██ HEINRICH BÖLL STIFTUNG
TEL AVIV

## Artificial Intelligence and Democracy

### Backgrounder

This backgrounder introduces readers to the research literature on AI's impact on democracy. It surveys literature in three distinct areas: AI and the democratic public sphere, the impact of AI on election campaigns, and the importance and accountability of automated decision-making systems in public services.



Artificial Intelligence and Democracy   licence infos

### Why is this topic important?

Over the course of the last decade, there has been an enormous interest in artificial intelligence (AI). Due to some breakthroughs in machine learning and new ways to gather and analyze large amounts of data, expectations abound that AI will be implemented in more and more everyday technologies and reshape societies in far-reaching and revolutionary ways. The massive interest in AI from a technological and economical perspective has been quickly followed by an intense discussion on its societal impact, especially on problems of fairness and accountability. Only recently has this sociological critique been complemented by a discourse on the effects that AI might have on politics, especially democratic politics and procedures. The discourse on AI and democracy has so far mostly been conducted in popular science books and opinion pieces, for instance by Jamie Bartlett or Yuval Noah Harari. More elaborate journal articles with a foundation in democratic theory or empirical analysis are still rare but their numbers are increasing and the debate is set to accelerate further once the impact of AI technologies becomes even more visible. Maybe due to the choice of outlets, the tone set in the debate so far has been rather alarmist. The discourse is dominated by rallying cries, which portray AI as a new and very serious threat to democracy although there is also a much smaller, segment of the literature that regards AI as a cure for the shortcomings of representative will formation. This "risk-versus-opportunity" framing resembles the frames applied in the older discourse on digitalization and democracy.

Nevertheless, a closer look reveals new and different ways in which AI is reshaping democratic politics. This is the focus of the present backgrounder.

After a brief explanation of how AI technologies work and transform societies, this backgrounder introduces readers to the research literature on AI's impact on democracy. It surveys literature in three distinct areas: AI and the democratic public sphere, the impact of AI on election campaigns, and the importance and accountability of automated decision-making systems in public services. In each strand of the literature, the text provides assessments of current trajectories and discusses to what extent democracy, as we know and practice it, is affected by the development and deployment of AI applications. Finally, the backgrounder will take a look at recent attempts to regulate AI and analyze how seriously the political dimension of AI is treated in regulatory discourses.

### AI and Democracy: Understanding the Socio-technical Background

As a field of research, AI was established long before our current debates took off. Its origins can be traced back at least to the mid-1950s. At that time, expectations were high that machines would soon be able to think and act like humans, but the predicted progress never materialized and a long so-called AI winter followed. During that period, AI mostly disappeared from public discourse. But although it was not visible, progress was made and many of the conceptual tools and algorithmic techniques on which our current expectations are based were developed or fine-tuned. At the same time, digitalization made enormous inroads into society, preparing the ground for AI's rise. Two factors were decisive: The availability of enormous computing power and the ever-expanding collection of data. New modes of communications, new technologies to sense, collect and store data, and strong economic incentives to gather data necessitated developing new ways of analyzing data. From 2010 onwards, AI took off as the now-dominant machine-learning methods proved their worth in a series of spectacular successes in fields such as speech and image recognition, and a machine surpassed humans in playing the board game Go, an achievement long thought to be impossible.

Today's deep learning techniques are characterized by an inductive approach. While in the early days AI development often entailed complex deductive classification and reasoning, current approaches work by analyzing large data sets, thereby generating or adapting decision rules in order to allow for optimization of pre-defined criteria. AI systems have in many domains become highly effective in detecting and categorizing patterns and have developed to allow for adaptation to new developments or patterns while running. These capabilities are best understood as learning processes; their sheer scope and complexity makes them superior to human analysis in some cases (the optimization of complex supply chains is a good example). Still, it is important to understand current approaches to AI as narrow or weak AI. Their narrowness lies in their non-transmissibility, i.e. AI must be trained to perform a certain kind of task and cannot apply itself or be applied to other problems without adaptive steps.[1]

In what follows, this text limits the discussion to societal developments that can be expected to spring from applications of narrow AI and will focus mainly on the next five to ten years. An underlying assumption is that, during that period, the main thrust of the development will come from narrow AI's widespread deployment in society. The use of AI techniques will be much more common, much more sophisticated and much more seamless than today – but it will not be profoundly different. Hence, we can already extrapolate the direction we are heading towards, even though we have many possibilities for steering and shaping the development, since it is as much dependent on political, societal and economical frames as on technological developments.

## AI and Democracy: Assessing the Most Important Developments

In what follows, this text will focus on democracy as an institutional arrangement that aims to bring about collective self-governance. Representation is key to democracy, and if we wish to have democratic representation, we need to have procedural arrangements that support the free and reflexive transmission of citizens' voices and preferences to the political institutions ("public will formation"), and institutions that can be held accountable even if no actual elections are taking place. Democracy is not mainly about outputs or effectiveness and therefore should not be equated with a just or fail-proof society (this false equivalency is sometimes the premise in AI debates, when results are characterized as good or just, and therefore democratic). On the other hand, democracy should also not be defined narrowly by electoral procedures only, since public will formation and the way power is exercised are also important for the complex appraisal of democracy.

Building on this understanding of democracy, one can roughly distinguish three main foci in the emerging debate on how AI transforms societal or political practices in a manner that directly affects the conduct of democracy: in the public sphere, with regard to elections, and in the administration of public services.[2]

## AI and the Democratic Public Sphere

The debate on AI's impact on the public sphere is currently the one most prominent and familiar to a general audience. It is also directly connected to long-running debates on the structural transformation of the digital public sphere. The digital transformation has already paved the way for the rise of social networks that, among other things, have intensified the personalization of news consumption and broken down barriers between private and public conversations. Such developments are often thought to be responsible for echo-chamber or filter-bubble effects, which in turn are portrayed as root causes of the intensified political polarization in democracies all over the world. Although empirical research on filter bubbles, echo chambers, and societal polarization has convincingly shown that the effects are grossly overestimated and that many non-technology-related reasons better explain the democratic retreat, the spread of AI applications is often expected to revive the direct link between technological developments and democracy-endangering societal fragmentation.

The assumption here is that AI will massively enhance the possibilities for analyzing and steering public discourses and/or intensify the automated compartmentalizing of will formation. The argument goes that the strengths of today's AI applications lie in the ability to observe and analyze enormous amounts of communication and information in real time, to detect patterns and to allow for instant and often invisible reactions. In a world of communicative abundance, automated content moderation is a necessity, and commercial as well as political pressures further effectuate that digital tools are created to oversee and intervene in communication streams. Control possibilities are distributed between users, moderators, platforms, commercial actors and states, but all these developments push toward automation (although they are highly asymmetrically distributed). Therefore, AI is baked into the backend of all communications and becomes a subtle yet enormously powerful structuring force.

The risk emerging from this development is twofold. On the one hand, there can be malicious actors who use these new possibilities to manipulate citizens on a massive scale. The Cambridge Analytica scandal comes to mind as an attempt to read and steer political discourses (see next section on electoral interference). The other risk lies in a changing relationship between public and private corporations. Private powers are becoming increasingly involved in political questions

and their capacity to exert opaque influences over political processes has been growing for structural and technological reasons. Furthermore, the reshaping of the public sphere via private business models has been catapulted forward by the changing economic rationality of digital societies such as the development of the attention economy. Private entities grow stronger and become less accountable to public authorities; a development that is accelerated by the endorsement of AI applications which create dependencies and allow for opacity at the same time. The 'politicization' of surveillance capitalism lies in its tendency, as Shoshana Zuboff has argued, to not only be ever more invasive and encompassing but also to use the data gathered to predict, modify, and control the behavior of individuals. AI technologies are an integral part in this 'politicization' of surveillance capitalism, since they allow for the fulfilment of these aspirations. Yet at the same time, AI also insulates the companies developing and deploying it from public scrutiny through network effects on the one hand and opacity on the other. AI relies on massive amounts of data and has high upfront costs (for example, the talent required to develop it, and the energy consumed by the giant platforms on which it operates), but once established, it is very hard to tame through competitive markets. Although applications can be developed by many sides and for many purposes, the underlying AI infrastructure is rather centralized and hard to reproduce. As in other platform markets, the dominant players are those able to keep a tight grip on the most important resources (models and data) and to benefit from every individual or corporate user. Therefore, we can already see that AI development tightens the grip of today's internet giants even further. Public powers are expected to make increasing use of AI applications and therefore become ever more dependent on the actors that are able to provide the best infrastructure, although this infrastructure, for commercial and technical reasons, is largely opaque.

The developments sketched out above – the heightened manipulability of public discourse and the fortification of private powers – feed into each other, with the likely result that many of the deficiencies already visible in today's digital public spheres will only grow. It is very hard to estimate whether these developments can be counteracted by state action, although a regulatory discourse has kicked in and the assumption that digital matters elude the grasp of state regulation has often been proven wrong in the history of networked communication. Another possibility would be a creative appropriation of AI applications through users whose democratic potential outweighs its democratic risks thus enabling the rise of differently structured, more empowering and inclusive public spaces. This is the hope of many of the more utopian variants of AI and of the public sphere literature, according to which AI-based technologies bear the potential of granting individuals the power to navigate complex, information-rich environments and allowing for coordinated action and effective oversight (e.g. Burgess, Zarkadakis).

## AI and (Data-driven) Elections

While the discourse on AI and the democratic public sphere focuses mostly on the societal requirements for a healthy democracy, an additional discourse looks at how we "practice" democracy, namely at elections and how they are conducted. While, in the public-sphere discourse, the assumed risks mostly stem from private actors, the relationship problematized in the discourse on AI and elections takes place more directly between those who govern and the governed.

Most theories of representative democracy conceive of elections as the single most important element of democracy. Elections should ensure that elites stay accountable and orient their actions towards the electorate. Elections are focal points for public discourses and allow the public to express its political preferences in a free and equal way. In order to do so, elections themselves must fulfil certain requirements: They must be conducted in specific ways and while, in the past, risks were mostly associated with the voting process itself, today our attention is also directed to the

electoral campaigns, i.e. the way citizens are addressed and informed in the run-up of the election. It is in this latter area that AI might undermine political discourse by creating effective means for political actors (such as parties or politicians but sometimes also foreign powers) to approach citizens deceptively and to prevent a democratic transmission of the interests and preferences of the citizens into the political system.

Two AI-related developments are often cited as posing threats to the integrity of democratic elections:

**The first trend** relates to the new possibilities for directly manipulating political controversies using AI-powered tools. Here, the risk is that an escalation in antagonistic methods would sway public opinion and discredit political opponents, thereby tarnishing the legitimacy of democracy and eroding the public faith in democratic procedures. Social bots and deep fakes are often seen as the most important instances of this development.

Social bots are social media agents that can be used to amplify opinions by automating certain communications. Especially in the aftermath of the 2016 elections in the US (Trump) and the UK (Brexit), the role of bots has been an important news item. Bots were portrayed as the weapon of choice for outsiders trying to influence elections or create rifts within the citizenry. Automating communication via bots is thought to manipulate the algorithmic logic of social media communication and thereby become an effective means of promoting disinformation.

Deep fakes, on the other hand, are forgeries that directly benefit from AI-techniques in the field of image and audio processing. It has become significantly easier and cheaper to alter audio-visual material convincingly and fast. This can become a weapon in the hands of political actors trying to tarnish the reputation of their opponents. While forgeries and propaganda have always had a place in the history of politics, deep-fake techniques redistribute the power to manipulate to many more actors, make it harder to prove small adjustments, and can further erode trust in all kinds of mediated communication.

While social bots and deep fakes are often identified as areas of concern that will become much more important with the further development of AI techniques, current research has mostly shown the problem to be overstated. Proponents of the position highlighting the risks usually use the technical means as the starting point of their argument, and tend to be oblivious to or underestimate the societal factors that help balance and counter the deployment of these techniques (from investigative journalism to the possibilities of sanctioning actors abusing them). While certain techniques, therefore, certainly have the potential to exacerbate the polarization of political discourse, their overall effect on democratic procedures for the near future will probably be weaker than the scenario painted by alarmists, especially in established democracies with a functioning and pluralistic media environment.

**The second trend** is the expansion of online political microtargeting. This strategy is less aggressive than social bots and deep fakes and much more in line with broader developments of political and commercial communication. Microtargeting depends on enormous amounts of data and elaborate techniques of pattern recognition and can therefore be expected to become even more sophisticated in the coming decade. For years, microtargeting has changed the nature of advertising in business-consumer relations and a driver in the development of the commercial attention economy. Within the political context, microtargeting is viewed with considerable skepticism, since the selectiveness it grants makes it highly manipulative. The argument goes that if political actors are able to tell constituents what they want to hear or to induce supporters of a particular candidate to refrain from voting in a particular election, microtargeting will become an invisible

hand whose effect on an election will be difficult to prove. As with social bots, the effectiveness of microtargeting has to be assessed with care: On the one hand, mobilizing voters and understanding their preferences might be even beneficial for democracy; on the other hand, manipulating voters with a few tailored messages is not as straightforward as suggested since political decisions are not one-off decisions, but often much entwined with personal identities that are hard to change (for more extensive discussions of microtargeting's promises and threats: Zuiderveen Borgesius et al 2018; Kreiss 2017) .

In the next couple of years, we can expect the discourse to transcend the electoral context and the focus on manipulation to include broader reflections about democracy and datafication of the *demos*. More sophisticated and widespread AI applications will bring about different knowledge about the citizens and their behavior and it will be important to see whether democracies decide to respond more actively to the automated detection of preferences and grievances. Elections might decline in importance relative to this continuous observation of the *demos*. While, at first glance, this shift might create more responsiveness and appear to be in line with democratic values, it might also result in a more technocratic version of society, wherein the reliance on observing and categorizing citizens supplants active participation (and the implicit value of a reflexive public debate).

## AI and Public Services

The third angle addressing the impact of AI on democracy takes its starting point from the way public services are conceived and operated. The call for the digitalization of politics often implies a surge in automating decision-making procedures in public administration. Examples reach from welfare administration to tax systems and border control. The hope is that in an ever more complex world a shift towards highly automated systems will result in a more efficient political system. Automation should eradicate failures and frustration, allow for more fine-grained and faster adjudication, and free up resources for other problems.[3]

But why is a change in the way services are administered even debated as a challenge to democracy? An often-unacknowledged premise of democracy is its reliance on the law and legal rules as the basic steering mechanisms of society. Written laws and their application by a judicially controlled executive are seen as a natural fit for democratic systems. Process-based institutionalization is an essential element of the idea of the people governing themselves. Laws and legal structures ensure that democratic regimes are comprehensible for their citizens and create the basis for contestations. Therefore, if states gradually replace or supplement legal processes with automated and adaptive processes, this affects the workings and legitimacy of democratic systems in ways that have to be critically examined.

Proponents of automating society basically look at the outcomes of political systems. They expect legitimacy to be enhanced by automation because of an assumed objectivity of the processes and the improvement of the system's output. This is only one way to look at the changes. Other research literature has highlighted at least two areas of concern:

- On the one hand, problems arise from the logic of machine learning applications as such. As has already been pointed out above and is widely discussed in the broader debate on the societal impact of AI, deep learning applications highly depend on the quality and the amount of the data that is fed into the system. Bias and opacity are therefore often perceived as side-effects that are very hard to prevent (or even discover) since they arise from the very foundations of inductive reasoning. Bias has its roots in the logic that pattern in

past data points are used to extrapolate developments in the future. And opacity, among other things, is a result of the dynamic complexity germane to AI models. AI applications are often imagined as neutral since human intervention can be rendered obsolete in their running, but bias and opacity are two main reasons why neutrality is not achieved. Neutrality is understood insufficiently if only the technical determination of a process is considered. One-sidedness and injustice can arise from many more conditions than just an incoherent or slow application. Automated systems optimize for speed and coherence but they will always be prone to conserve the status quo and are hard to question due to their technical nature and complexity.

- The second issue is related, but more abstract. It pertains to the difference between laws that have to be rather explicit and, therefore, can be contested and explained through interpretative reasoning, and automated decision-making systems that create more frictionless processes by streamlining or personalizing choice architectures. If we assume that citizens will increasingly interact with public administrations through AI-based applications – such as in the case of smart cities where many everyday actions are actively managed by algorithmic interventions – the relationship between the state and its citizens changes (see John Danaher "The Threat of Algocracy"). We are no longer addressed as an abstract and pluralistic entity (the citizens or *the demos*), mostly capable of managing our own affairs within the boundaries of explicitly communicated rules. Instead our interaction with public authorities becomes more personalized, context-dependent and granular. Making use of AI systems to tailor our individual experience in the public sphere constitutes a new and potentially intrusive way of governing and it changes the nature of authority in our systems (from something we collectively decided to trust and control to something we individually rely on).

In sum, the discourse on automating public services highlights the heightened risks for structural domination in highly automated settings. AI allows for targeted interventions into everyday life, and this kind of micromanagement is likely to further intensify if more sophisticated steering instruments become available. While this might be desirable in some regards – think of combatting the climate crisis and how it demands, among other things, society-wide coordination of behavior – from a democratic perspective the question is whether an AI-powered public administration can be held accountable in the same way as a law-based administration could be. Furthermore, the question is whether a focus on prediction does not –  paradoxically – tie our politics too much to the past. AI is rather different from traditional modes of governance since reason-giving is hard to realize within it and the mix of private and public forces that drive the current development constitutes a serious problem with regard to its democratization (see Djeffal on some ideas on how to tackle these problems).

## AI and Democracy: Regulatory Attempts

As the chapter above has shown, there is a range of challenges that the widespread adaptation of AI-applications in the realm of the public sphere, democratic politics and public services might create for democracy. Although many of the risks highlighted so far are rather speculative and do not sufficiently take into account countervailing forces and other balancing factors, we certainly should be aware of the risks that even narrow AI poses with regard to democratic politics. Therefore, this final part of the backgrounder will take a quick glance at current initiatives to regulate AI and see how far they address the challenges to democracy.

That AI needs to be regulated has been a major claim in world politics for about five years. There has been a strong proliferation of regulatory proposals in political systems as diverse as China,

the United States and the European Union, each of which have worked on a comprehensive approach to regulation of AI and its applications. These proposals share some similarities in that they create a master narrative of AI as an inevitable and disrupting development. All proposals point out the high degree of uncertainty regarding the uses and impacts of AI and then go on to translate these into a high demand for regulatory leadership (see Bareis/Katzenbach). Although similar in the overall narrative, national emphasis on what to regulate and in what direction to steer differs strongly: While the American approach – predominantly developed during the Trump administration – is mostly concerned with economic opportunities for American industry, the Chinese AI development plan is more focused on the question of how to best govern society, and carries a more behaviorist logic. European attempts to regulate AI – such as the Draft EU AI Act, the Study Commission on Artificial Intelligence (deployed by the German Bundestag), the German AI strategy, or the French Villani report for A Meaningful Artificial Intelligence – make the strongest use of normative language and most explicitly claim to strike a balance between economic demands and ethical considerations. Therefore, European approaches will be the main focus of the remainder of this paper.

The European proposals are characterized by highlighting the importance of a genuine approach with regard to future competitiveness, but without losing sight of European values that are referred to under broad umbrella terms such as the common good or individual autonomy. Under the wider heading of digital sovereignty, Europe should not only ensure that it remains able to develop and maintain its own digital infrastructure, but also that AI is only deployed if its societal impact is properly reflected and sufficiently controlled. There are two main approaches to achieve this goal: The first is that nearly all recent European documents are risk-based approaches, delineating different types of usage. For instance, the EU draft AI Act distinguishes between prohibited uses such as social scoring, high-risk usages – where among others things, health or security questions are prevalent – and lower risk usages. The second main – and often related – approach is to demand high standards of transparency in order to allow but supervise and qualify the use of AI applications by private actors.

How important are references to democracy and the risks portrayed above within these documents? Democracy is mostly present as a background condition that is strongly tied to the more abstract consideration of European values. It is not in itself a major factor in any of those documents and they rarely focused on it as an institutional setup. Still, each of the three areas designated above is addressed in many of the main documents.

Regarding the structural transformation of the public sphere, the Draft EU AI Act, for instance, prohibits the use of AI for deliberative manipulative actions and creates transparency demands for the use of social bots and deep fakes. The long report of the German Study Commission on Artificial Intelligence is very broad in its recommendations, but discusses all of the issues addressed above and, specifically, stresses the need for a better monitoring of the developments concerning the public sphere. The Villani report additionally emphasizes the need to regulate private powers, but mostly focus on economic aspects and securing Europe's competitiveness.

Topics that are directly linked to democratic elections are rarely addressed outright in the AI regulation proposals. Although the reason could be that elections and campaigns are seen as more of an issue in national election laws, since these most often only slowly adapt to technological changes, more attention should be given to this topic. Microtargeting as a specific practice is indirectly tackled in the General Data Protection Regulation (GDPR) – for an overview see Dobber et al.

With regard to public services, the main line in the European approaches can be described as

leading by example. This means that the usage in public administrations in general is seen as desirable and is encouraged, but the regulatory proposals try to formulate abstract quality standards to ensure that the usage of AI technology is transparently communicated and that non-AI procedures to contest decisions are established. The ethical application of AI in the high-risk environment of public administration should also create an incentive for AI development to fulfill criteria of transparency and accountability.

The developments discussed in this backgrounder show that the discourse on AI and democracy is still in its infancy. Academic treatments and policy adaptation started around the same time and are by now still mostly driven by broader debates on digitalization and democracy and exemplary cases of misuse. The expected spread of AI applications in society will lead to more thorough inspections, and one can expect that in particular, the topic of the use of AI in public services will become a more important issue in the next decade.

### Further Reading

- AlgorithmWatch (2020). *Automating Society Report* 2020. Algorithm Watch / Bertelsmann Stiftung.
- König, P. D., & Wenzelburger, G. (2021). "Between Technochauvinism and Human-centrism: Can Algorithms Improve Decision-making in Democratic Politics?" *European Political Science*, https://doi.org/10.1057/s41304-020-00298-3
- Marcus, G., & Davis, E. (2019). Rebooting AI: Building Artificial Intelligence We Can Trust. Pantheon.
- Nemitz, Paul 2018: Constitutional Democracy and Technology in the Age of Artificial Intelligence, in: Philosophical Transactions of the Royal Society A 376: 2133.
- Nowotny, Helga 2021: In AI We Trust: Power, Illusion and Control of Predictive Algorithms, Medford.
- Risse, Mathias 2021: Artificial Intelligence and the Past, Present, and Future of Democracy (Carr Center Discussion Paper), Cambridge, MA.
- Sudmann, Andreas (ed.): The Democratization of Artificial Intelligence: Net Politics in the Era of Learning Algorithms, Bielefeld.
- Susskind, J. (2018). Future Politics: Living Together in a World Transformed by Tech. OUP.
- Veale, Michael/Borgesius, Frederik Zuiderveen 2021: Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International*, https://doi.org/10.9785/cri-2021-220402

[1] This text does not speculate as to whether there is a path from narrow AI to artificial general intelligence (AGI) and whether AGI would renders necessary a complete rethinking of our societal and political relationships. The literature on AI and democracy can broadly be divided into research that deals with the impact of current developments in AI and Machine Learning (which is the focus of this backgrounder), more far-reaching literature that speculates often from a philosophical stance about the effects of AGI, and literature that uses AI as a catchword that refers to all kinds of digital developments, and, for example, treats AI and algorithms synonymously. While this last strand of literature does not reach beyond very general concerns and will be ignored in what follows, the more philosophical literature on AGI is more compelling, but as far as the focus of this paper is concerned, less relevant as it deals more with abstract questions of humanism and society than on politics and democracy.

[2] A fourth area would be the more abstract idea of to what extent our use of AI techniques transforms our understanding of temporality and our relationship to the future (e.g. Esposito; Nowotny). Approaches in this line discuss whether humanity, by growing accustomed to certain standards of predictive forecasting and adapting our behavior towards them, loses its capacity for individual autonomy and whether this in turn heralds the end of democratic politics as a participatory and contingent practice. Although important and insightful, this line of reasoning will not be included in our further considerations since it pertains to a longer time frame, and is more speculative and less tangible than the issues at hand.

[3] While automated decision-making is not synonymous with AI applications (many of the technologies currently in use are just rule-bound applications and have no element of 'learning'), the use of AI in automated decision-making systems can be expected to grow over time since it allows for the scaling of those systems and for making them more flexible.

---

**6 January 2022**

**by Thorsten Thiel**