner. Other – moral – value notions behind data protection are human dignity and personal autonomy, which are notions with an obvious ethical dimension. In addition, ethical considerations play a role in the application of data protection law, including the GDPR". An interesting debate around the GDPR *in* Evas, "European Framework on Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies: European Added Value Assessment," 6.

[167] "What Is GDPR, the EU's New Data Protection Law?," GDPR.eu, November 7, 2018, https://gdpr.eu/what-is-gdpr/.

[168] Coeckelbergh, *AI Ethics*, 136.

[169] **Article 17**, which claims that in various circumstances "The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay (…)". "General Data Protection Regulation (GDPR) – Official Legal Text," General Data Protection Regulation (GDPR), accessed June 29, 2021, https://gdpr-info.eu/.

[170] **Article 15:** "The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data (...)".

[171] **Article 22: "**The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."

"General Data Protection Regulation (GDPR) – Official Legal Text."

[172] Evas, "European Framework on Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies: European Added Value Assessment," 8.

contribute to the shaping of a normative approach to AI. The EU is already developing this framework of an ethics-driven and trustworthy development of AI technology.

The AI HLEG, defines 'trustworthy AI' as lawful (laws and regulations are followed); ethical (ethical values and principles are obeyed), and robust (from technical and social perspectives no harm is created). The AI HLEG definition and ethical understanding is based on four ethical principles:

1) Respect for human autonomy – only attributable to human beings and a central aspect of dignity and agency;
2) Prevention of harm – avoidance of harmful practices and their aggravation;
3) Fairness – ensuring equality and justice, absence of unfair bias, discrimination and stigmatization; freedom of choice and proportionality between means and ends;
4) Explicability – transparency of processes, open communication of capabilities and purpose and explainable decisions for those directly and indirectly affected.[173]

The EU has also been developing its Digital Strategy. As part of its Digital Services Act package, it recently launched two regulations, the Digital Services Act (DSA) and the Digital Market Act (DMA).

The DSA will introduce new obligations on platforms to reveal information and data to regulators about how their algorithms work, how decisions are made to remove content, and how adverts are targeted at users. Many of its provisions only apply to platforms with more than 45 million users, a threshold surpassed by several services,

[173] Evas, "European Framework on Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies: European Added Value Assessment."

including Facebook, YouTube, Twitter, and TikTok (Perrigo, 2021). It aims at transparency and, consequently, accountability - one of the riskiest absences of the current framework. The DMA aims to give smaller companies a greater ability to compete with big tech platforms, which some European lawmakers have long thought of as monopolistic entities. This regulation aims at diluting some of the oligopolistic unbalanced advantages: it forces "gatekeeper"[174] companies (such as Amazon, Google, Facebook) to act fairly by not using competitors' data to disadvantage them; by enforcing interoperability, allowing users to take their data elsewhere and still interact with their services; and by not treating their own services more favorably than competitors that use their platform. (Perrigo, 2021). It aims at a fairer playing field to foster competition and innovation.

Regulatory efforts are not enough, especially in the EU, where many of its regulations are not mandatory or enforceable, and frequently in a strict top-down approach. However, the EU's strategy is distinctive in emphasizing the trustworthy and secure development of "human-centered" AI in full respect of citizens' rights, freedoms, and values, which may lay the basis for a change of the digital panorama. Notwithstanding, it remains to be seen whether these concepts will actually be effective and practical or merely empty slogans. Moreover, it is essential to avoid hollow rhetorical declarations or overregulation that could transform this framework into a bureaucratic apparatus and hamper innovation, commercialization, and uptake.

---

[174] The European Parliament writes: "Recent reports and studies have shown how a few large platforms have become online gatekeepers, controlling key channels of distribution because of a variety of factors, including strong network effects in the digital environment (i.e. users are more likely to value and choose platforms with a large user base), their intermediary role (i.e. between sellers and customers), and their ability to access and accumulate large amounts of data (e.g. users' personal and non- personal data and competitors' sales data). These characteristics may provide online gatekeepers with a dominant position and market power detrimental to fair competition." European Parliament, "Regulating Digital Gatekeepers - Background on the Future Digital Markets Act," 2.

Policies and Possibilities

The EU might be well-positioned to help establish best practices and standards to steer AI's present and future. However, it is fundamental that it does not lead to increased fragmentation and a real barrier in the Atlantic Alliance. Both the EU and the US should be available to harmonize and develop a consistent approach that enables building a robust democratic AI development that increases the societal value and guarantees citizens and countries' rights and security[175]. It is also fundamental to engage with a broader scope of democracies. Countries like Japan are also moving towards this direction, having developed in 2019 the 'Social Principles of Human Centric AI'[176]. In 2019, it also fostered the G20 signing of the Osaka Track[177], which aims to intensify

---

[175] In June 2021, the two sides showed positive signs: the first EU-US summit since 2014 was held, as well as and the first visit by a US President to the EU institutions since 2017. One of the main points of this renewed cooperation, as defined in the final statement, "Towards a Renewed Transatlantic partnership", is that of strengthen technological cooperation: "We resolve to stand together to protect our businesses and workers from unfair trade practices, in particular those posed by non- market economies that are undermining the world trading system." "EU-US Summit, Brussels, 15 June 2021," accessed June 29, 2021, https://www.consilium.europa.eu/en/meetings/international-summit/2021/06/15/.

[176] In its 2019 document, Japan's Integrated Innovation Strategy Promotion Council establishes basic principles each stakeholder should keep in mind in order for AI to be accepted and properly used by society: the human-centric principle, which claims that "the utilization of AI must not infringe upon the fundamental human rights" and that "AI should be developed, utilized, and implemented in society to expand the abilities of people and allow diverse people to pursue their own well-being"; The Principle of Education/Literacy, which claims that "policy makers and managers of businesses involved in AI must have an accurate understanding of AI, knowledge and ethics permitting appropriate use of AI in society and "AI users should have a general understanding of AI and should acquire sufficient education to use it properly"; The Principle of Privacy Protection, which claims that "any AI using personal data and any service solutions that use AI, including use by the government, do not infringe on a person's individual freedom, dignity or equality"; The Principle of Ensuring Security, which claims "the use of AI poses a new set of risks to security. Society should always be aware of the balance between the benefits and risks, and endeavor to improve social safety and sustainability as a whole."; the Principle of Fair Competition which claims "A fair competitive environment must be maintained in order to create new businesses and services, to maintain sustainable economic growth, and to present solutions to social challenges"; the Principle of Fairness, Accountability, and Transparency which claims "it is necessary to ensure fairness and transparency in decision-making, appropriate accountability for the results, and trust in the technology," and finally the Principle of Innovation, which claims that "[to] aim for continuous innovation that advances as people evolve together with AI development, we should transcend boundaries such as national borders, industries, academia, governments, race, gender, nationality, age, political convictions and religion. We should promote total globalization, diversification, and industry-academia-government cooperation (...)." https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf

[177] Osaka Declaration on Digital Economy, signed by the leaders of Argentina, Australia, Brazil, Canada, China, the European Union, France, Germany, Italy, Japan, Mexico, Republic of Korea, Russian

international rule-making efforts on the digital economy while promoting the protection of intellectual property, personal data, and cybersecurity.

However, the sight must go beyond the transatlantic and embrace other nations with liberal democratic and human-rights-based systems. Moreover, as Feldstein and Pomerantsev (2021) point out, it would be essential to go to human-rights-based democracies but also try to encompass 'swing states' like India or Brazil, persuade states in the middle to align with the liberal democratic agenda, so the digital authoritarian model does not co-opt them[178] – which is specially pressing given the authoritarian "export" of their AI technology and agenda[179].

A multilateral approach also encompasses international forums such as the OECD and its AI Policy Observatory[180] for sharing good practices, the United Nations and the

Federation, Saudi Arabia, Turkey, United Kingdom, United States, Spain, Chile, Netherlands, Senegal, Singapore, Thailand, and Viet Nam and which affirms "the importance of promoting national and international policy discussions for harnessing the full potential of data and digital economy to foster innovation, so that we can keep pace with the fast-growing digital economy and maximize the benefits of digitalization and emerging technologies." "Osaka Declaration on Digital Economy," accessed June 1, 2021, https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/special_event/en/special_event_01.pdf .

[178] As we have seen, the digital realm can be an area of authoritarian attractiveness. In its 2018 "The Rise of Digital Authoritarianism", Freedom House alerts to the fact that "a cohort of countries is moving toward digital authoritarianism by embracing the Chinese model of extensive censorship and automated surveillance systems". In fact, "Beijing took steps to propagate its model abroad by conducting large-scale trainings of foreign officials, providing technology to authoritarian governments, and demanding that international companies abide by its content regulations even when operating outside of China" Freedom House, "Freedom on the Net 2018 - The Rise of Digital Authoritarianism."

[179] "Overall, China is making a sustained push for leadership and primacy in AI.A growing consensus singles out China as a global driver of "authoritarian tech." (…) China is exporting surveillance tech to liberal democracies as much as it is targeting authoritarian markets." Feldstein, "The Global Expansion of AI Surveillance," 13–14.

[180] Whose aim is to "Shape and share public policies for responsible, trustworthy and beneficial AI" *in* "The OECD Artificial Intelligence Policy Observatory - OECD.AI," accessed June 25, 2021, https://www.oecd.ai/.

High-Level Panel on Digital Cooperation[181], and the G7 and its Global Partnership on AI[182].

Civil society – one of the vital pillars of liberal democracies[183] – must have a central place in this landscape. Civil society organizations and movements, such as Algorithm Watch[184], Access Now[185] and the Center for Humane Technology[186] (a few cases which contributed to the research on this dissertation), have been fundamental in demanding a

[181] "The High-level Panel on Digital Cooperation was convened by the UN Secretary-General to advance proposals to strengthen cooperation in the digital space among Governments, the private sector, civil society, international organizations, technical and academic communities and other relevant stakeholders.

The Panel is expected to raise awareness about the transformative impact of digital technologies across society and the economy, and contribute to the broader public debate on how to ensure a safe and inclusive digital future for all, taking into account relevant human rights norms." *In* United Nations, "Secretary-General's High-Level Panel on Digital Cooperation," United Nations (United Nations), accessed June 25, 2021, https://www.un.org/en/civil-society/secretary-general%E2%80%99s-high-level-panel-digital-cooperation.

[182] Launched in June 2020, GPAI is the fruition of an idea developed within the G7, under the Canadian and French presidencies. GPAI's 15 founding members are Australia, Canada, France, Germany, India, Italy, Japan, Mexico, New Zealand, the Republic of Korea, Singapore, Slovenia, the United Kingdom, the United States and the European Union. They were joined by Brazil, the Netherlands, Poland and Spain in December 2020.

GPAI provides a mechanism for sharing multidisciplinary research and identifying key issues among AI practitioners, facilitating international collaboration and promoting the adoption of trustworthy AI.

"Global Partnership on Artificial Intelligence - GPAI," accessed June 25, 2021, https://gpai.ai/.

[183] As beautifully pointed out by Alexis de Tocqueville, in his great work "Democracy in America": "There is nothing, in my opinion, that merits our attention more than the intellectual and moral associations of America. The political and in- dustrial associations of the Americans easily fall within our grasp, but the others escape us; and, if we discover them, we understand them badly, because we have hardly ever seen anything analogous. You must recognize, however, that the intellectual and moral associations are as necessary as the political and industrial ones to the American people, and perhaps more. In democratic countries, the science of association is the mother science; the progress of all the others depends on the progress of the former." Alexis de Tocqueville, *Democracy in America*, ed. Eduardo Nolla, trans. James T. Schleifer, English ed (Indianapolis: Liberty Fund, 2012), 902.

[184] "AlgorithmWatch is a non-profit research and advocacy organization committed to evaluating and shedding light on algorithmic decision-making processes that have a social relevance, meaning they are used either to predict or prescribe human action or to make decisions automatically." Algorithm Watch, "What We Do," *AlgorithmWatch* (blog), accessed June 29, 2021, https://algorithmwatch.org/en/what-we-do.

[185] Access Now "defends and extends the digital rights of users at risk around the world" in different areas such as Policy, Advocacy, Legal and grants. "About Us," *Access Now* (blog), accessed September 21, 2021, https://www.accessnow.org/about-us/.

[186] "Our mission is to drive a comprehensive shift toward humane technology that supports our well-being, democracy, and shared information environment."'"Who We Are," accessed September 21, 2021, https://www.humanetech.com/who-we-are#our-story.

more accountable and transparent AI conception and deployment and are essential stakeholders in the development of a fair and accountable system. They can also be great beneficiaries of a freer digital environment. Civil society flourishes in freedom, and its vitality is a timeless sign of a healthy democracy.

Moreover, civil society organizations may contribute to a fundamental element in a more democratic AI: literacy and awareness. They can contribute to guarantee that the increasing complexity of technology does not mean incomprehensibility[187].

Last but certainly not least, a human-centric approach is fundamentally focused on using AI to empower the individual. This approach is fundamentally focused on protecting citizens' civil liberties and rights. Furthermore, it is fundamental to develop a structure that empowers citizens both on their individual and social dimension. One that enables them to understand and follow the learning, reasoning, and planning of AI systems, but also to interact and cooperate with and within the system contributing therefore to its openness, legitimacy, and progress.

If democracy is to survive and strive in the digital age, technology companies, governments, and civil society must work together to find real solutions. Multilateral and cross-sectoral coordination is required. Data protection, algorithmic transparency, and privacy should be strengthened and assured by both the government and corporations. A free and responsible digital realm can be the antidote to the darkness, fear, and repression of digital authoritarianism. The health of the world's democracies depends on it.

---

[187] That is, in fact, one of the main mottos of AlgorithmWatch. In its manifesto, Algorithm Watch claims "The fact that most ADM procedures are black boxes to the people affected by them is not a law of nature. It must end.". Algorithm Watch, "What We Do."

Conclusion

> Surviving the present means rebuilding the legitimate authority of the institutions of liberal democracy, while resisting those powers that aspire to make nondemocratic institutions central.
>
> Fukuyama, "30 Years of World Politics: What Has Changed?", 2020, (p. 21)

Artificial Intelligence has been called by some "the new electricity", due to its power to transform virtually every industry and in fact almost every human action, to the broadness of its impact and utility. The complexity of its impact is, however, very different.

AI is not simply a "tool" that is used and leaves everything as it is. AI shapes and alters the realms in which it is deployed. It impacts minds and actions of individuals, in a political system in which these individuals have the ultimate decisional power. It alters the ways institutions function. It changes their human dimension, having influence to either increase or decrease their fairness and trustability. It impacts, therefore, the way core institutions are perceived and perceive themselves. AI fundamentally changes key concepts of any political regime: safety, efficiency, liberty, well-being. Its ubiquitous nature can be both a dark cloud or a ray of light. It all depends on what we do now.

For liberal democracies to compete in the technological arena without crumbling, they must struggle to build a new narrative to sustain their idea of technological progress. A narrative that embraces liberal checks and human rights as promoters of a more prosperous networked society and propellers of progress. Even though the tools (or weapons) that democracies and authoritarians have access to are the same, their uses can be radically different. These differences do not have to be perceived *weaknesses* in what regards international competition, any more than free speech is a "weakness" in ideological competition. The same technological changes that are currently harming

fundamental rights and eroding liberal institutions can be used to protect and strengthen them, empowering civil society and advancing pluralism and liberty.

Liberal democracies must focus on building a social, legal, and regulatory framework in which artificial intelligence deployment is a motor of social value. Innovation must be shaped by democratic ideals, not the other way around. The survival of the democratic ideal is and will be fundamentally intertwined and dependent on the democratic practice. The latter, on its hand, must be constantly and relentlessly embedded with the essence of the democratic ideal.

Once again in History, and maybe more than ever, liberal democracies have to hold on to their essence and use their own principles as an impulse towards progress and prosperity.

BIBLIOGRAPHY

Access Now. “About Us.” Accessed September 21, 2021. https://www.accessnow.org/about-us/.

AI HLEG. “Ethics Guidelines for Trustworthy AI.” Ebook. Bruxelles: European Comission, April 8, 2019. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

Algorithm Watch. “Automating Society Report 2020.” Algorithm Watch, October 2020. https://automatingsociety.algorithmwatch.org.

———. “What We Do.” *AlgorithmWatch* (blog). Accessed June 29, 2021. https://algorithmwatch.org/en/what-we-do.

Allen-Ebrahimian, Bethany. “The Man Who Nailed Jello to the Wall.” Foreign Policy, June 29, 2016. https://foreignpolicy.com/2016/06/29/the-man-who-nailed-jello-to-the-wall-lu-wei-china-internet-czar-learns-how-to-tame-the-web/.

Andersen, Lindsey. “Human Rights in the Age of Artificial Intelligence.” Access Now, November 2018. https://www.accessnow.org.

Andrews, Evan. “Who Were the Luddites?” HISTORY, June 26, 2019. https://www.history.com/news/who-were-the-luddites.

Simon & Schuster. “Bao Pu.” Accessed October 11, 2021. https://www.simonandschuster.com/authors/Bao-Pu/66810478.

Bucher, Taina. *If...Then: Algorithmic Power and Politics*. New York: Oxford University Press, 2018.

Cadwalladr, Carole, and Emma Graham-Harrison. “Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach.” *The Guardian*, March 17, 2018, sec. News. http://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

Carr, Nicholas G. *The Glass Cage: Automation and Us*. New York: W. W. Norton & Company, Inc., 2014. http://ebook.3m.com/library/BCPL-document_id-hugsyg9.

Castelluccia, Claude, and Daniel Le Métayer. “Understanding Algorithmic Decision-Making: Opportunities and Challenges.” European Parliamentary Research Service, March 2019.

“Charter of Fundamental Rights of the European Union.” Accessed June 29, 2021. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT.

Free speech vs Maintaining Social Cohesion: A Closer Look at Different Policies. “China’s Great Firewall.” Accessed June 18, 2021. https://cs.stanford.edu/people/eroberts/cs181/projects/2010-11/FreeExpressionVsSocialCohesion/china_policy.html.

Chinese State Council. “Planning Outline for the Construction of a Social Credit System (2014-2020).” *China Copyright and Media* (blog), June 14, 2014. https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/.

Christian, Brian. *The Alignment Problem: Machine Learning and Human Values*. First edition. New York, NY: W.W. Norton & Company, 2020.

Coeckelbergh, Mark. *AI Ethics*. The MIT Press Essential Knowledge Series. Cambridge, MA: The MIT Press, 2020.

Committee of Experts on Internet Intermediaries (MSI-NET). “Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications.” Council of Europe, March 2018.

Conger, Katy. “Google Removes ‘Don’t Be Evil’ Clause From Its Code of Conduct.” Gizmodo, May 18, 2018. https://gizmodo.com/google-removes-nearly-all-mentions-of-dont-be-evil-from-1826153393.

Cook, Sarah. “China’s Cyber Superpower Strategy: Implementation, Internet Freedom Implications, and U.S. Responses.” Freedom House, September 28, 2018. https://freedomhouse.org/article/chinas-cyber-superpower-strategy-implementation-internet-freedom-implications-and-us.

Crawford, Kate. *Atlas of Ai: Power, Politics, and the Planetary Costs of Artificial*

Intelligence. New Haven: Yale University Press, 2021.

Crawford, Matthew. "Algorithmic Governance and Political Legitimacy." *American Affairs* III, no. 2 (Summer 2019). https://americanaffairsjournal.org/2019/05/algorithmic-governance-and-political-legitimacy/.

Data Reportal. "Digital 2021: Global Overview Report." DataReportal – Global Digital Insights, January 2021. https://datareportal.com/reports/digital-2021-global-overview-report.

David, Marie. "AI and the Illusion of Human-Algorithm Complementarity." *Social Research: An International Quarterly* 86, no. 4 (2019): 887–908.

Dyson, George, and Inc. *Turing's Cathedral*. New York: Pantheon Books, 2012. http://api.overdrive.com/v1/collections/v1L2BMAAAAM0GAAA19/products/f860601c-e462-49ef-b462-99f36221801c.

THE ETHICS CENTRE. "Ethics Explainer: The Panopticon - What Is the Panopticon Effect?," July 18, 2017. https://ethics.org.au/ethics-explainer-panopticon-what-is-the-panopticon-effect/.

European Parliament. "Regulating Digital Gatekeepers - Background on the Future Digital Markets Act," December 8, 2020.

"EU-US Summit, Brussels, 15 June 2021." Accessed June 29, 2021. https://www.consilium.europa.eu/en/meetings/international-summit/2021/06/15/.

Evas, Tatjana. "European Framework on Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies: European Added Value Assessment." Brussels: European Parliamentary Research Service, September 2020.

"Everything You Need to Know about Trade Secrets | IPTSE," October 19, 2020. https://iptse.com/understanding-trade-secrets/.

Feldstein, Steven. "The Global Expansion of AI Surveillance." Carnegie Endowment for International Peace, September 2019. https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847.

———. “The Road to Digital Unfreedom: How Artificial Intelligence Is Reshaping Repression.” *Journal of Democracy*, January 2019.

Feldstein, Steven, and Peter Pomerantsev. “Democracy Dies in Disinformation.” American Purpose, February 10, 2021. https://www.americanpurpose.com/articles/democracy-dies-in-disinformation/.

Florin, Marie-Valentine, and Kujtese Bejtullahu-Michalopoulos. “The Governance of Decision-Making Algorithms.” EPFL International Risk Governance Center, November 2018.

FRA. “Facial Recognition Technology: Fundamental Rights Considerations in the Context of Law Enforcement.” Vienna – Austria: FRA – EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 2019.

Frank Pasquale. *New Laws of Robotics*. Harvard University Press, 2020.

Freedom House. “Freedom in the World: Democracy under Siege.” Freedom House, 2021. https://freedomhouse.org/report/freedom-world/2021/democracy-under-siege.

———. “Freedom on the Net 2018 - The Rise of Digital Authoritarianism.” Freedom House, October 2018. https://freedomhouse.org/report/freedom-net/2018/rise-digital-authoritarianism.

———. “Freedom on the Net 2020 - The Pandemic’s Digital Shadow.” Freedom House, 2020. https://freedomhouse.org/sites/default/files/2020-10/10122020_FOTN2020_Complete_Report_FINAL.pdf.

Fry, Hannah. *Hello World*. New York: W.W. Norton & Company, 2018. https://api.overdrive.com/v1/collections/v1L1BcAAAAA2A/products/3fd6079f-e6dc-4e76-9900-ef1f235aae45.

Fukuyama, Francis. “Making the Internet Safe for Democracy.” *Journal of Democracy* 32, no. 2 (April 2021): 37–44.

Fukuyama, Francis, Barak Richman, and Ashish Goel. “How to Save Democracy From Technology | Foreign Affairs.” *Foreign Affairs*, February 2021. https://www.foreignaffairs.com/articles/united-states/2020-11-24/fukuyama-how-save-

democracy-technology.

General Data Protection Regulation (GDPR). “General Data Protection Regulation (GDPR) – Official Legal Text.” Accessed June 29, 2021. https://gdpr-info.eu/.

“Global Partnership on Artificial Intelligence - GPAI.” Accessed June 25, 2021. https://gpai.ai/.

Hoffman, Samantha. “Engineering Global Consent: The Chinese Communist Party’s Data-Driven Power Expansion.” Policy Brief. Australian Strategic Policy Institute, October 2019.

———. “Programming China: The Communist Party’s Autonomic Approach to Managing State Security.” PhD Thesis, University of Nottingham, 2017.

———. “Social Credit: Technology-Enhanced Authoritarian Control with Global Consequences.” Policy Brief. Australian Strategy Policy Institute, June 28, 2018. https://www.aspi.org.au/report/social-credit.

Kearns, Michael, and Aaron Roth. *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. New York: Oxford University Press, 2020.

Kenyon, Tilly. “How Is the Splinternet Changing the World Wide Web? | Digital Transformation.” Technology. Accessed June 18, 2021. https://technologymagazine.com/digital-transformation/how-splinternet-changing-world-wide-web.

Lanier, Jaron, and Glen Weyl. “A Blueprint for a Better Digital Society.” *Harvard Business Review*, September 26, 2018. https://hbr.org/2018/09/a-blueprint-for-a-better-digital-society.

———. “AI Is an Ideology, Not a Technology.” *Wired*, March 15, 2020. https://www.wired.com/story/opinion-ai-is-an-ideology-not-a-technology/.

Lee, Kai-Fu. *AI Superpowers: China, Silicon Valley, and the New World Order*. Boston: Houghton Mifflin Harcourt, 2018.

Lum, Kristian, and William Isaac. “To Predict and Serve?” *Significance* 13, no. 5

(October 2016): 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x.

Nations, United. “Secretary-General’s High-Level Panel on Digital Cooperation.” United Nations. United Nations. Accessed June 25, 2021. https://www.un.org/en/civil-society/secretary-general%E2%80%99s-high-level-panel-digital-cooperation.

“Osaka Declaration on Digital Economy.” Accessed June 1, 2021. https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/special_event/en/special_event_01.pdf.

Pariser, Eli. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. New York: Penguin Books, 2014. http://rbdigital.oneclickdigital.com.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press, 2015.

“Recital 58 - The Principle of Transparency.” Accessed June 29, 2021. https://gdpr-info.eu/recitals/no-58/.

Richard Ghiasy, and Rajeshwari Krishnamurthy. “China’s Digital Silk Road and the Global Digital Order.” *The Diplomat*, April 23, 2021. https://thediplomat.com/2021/04/chinas-digital-silk-road-and-the-global-digital-order/.

Scheiber, Noam, and Mike Isaac. “Facebook Halts Ad Targeting Cited in Bias Complaints.” *The New York Times*, March 19, 2019. https://www.nytimes.com/2019/03/19/technology/facebook-discrimination-ads.html.

Smith, Brad, and Carol Ann Browne. *Tools and Weapons: The Promise and the Peril of the Digital Age*. New York: Penguin Press, 2019.

Southerland, Vincent. “With AI and Criminal Justice, the Devil Is in the Data.” American Civil Liberties Union, April 9, 2018. https://www.aclu.org/issues/privacy-technology/surveillance-technologies/ai-and-criminal-justice-devil-data.

Strittmatter, Kai. *We Have Been Harmonised: Life in China’s Surveillance State*. United Kingdom: Old Street Publishing, 2019. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk

&AN=2121867.

Temperton, James. “Facebook’s Australia News Ban Is the Best Decision It’s Ever Made.” *Wired UK*. Accessed June 29, 2021. https://www.wired.co.uk/article/facebook-australia-rupert-murdoch.

“The OECD Artificial Intelligence Policy Observatory - OECD.AI.” Accessed June 25, 2021. https://www.oecd.ai/.

Tocqueville, Alexis de. *Democracy in America*. Edited by Eduardo Nolla. Translated by James T. Schleifer. English ed. Indianapolis: Liberty Fund, 2012.

Toulson, Lauren. “Why Algorithms Erode Democracy. The Way You Navigate the Internet Is… | by Lauren Toulson | CARRE4 | Medium,” February 23, 2021. https://medium.com/carre4/why-algorithms-erode-democracy-d1fcdeab63a5.

United Nations. “Universal Declaration of Human Rights,” 1948. https://www.un.org/en/about-us/universal-declaration-of-human-rights.

Ünver, H. Akın. “Artificial Intelligence, Authoritarianism and the Future of Political Systems.” Cyber Governance and Digital Democracy 2018/9. Centre for Economics and Foreign Policy Studies, July 2018.

Venkataramakrishnan, Siddharth. “Algorithms and the Coronavirus Pandemic.” *Financial Times*, January 10, 2021. https://www.ft.com/content/16f4ded0-e86b-4f77-8b05-67d555838941.

West, Darrell M., and John R. Allen. *Turning Point: Policymaking in the Era of Artificial Intelligence*. Washington: Brookings Institution Press, 2020.

GDPR.eu. “What Is GDPR, the EU’s New Data Protection Law?,” November 7, 2018. https://gdpr.eu/what-is-gdpr/.

“Who We Are.” Accessed September 21, 2021. https://www.humanetech.com/who-we-are#our-story.

American Civil Liberties Union. “Will Artificial Intelligence Make Us Less Free?” Accessed June 14, 2021. https://www.aclu.org/issues/privacy-technology/will-artificial-

intelligence-make-us-less-free.

World Health Organization. “Managing the COVID-19 Infodemic: Promoting Healthy Behaviours and Mitigating the Harm from Misinformation and Disinformation.” World Health Organization, September 23, 2020. https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation.

Xiao, Qiang. “The Road to Digital Unfreedom: President Xi’s Surveillance State.” *Journal of Democracy* 30, no. 1 (January 2019): 53–67.

X-Road® Data Exchange Layer. “X-Road® Technology Overview.” Accessed March 24, 2021. https://x-road.global/x-road-technology-overview.

Zuboff, Shoshana. “Big Other: Surveillance Capitalism and the Prospects of an Information Civilization.” *Journal of Information Technology*, 2015.

———. “Opinion | Facebook and the Surveillance Society: The Other Coup - The New York Times.” New York Times, January 29, 2021. https://www.nytimes.com/2021/01/29/opinion/sunday/facebook-surveillance-society-technology.html.