UDC 321.7:004.8(4-672EU)
DOI: 10.5937/spm94-61498
Original research article

Српска политичка мисао
(Serbian Political Thought)
No 6/2025.
Vol. 94
pp. 1-23

Ivana Damnjanović*

Faculty of Political Science, University of Belgrade

AI *VS.* DEMOCRACY: PRESENT AND POSSIBLE FUTURES**

Abstract

The aim of this paper is to explore and systematize current research on the relationship between artificial intelligence (AI) and democracy, primarily focusing on existing AI applications, in order to understand the general avenues of exploration conceived within the field of political science on this topic, and to evaluate their results so far. However, some thought will also be given to the more far-fetched possibilities of AI development and their implications for democracy. Using Robert Dahl's theory of democracy, the paper outlines why the AI is perceived as a threat to liberal democracy, especially in regard to principles of equality, personal autonomy, and autonomous choice, and the functioning of the public sphere. Finally, the paper considers some possible, if not very probable, future developments in AI, namely the inception of artificial general intelligence (AGI) and artificial superintelligence (ASI), and their implications for democracy. While existing AI applications do seem to erode certain preconditions for functioning democracy, it is nevertheless possible to overturn this trend by restructuring AI development based on the democratic debate and participatory design.

* E-mail: ivana.damnjanovic@fpn.bg.ac.rs; ORCID: 0000-0002-8553-492X

** This work is supported by the Ministry of Science, Technological Development, and Innovation of the Republic of Serbia (grant number 451-03-136/2025-03 from 27 January 2025).

Keywords: artificial intelligence, democracy, equality, autonomy, public sphere, artificial general intelligence, artificial superintelligence

INTRODUCTION

In recent years, artificial intelligence (AI) has been increasingly moving from the realm of science fiction to reality, even in its most mundane, almost quotidian aspects. While there is still a significant discrepancy between the visions of artificial intelligence promoted for decades, even centuries[1], by artists and futurists – images of superintelligent machines and humanoid robots – and the current reality of chatbots, image generators, and algorithmic recommendations, artificial intelligence seems to be conquering our daily lives.

Politics being an important (perhaps the most important) aspect of our shared reality, it is not surprising that the body of scholarly literature investigating different aspects of the potential impacts of AI on politics is rapidly and steadily growing. While it is still somewhat underdeveloped compared to, for example, the study of ethical issues related to AI, this research nevertheless already points towards some of the more pressing questions that need to be answered. Combined with the justified concern for democracy in the 21st century world, from the democratic malaise (Newton 2012, 7; Di Gregorio 2021, 12; Kupchan 2012, 62) to the democratic hollowing (Mair 2023, 2) and backsliding (Bermeo 2016, 6; Wolkenstein 2022, 265) in the last decade, the new momentum of AI research, and more importantly, AI applications, prompted the scholars in the field of political science to investigate whether it poses a threat to democracy or can, maybe, be used to strengthen it and make it better.

This is not surprising – after all, artificial intelligence is a technology (or, perhaps more accurately, a family of technologies), and technologies are not politically neutral (Winner 1980, 123; Kranzberg 1986 545; Coeckelbergh 2022, 4). Furthermore, as Narayanan and Kapoor argue, it is a *normal* technology (Narayanan and Kapoor 2025). And every new technology, according to Winner, is lauded as the great democratizer, while eventually serving those already in power to further

[1] One of the first descriptions of superintelligent machines and their potential to enslave humanity can be found in Samuel Butler's Erewhon, first published in 1872 (Butler [1872] 1901).

strengthen and entrench their position (Winner 2010, 107). Just like with the other technologies with socially transformative potential, such as the nuclear power and the Internet, AI's democratizing, or de-democratizing potential merits serious consideration (see, for example, Hand and Sandywell 2002, 198; Dahl 1953, 1; Bartoletti 2020, 127).

The aim of this paper is to explore and systematize current research on the relationship between AI and democracy, primarily focusing on existing AI applications, in order to understand what are the general avenues of exploration conceived within the field of political science on this topic, and to evaluate their results so far. However, some thought will also be given to the more far-fetched possibilities of AI development and their implications for democracy. Both of these lines of inquiry should help in identifying the overlooked issues and outlining possibilities for further research.

In order to properly assess the existing scholarship on the topic, it is necessary to start with conceptual issues, given that both artificial intelligence and democracy are notoriously hard to define. Afterwards, the paper will consider AI as a threat to democracy, by reexamining some of the issues identified so far in the scholarly literature – AI's discriminatory potential, its relation with political participation, as well as its impact on the public sphere. Another section of the paper will be devoted to the examination of the ways in which artificial intelligence could be used to strengthen democracy. Finally, a section of the paper will consider the possible futures of artificial intelligence, such as the emergence of artificial general intelligence (AGI) or artificial superintelligence (ASI), and their potential implications for democracy.

COMPLICATED CONCEPTS: DEFINING AI AND DEMOCRACY

In order to determine how artificial intelligence relates to democracy, it is necessary to clear up the terminological confusion and provide at least working definitions of both *artificial intelligence* and *democracy*. This is not an entirely trivial task, given that democracy is notoriously an essentially contested concept, singled out as such by Gallie himself in his seminal paper (Gallie 1955, 168). While a certain degree of conceptual confusion is a rule rather than an exception in the social sciences, definitions of "technical terms", and one may argue that artificial intelligence *is* one of them, are usually more precise and less

controversial. And yet, there seems to be no consensual definition of AI. The body of literature on artificial intelligence struggles, time and again, with the very definition of its basic concept. Some authors even claim that it is intrinsically, and by design, nebulous, vague, and empty of objective meaning (Katz 2020, 6).

The origin of the term *artificial intelligence* is well documented: it was introduced at the conference at Dartmouth College in 1956 by John McCarthy. However, its content has been changing ever since. Some authors point out that at least part of the problem stems from the fact that there is no agreement on the definition of intelligence as such (Legg and Hutter 2007, 392; Warwick 2013, 12). While this may well be a significant obstacle for achieving a consensual definition of AI, there are other, equally important issues at play.

Nearly four decades ago, Negrotti[2] collected about 180 definitions of artificial intelligence from academics and practitioners gathered for a major conference. These definitions turned out to be quite different from one another, yet could be grouped, according to Trappl's interpretation, into two major categories, based on the perceived *objectives* of AI. The first group sees AI as "the science and technique to make computers smart", which would enable them to "perform tasks which normally require human intelligence", while the second sees AI as a tool to "better understand model human intelligence" (Trappl 1992, 3). Others point out that not only objectives, but also *technologies* that constitute AI are disputed – Katz, for example, shows that in the 1960s neural networks were not seen by practitioners as part of AI, while today this is considered one of the basic technologies beyond AI applications (Katz 2020, 49). In his view, AI "quickly became an evocative label for an academic field, a concept, and an industry" (Katz 2020, 3).

Another potential problem is that public perceptions of AI stem primarily from popular culture, where artificial intelligence is presented as general, usually smarter than humans, conscious, and sometimes anthropomorphic. In his paper, Preethi suggests that "[s]ome industry experts believe that the term artificial intelligence is too closely linked to popular culture, causing the general public to have unrealistic fears about how it will change the workplace and life in general" (Preethi

[2] His original research was published in 1983 (Negrotti 1983). It is cited here according to Katz (Katz 2020, 37) and Trappl (Trappl 1992, 6). Some of Negrotti's results relevant to the topic of this paper are also published later (see Negrotti 1987, 114).

2020, 40). And yet, the current state-of-the-art in the AI field is very far from these depictions.

While it is common that the very concept of *technology* is frequently equated exclusively with the latest generation of technological innovations,[3] this trend seems to be even more amplified when it comes to AI, and not only among the general population, but also among the researchers or practitioners. As Narayanan and Kapoor remind us, there is a running joke that AI is "what hasn't been done yet", suggesting that earlier successful and "domesticated"[4] applications are simply not referred to as AI anymore (Narayanan and Kapoor 2024, 14).[5] In their view, artificial intelligence is an umbrella term covering a variety of technologies, only loosely related, and its use is frequently influenced by historical usage, marketing, etc.[6] These authors, however, do not see the lack of conceptual clarity as a major issue and suggest focusing on specific problems and applications.

In order to provide a clear focus for the research, in this paper, the definition of AI provided by the High-level expert group on artificial intelligence[7] appointed by the European Commission will be used. According to this group of experts, "artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals" (European Commission's High-Level Expert Group on Artificial Intelligence 2018). This definition is precise enough to distinguish AI from other technologies, and still broad enough to encompass the current forms of AI, such as, among others, generative and predictive AI used in most commercially available applications, as well as potentialities such as artificial general intelligence and artificial superintelligence. It thus covers all the major varieties of AI that this paper will consider.

Debates about the meaning of democracy are much longer-running and more complicated. There are competing theories of democracy, which reflect the complexity of the phenomenon itself, as well as its

[3] It could be argued that what people considered by technology in, for example, the 1950s is very different from what they consider by technology today.

[4] Domestication of technology is a framework developed by Berker *et al.* to explain how technologies transition, in the eyes of the public, from new and exciting/dangerous to mundane and "boring" (Berker *et al*. 2005, 2–9).

[5] Many other authors have made similar points (see Sheikh *et al*. 2023, 17).

[6] In this, they basically concur with Katz, who claims, *inter alia*, that AI is what AI people, at a given point of time, do (Katz 2020, 33).

[7] For more details about this group, see European Commission 2025.

capacity for transformation (see, for example, Cunningham 2002, 15–26). Furthermore, the notion of democracy is always, at least to some extent, normative. This means that every discussion about democracy is a partial discussion about values, thus earning it the already mentioned qualification of essentially contested concept. It is not surprising, then, that there is a vast amount of scholarly literature on the topic, which cannot be adequately addressed within the scope of this paper.

For the purposes of this paper, the term *democracy* will be used to denote primarily representative, or *liberal* democracy, as defined most famously by Robert Dahl (Dahl 1989, 83–118). Drawing from the broad principles of intrinsic equality and effective participation, he defines democratic process as a normative concept that can, in theory, be applied in all settings requiring collective and binding decisions. When applied to modern nation-states, it manifests as a polyarchy, a political system where institutions embodying and enabling the democratic process are present over a certain, but not firmly fixed, threshold. Inclusion, personal autonomy, freedom of expression, as well as access to alternative sources of information, are important features of his (and most other) conceptions of democracy.

Focusing on liberal democracy does not, of course, imply that other conceptualizations of democracy are irrelevant for the discussion about AI and democracy, but including all, or even some of them, would dramatically expand the scope of this paper. Exploration of diverse theoretical approaches to democracy and their relationship with artificial intelligence is in itself a worthy research goal, and some important steps in that direction have already been made by Coeckelbergh, who discusses the issues that AI can cause within the frameworks of deliberative, participative, and agonistic theories of democracy (Coeckelbergh 2022, 62–84; 2024, 29–38).

IS ARTIFICIAL INTELLIGENCE A THREAT TO DEMOCRACY?

Most of the scholarly research on AI and democracy explicitly sees AI, in its current form and its current applications, as a potential or actual threat to democracy. In part, it is probably caused by the observation that we are witnessing, in Huntington's terms, a reverse wave of democracy (Huntington 2012, 15–16). But also, AI is seen as part of the debate on the

role of *Big Tech*[8] in society. The concentration of economic power and, consequently, the political influence these companies wield, is viewed by many as a major threat to democratic politics and, more widely, social and political equality as its underlying condition (see, for example, Zuboff 2019, 497; Giblin and Doctorow 2022, 94; Varoufakis 2023, 27). From this perspective, to put it succinctly, the threat to democracy is not really AI, it's rampant capitalism (Coeckelbergh 2022, 21–22; 2024, 48).

There are, however, certain domains in which artificial intelligence is currently, and increasingly, applied that seriously undermine democracy by degrading social and political equality, personal autonomy, and the free and diverse public sphere as preconditions for political participation and democratic decision-making. In the following pages, some of these issues will be explained in more depth.

AI *vs.* Equality

Robert Dah's theory builds on the principle of *intrinsic equality* – the notion that each person's interests are equally deserving of consideration, and that no group should be privileged, in terms of influence on political decision-making, above the others, based on features such as gender, wealth, race etc. In this sense, the issue of political equality is closely tied to the previously mentioned concentration of wealth in the hands of the few.[9]

But the modern digital technologies are eroding equality in much more specific ways. In her seminal work, O'Neil has demonstrated that algorithms and big data entrench and deepen existing inequalities, targeting and affecting the most vulnerable groups disproportionately (O'Neil 2016, 11). While she insists that her examples do not include AI, both algorithms and big data are the necessary building blocks of all AI applications, and it has been shown that the use of AI only exacerbates this trend (Narayanan and Kapoor 2024, 53; Coeckelbergh 2024, 47). If automated sentencing systems[10], for example, consistently propose

[8] *Big Tech* is a shorthand term denoting a small number of powerful technological companies, mostly based in a few hubs in the USA and China. For a more detailed explanation, see Birch and Bronson 2022, 3–5.

[9] It should be noted that throughout his work, Karl Marx argued that political equality without meaningful social (including economic) equality is just an empty shell.

[10] Such systems are already in use in a number of jurisdictions in the USA, and there is a growing body of academic, although primarily legal literature, on their features as well as their intended and unintended consequences.

harsher penalties for the people belonging to certain social groups, based on indicators such as race or income, they are not only propagating existing biases, but also potentially disenfranchising them, for example, in the states where prisoners or convicts are excluded from participating in the elections.

Even though the most common justification for the use of such systems in all domains of public services is to provide impartial and objective results, eliminating bias by removing humans (who are prone to bias) from the equation, predictive systems based on AI are repeatedly reported to produce biased results. These biases can be "in the training data, in the algorithm, in the data the algorithm is applied to, and in the teams that program the technology" (Coeckelbergh 2022, 38). Some authors are very pessimistic about the possibility of fixing those systems, pointing out the inherent limits and flaws of the so-called *predictive* AI (see Narayanan and Kapoor 2024, 10).

Even more directly, the manipulation of voters through micro-targeting, revealed in the 2016 Cambridge Analytica scandal, could influence the outcome of the elections. Whether that was the case in the USA presidential election is unclear, but the testimonies suggest that it *is* possible to influence the results by targeting a small number of "swing" voters in a small number of "swing states". According to O'Neil, this effectively creates "the political 1%", a tiny number of voters on whom the majority of resources of political campaigns are focused (O'Neil 2016, 196).

It could be argued that such targeted manipulation is a concern only for particular electoral systems and can thus be easily solved through electoral engineering, but, scale-wise, there is an issue with AI on the opposite end of the spectrum. Regardless of the type of electoral system, generative AI can be used to produce an extremely vast quantity of text impersonating citizens. If the policy-makers are faced with such an influx of materials, for example, letters from concerned citizens, it could influence their decision, especially if they are unable to distinguish between genuine and AI-generated e-mails. Several such instances were already recorded, albeit on a smaller scale and related to single political issues (Narayanan and Kapoor 2024, 140–141).

AI *vs*. Autonomy

There are two main applications of AI that restrict personal autonomy as a prerequisite for democracy: surveillance and recommendation systems. Surveillance has been a long-standing topic of the political debate. Since Jeremy Bentham's invention of the *panopticon*, and later adaptation of the term by Foucault, it is a widely accepted fact that people tend to behave differently when they believe that they are being watched (Bentham 2020, 35; Foucault 1977, 195). Similarly, Ellul warned about the dangers of technologically enabled surveillance turning modern societies into concentration camps (Ellul 1964, 100). The increase of surveillance and data collection about the population indeed went hand in hand with the very construction of the modern nation state. "Nation states," Coeckelbergh points out, "especially, are data hungry" (Coeckelbergh 2024, 14–15). They are thus very keen on quickly adopting technologies that enable more widespread and more efficient ways to track their own citizens (as well as the others who are within reach).

While such technologies can obviously be abused by authoritarian regimes and malicious actors within democracies, even the most democratic governments tend to see them as, at least, a necessary evil contributing to better security. Artificial intelligence was promptly mobilized for the task, and today, one of the most widespread applications of AI is in the field of facial recognition. Combined with the variety of other data compiled about citizens by state and private actors, complete surveillance seems to be within reach. Even without further repressive actions from the state, the awareness that they are being watched could seriously alter the manner in which people behave and the extent to which they are willing to participate in various types of political activities – from attending protests to speaking their mind on social media, but also in private conversations.

Another way for AI applications to degrade personal autonomy is by changing the choice architecture. This can be done by several means: recommendation systems, algorithmic feeds on social media, or "nudging". Recommendation systems are probably the most pervasive and the most underestimated use of artificial intelligence today. After all, streaming platforms, online stores, and other businesses seem to know us so well (thanks to the vast amount of data about us that they have collected) that their recommendations of what to listen to, watch, or buy

are rarely completely off the mark. For this reason, they are frequently implicitly trusted. This trust could, however, be abused to skew the election results. Similar technologies can be used in political marketing, recommending parties and candidates who are the most in line with the interests and ideologies of a specific voter, or tailoring the candidates to suit the particular group or individual.[11]

In the early years of social media, the feeds, that is, the posts that are shown to a user when they log in, were quite straightforward. They would show posts from their friends on the platform, and from the pages they decided to follow. But by 2016, most platforms had switched to so-called algorithmic feed, where the platform would show you a personalized feed based not exclusively on your preferences, but on engagement optimization, and not in chronological order. The exact extent to which the algorithmic feeds on social media are based on artificial intelligence is not certain, since the software running them is proprietary and, by and large, not available for external scrutiny. The consequences of the algorithmic feeds will be discussed in greater detail in the next section, but in the context of autonomy and autonomous choices, the important issue is limiting the range of choices available to citizens. If the visibility of political candidates and their positions on a range of social issues is determined by an opaque algorithm, it could easily affect their chances in the electoral race. Some of the companies behind platforms, most prominently Meta, the owner of Facebook, have experimented with their ability to influence citizens' interest in voting, but also with their voting preferences.[12]

Finally, *nudging* is another technique that can be enhanced by artificial intelligence and used to circumscribe autonomous choices. Introduced by Thaler and Sunstein in their highly influential book, it suggests that human behavior can be influenced without coercion and without limiting personal autonomy of the choice, by making simple changes in the choice architecture – for example, by switching the default answer from *no* to *yes* (Thaler and Sunstein 2021, 103–130).

[11] This kind of tailoring is not a new development in political marketing, but the use of AI could make it much more successful (see O'Neil 2016, 187; McGinniss 1988, 26).

[12] The general description of these experiments can be found in O'Neil (O'Neil 2016, 180–184). For more detailed analysis of the different aspects of the voter turnout (see Theocharis and Lowe 2016, 1465–1486; Haenschen 2023, 1661–1681; 2016, 542–563).

Possible uses in the public policy have drawn much attention both from scholars and policy-makers (for a nuanced discussion, see Schmidt 2017, 405), as well as much criticism (for example, Morozov 2013, 198). Even if there is no overt coercion, the subtle persuasion of nudging can still be construed as manipulation, and, as such, antithetical to autonomous choice. Again, while AI is not a necessary component in nudging, its use in this domain could exacerbate the scale of the problem (Coeckelbergh 2022, 17).

AI and the Public Sphere

Even among the scholars who do not subscribe to the theory of deliberative democracy, grounded in Habermas's work on the public sphere (Habermas 1991, 236) and communicative action (Habermas 1985b, 43–76; 1985a, 273–338), there is a consensus about freedom of expression and an informed public being the prerequisites of a functional democracy. In Dahl's theoretical framework, both freedom of expression and access to alternative sources of information are among the institutional guarantees that make a state a polyarchy. There is already much debate about the role of social media in the degradation of the public sphere.[13]

The conventional wisdom is that the algorithmic feeds work towards reducing the public sphere to many small private spheres, creating so-called echo chambers and epistemic bubbles, thus reducing the space for the debate about social and political issues (Coeckelbergh 2022, 76). Focus on engagement fuels polarization, and the use of AI to boost or game social media algorithms is likely to further aggravate this problem. Even though AI is increasingly used in content moderation across all platforms, it seems that it will not be the solution to polarization and the dissolution of the public sphere (Narayanan and Kapoor 2024, 179).

[13] Interestingly, until 2016, most of the scholarly discourse was along the lines of the Internet and the social media being the new, and potentially better public sphere. The turning point seems to be the Cambridge Analytica scandal, and the subsequent scholarship is much more interested in the dangers social media poses to the public sphere.

WHAT COULD THE FUTURE BRING?

All of the applications considered artificial intelligence that are in use today fall under what is usually referred to as the *narrow* AI – namely, a system that is designed to perform a single task or several related tasks. This pertains to generative AI as well, despite the illusion of generality it projects. By many researchers and practitioners in the field, narrow AI is seen as just a step on the road towards a true goal: artificial general intelligence. Whether it is conceived as an imitation of human intelligence or in some other manner, AGI is generally understood as artificial intelligence that is capable of performing any[14] task a human being can perform. The predictions about AGI differ significantly – some researchers believe that this goal could now be achieved in just a couple of decades, while others doubt that it can be achieved at all (Armstrong and Sotala 2015, 24; Muehlhauser and Salamon 2012, 25; Mueller 2024).

The emergence of AGI potentially opens a new set of issues regarding democracy, especially in relation to the concept of inclusion. It is illustrative that when, at a lecture some years ago, Ray Kurzweil was asked, "In a world where AIs passed the Turing test,[15] who gets to vote? Does democracy make sense?"[16] He immediately interpreted the question as "should AI be allowed to vote?" (and gave no clear answer).

Although the original question can be interpreted in different, and perhaps more interesting ways, the issue of political rights for AGI is important for the discussion about democracy. A crucial part of democracy

[14] Narayanan and Kapoor would add "economically relevant" tasks (Narayanan and Kapoor 2024, 150).

[15] The Turing test is a popular name given to the "imitation game" introduced by Alan Turing in his seminal paper (Turing 1950). He proposes that if the way in which the machine communicates is indistinguishable from human communication, then the machine can be labeled as intelligent. However, chatbots, from the early and rudimentary ones such as ELIZA (for more information on this program developed in the 1960s, see Weizenbaum 1966, 36), to popular apps such as ChatGPT or Claude, have been passing this test for some time, without being intelligent in any meaningful way. In recent decades, there has been a wide consensus about the inadequacy of the Turing test for assessing machine intelligence, with some authors going so far as to call it a "blind alley" (Whitby 1996, 53–65) as well as an "ideology" (Halpern 1987, 79–93) of AI research. However, its significance as an inspiration to the entire field can not be disputed.

[16] The video can be seen at Vimeo (Vimeo 2015). Ray Kurzweil is sometimes referred to as "the prophet of Singularity" (see, for example, Tirosh-Samuelson 2012, 722) and has authored several books that deal, among other topics, with artificial intelligence (see Kurzweil 2005, 203–226; 2001, 40–57; 1992, 401–416).

is, obviously, *the demos* – a body politic of citizens who are allowed to participate in decision-making. Who constitutes the demos is, thus, not a trivial question, even though some authors have tried to dismiss it.[17]

For democracy to function, the demos must be clearly outlined. Both Dahl and Sartori insist that, although democracy must be as inclusive as possible – that is, the largest possible number of people within a democracy should be included in political decision-making – the demos is also by definition *exclusive* (Dahl 1989, 119; Sartori 1987, 21–25). In other words, it is necessary to determine precisely who is and who isn't part of the demos in a democratic political unit. The composition of the demos, obviously, varies throughout both time and space, but some categories are consistently excluded even in our age of universal suffrage: for example, children and non-citizens. Both of these grounds for exclusion bear on the issue of AGI and its political rights.

The suffrage is today usually tied to the age of majority, which is also the moment when the full legal capacity is acquired. Children are excluded from the demos based on their incompetence. However, it is not necessary self-evident that this incompetence relates to levels of *intelligence* and not, for example, emotional immaturity. So, even if AGI is a human-level intelligence, is this enough for it to be recognized as part of the demos?

Part of the problem lies in the conceptual confusion surrounding the very concept of intelligence, and its relationship with other concepts, such as consciousness and sentience. Is self-awareness and the capacity to feel necessary for one to be recognized as a subject of political rights? Can AGI be truly intelligent without being conscious and/or sentient? While there are strong arguments for AI, even if it never achieves general intelligence, consciousness and sentience, it could be eligible for some kind of political standing (Coeckelbergh 2022, 142), right to vote is a completely different matter.

Another ground for exclusion from a specific demos is citizenship. As a rule, only citizens can vote in the national election. Even if we argue that there are no good reasons why AGI should not be granted citizenship,[18] and the rights stemming from the citizen's status (see

[17] Most famous among them is probably Joseph Schumpeter, who argues that the demos defines itself (Schumpeter 2018, 258–259).

[18] In 2017, part of a publicity stunt, a "social robot" named Sophia was granted citizenship of Saudi Arabia, spurring a series of opinion pieces (see, for example, Reynolds 2018).

Jaynes 2020, 346), it is not clear how AGI's citizenship would be determined. Should it, by default, be given by the state where the legal entity responsible for the AGI's inception resides? If the AGI in question is not embodied within a, for example, robot, but exists only as software on the Internet, where does it reside (and, consequently, where should it vote)? If the AGI is capable of making copies of itself and storing them on servers around the world, should every one of those copies be granted suffrage under appropriate state laws?

Finally, if the current AI applications are endangering democracy by limiting human autonomy, the question of AGI's autonomy could also be posed. It could be argued that its choice architecture would be even more severely limited by original programming (see Mueller 2024). Would the AGI then be capable of autonomous action and autonomous choice?

Many authors have warned that, even though it may be the Holy Grail of AI research, AGI would not be the end of the AI evolution. They argue that if AGI is achieved and capable of self-improvement, it would quickly evolve into artificial superintelligence, possibly leading to an intelligence explosion and/or singularity. While these two concepts are frequently used interchangeably, they are not quite synonymous. The intelligence explosion, according to Muehlhauser, refers to the moment when intelligent machines will surpass the human level of intelligence (Muehlhauser 2013, 79). The technological singularity, on the other hand, seems to be a more vague and somewhat contested term. In Sandberg's words, some definitions of technological singularity "stress the role of artificial intelligence, others refer to more general technological change (Sandberg 2013, 377). These meanings can overlap, and many writers use combinations of meanings." However, it seems that it is mostly understood as a moment of extremely accelerated technological change, which renders human comprehension and prediction impossible. This is frequently seen as a direct consequence of the intelligence explosion. Optimists place the possible occurrence of such an event somewhere in the second half of the 21st century.[19] There is a certain logic in the described trajectory from AGI to ASI: if AGI were able to take over *any* task, one of those tasks would probably be further AI research, thus resulting in more and more sophisticated and powerful AIs, which would

[19] Pessimists, on the other hand, believe that neither AGI nor ASI can ever be achieved.

quickly drastically surpass humans in every possible way (Narayanan and Kapoor 2024, 151).

The emergence of ASI is considered to pose an existential risk to humanity by a significant number of academics and public figures. The most prominent among them are probably Eliezer Yudkowsky and Nick Bostrom (Yudkowsky 2008, 308–345; Bostrom 2014, 115–116).[20] Even if it turns out that the inception of ASI is not an extinction-level event, it would still profoundly influence and shape human society and politics, with possible grave consequences for democracy.

One of the first things the superintelligence would strive to do, according to Bostrom, is to secure its position as a singleton. Singleton, in this context, means an entity with the global decision-making capacity, with no competing peers. In other words, perhaps more recognizable in the political science literature, it would become a global sovereign. What would that mean for human politics as such and for democracy specifically is hard to determine. Perhaps ASI would introduce totalitarian control beyond everything humanity has experienced so far. On the other hand, perhaps it would not be interested in human affairs at all, as long as they are not in the way of ASI's own goals.

But democracy would nevertheless be in peril, even if ASI wanted to preserve it. The chasm between the knowledge and processing power available to ASI and that available to humans would be so vast that the most of the arguments for democracy and against guardianship presented by Dahl could not hold anymore (Dahl 1989, 52–65).[21] If we do not allow children to participate in political decision making, why would ASI let us do it, being presumably much more superior to us than we are to children?

Of course, every debate about AGI and ASI is at this point completely speculative. It was already noted that insisting on risks stemming from these hypothetical future forms of AI could very well distract us from real harms that AI applications are causing now. Nevertheless, thinking about these matters can be inspiring and useful for testing the limits of justifications for democracy and its ability to cope with new and unforeseen challenges.

---

[20] Bostrom does not necessarily see superintelligence as machine intelligence. He identifies three possible paths that could lead to superintelligence: artificial intelligence, biological cognition enhancement, and collective intelligence.

[21] For a more elaborate discussion on this particular topic, see Damnjanović 2015, 79–80.

CONCLUSION

The scholarly debate about relationships between artificial intelligence and democracy is thriving, even if it is still in its early stages. The importance of the topic is recognized not only by scholars but also by policy-makers. For example, the European Union has explicitly classified AI systems "intended to be used for influencing the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda" as high-risk (Regulation 2024/1689, Annex III). In the wider context of the ongoing crisis of (liberal) democracy, rapidly evolving AI applications and their impact on societies in general and on democracy in particular are rightly seen as a cause for concern.

Main reasons why the AI, in its current form, is seen as a threat to democracy are its already documented roles in increasing or entrenching inequalities, restricting personal autonomy through mass surveillance and nudging, further eroding the public sphere, and concentration of economic (and, consequently, possibly political) power.

However, these harmful impacts could be overturned by putting existing AI systems to different use: by making them work in such a way as to improve, not degrade, democratic values and practices. An important part of this would be a wider democratic debate about AI, as well as a more participatory approach to AI development (see Coeckelbergh 2022, 152; 2024, 81; Narayanan and Kapoor 2024, 17, 265).

Future threats to democracy posed by the development of artificial general intelligence, and, subsequently, artificial superintelligence are, for now and maybe forever, mostly useful as thought experiments. It could be argued that most of the interest stems from dire doomsday scenarios, as well as from anthropomorphizing AI, that is, ascribing to it the humanlike features it does not possess. While the value of such thought experiments for examining and predicting the future of AI research and application is limited, they can help us think about the limits and weak points of our understanding of democracy and its justifications. They should not, however, divert us from the search for solutions to more pressing problems presented by AI, and for ways to reimagine and restructure AI research and application in ways more supportive and conducive to democracy.

REFERENCES

Armstrong, Stuart, and Kaj Sotala. 2015. "How We're Predicting AI–or Failing To." In *Beyond Artificial Intelligence*, edited by Jan Romportl, Eva Zackova, and Jozef Kelemen, 11–29. Cham: Springer.

Bartoletti, Ivana. 2020. *An Artificial Revolution: On Power, Politics and AI*. London: Indigo Press.

Bentham, Jeremy. 2020. *The Panopticon Writings*. London: Verso Books.

Berker, Thomas, Maren Hartmann, and Yves Punie [Berker *et al*.]. 2005. *Domestication of Media and Technology*. Maidenhead: McGraw-Hill Education.

Bermeo, Nancy. 2016. "On Democratic Backsliding." *Journal of Democracy* 27 (1): 5–19. DOI: 10.1353/jod.2016.0012

Birch, Kean, and Kelly Bronson. 2022. "Big Tech." *Science as Culture* 31 (1): 1–14. DOI: 10.1080/09505431.2022.2036118

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Butler, Samuel. [1872] 1901. *Erewhon: Or, Over the Range*. Project Gutenberg. https://gutenberg.org/cache/epub/1906/pg1906-images.html

Coeckelbergh, Mark. 2022. *The Political Philosophy of AI: An Introduction*. Hoboken, NJ: John Wiley & Sons.

Coeckelbergh, Mark. 2024. *Why AI Undermines Democracy and What To Do About It*. Hoboken, NJ: John Wiley & Sons.

Cunningham, Frank. 2002. *Theories of Democracy: A Critical Introduction*. London: Routledge.

Dahl, Robert A. 1953. "Atomic Energy and the Democratic Process." *The ANNALS of the American Academy of Political and Social Science* 290 (1): 1–6. DOI: 10.1177/000271625329000101

Dahl, Robert A. 1989. *Democracy and Its Critics*. New Haven, CT: Yale University Press.

Damnjanović, Ivana. 2015. "Polity Without Politics? Artificial Intelligence Versus Democracy: Lessons From Neal Asher's Polity Universe." *Bulletin of Science, Technology & Society* 35 (3–4): 76–83. DOI: 10.1177/0270467615623877

Di Gregorio, Luigi. 2021. *Demopathy and the Democratic Malaise: Symptoms, Diagnosis and Therapy*. Cheltenham: Edward Elgar Publishing.

Ellul, Jacques. 1964. *The Technological Society*. New York: Vintage Books.

European Commission's High-Level Expert Group on Artificial Intelligence. 2018. "A Definition of AI: Main Capabilities and Scientific Disciplines." *European Commission*. December 18, 2018. https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf

European Commission. 2025. "High-level expert group on artificial intelligence". *Digital-strategy*. May 6, 2025. https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

Foucault, Michel. 1977. *Discipline and Punish: The Birth of the Prison*. New York: Penguin Books.

Gallie, Walter Bryce. 1955. "Essentially Contested Concepts." *Proceedings of the Aristotelian Society* 56: 167–198.

Giblin, Rebecca, and Cory Doctorow. 2022. *Chokepoint Capitalism: How Big Tech and Big Content Captured Creative Labor Markets and How We'll Win Them Back*. Boston: Beacon Press.

Habermas, Jürgen. 1985a. *The Theory of Communicative Action: Lifeworld and System: A Crituque of Functionalist Reason*. Vol. 2. Boston: Beacon Press.

Habermas, Jürgen. 1985b. *The Theory of Communicative Action: Reason and the Rationalization of Society*. Vol. 1. Boston: Beacon Press.

Habermas, Jurgen. 1991. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Cambridge, MA: MIT Press.

Haenschen, Katherine. 2016. "Social Pressure on Social Media: Using Facebook Status Updates to Increase Voter Turnout." *Journal of Communication* 66 (4): 542–563. DOI: 10.1111/jcom.12236

Haenschen, Katherine. 2023. "The Conditional Effects of Microtargeted Facebook Advertisements on Voter Turnout." *Political Behavior* 45 (4): 1661–1681. DOI: 10.1007/s11109-022-09781-7

Halpern, Mark. 1987. "Turing's Test and the Ideology of Artificial Intelligence." *Artificial Intelligence Review* 1 (2): 79–93. DOI: 10.1007/BF00130010

Hand, Martin, and Barry Sandywell. 2002. "E-Topia as Cosmopolis or Citadel: On the Democratizing and De-Democratizing Logics of the Internet, or, Toward a Critique of the New Technological Fetishism." *Theory, Culture & Society* 19 (1–2): 197–225. DOI: 10.1177/026327640201900110

Huntington, Samuel P. 2012. *The Third Wave: Democratization in the Late 20th Century*. Norman: University of Oklahoma Press.

Jaynes, Tyler L. 2020. "Legal Personhood for Artificial Intelligence: Citizenship as the Exception to the Rule." *AI & SOCIETY* 35 (2): 343–354. DOI: 10.1007/s00146-019-00897-9

Katz, Yarden. 2020. *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. New York: Columbia University Press.

Kranzberg, Melvin. 1986. "Technology and History:"Kranzberg's Laws"." *Technology and Culture* 27 (3): 544–560.

Kupchan, Charles A. 2012. "The Democratic Malaise: Globalization and the Threat to the West." *Foreign Affairs* 91 (1): 62–67.

Kurzweil, Ray. 1992. *The Age of Intelligent Machines*. New York: Viking.

Kurzweil, Ray. 2001. *The Age of Spiritual Machines: How We Will Live, Work and Think in the New Age of Intelligent Machines*. New York: Texere.

Kurzweil, Ray. 2005. *The Singularity Is Near: When Humans Transcend Biology*. New York: Penguin Publishing Group.

Legg, Shane, and Marcus Hutter. 2007. "Universal Intelligence: A Definition of Machine Intelligence." *Minds and Machines* 17 (4): 391–444. DOI: 10.1007/s11023-007-9079-x

Mair, Peter. 2023. *Ruling the Void: The Hollowing of Western Democracy*. London: Verso books.

McGinniss, Joe. 1988. *The Selling of the President: The Classic Account of the Packaging of a Candidate*. New York: Penguin Publishing Group.

Morozov, Evgeny. 2013. *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems That Don't Exist*. London: Penguin Books.

Muehlhauser, Luke. 2013. *Facing the Intelligence Explosion*. Berkley, CA: Machine Intelligence Research Institute.

Muehlhauser, Luke, and Anna Salamon. 2012. "Intelligence Explosion: Evidence and Import." In *Singularity Hypotheses*. edited by Amnon H. Eden, James H. Moor, Johnny H. Søraker, and Eric Steinhart, 15–42. Berlin: Springer.

Mueller, Milton. 2024. "The Myth of AGI." *Internet Governance Project*. Last accessed January 17, 2025. https://www.internetgovernance.org/wp-content/uploads/MythofAGI.pdf

Narayanan, Arvind, and Sayash Kapoor. 2024. *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*. Princeton, NJ: Princeton University Press.

Narayanan, Arvind, and Sayash Kapoor. 2025. "AI as Normal Technology." *The Knight First Amendment Institute*. April 15, 2025. https://knightcolumbia.org/content/ai-as-normal-technology

Negrotti, Massimo. 1983. *How AI People Think*. Report III/2203/En 83. Chair of Sociology of Knowledge, University of Genoa.

Negrotti, Massimo. 1987. "The AI Peoples Way of Looking at Man and Machine." *Applied Artificial Intelligence: An International Journal* 1 (1): 109–116. DOI: 10.1080/08839518708927963

Newton, Kenneth. 2012. "Curing the Democratic Malaise with Democratic Innovations." In *Evaluating Democratic Innovations*, eds. Kenneth Newton and Brigitte Geissel, 3–20. London: Routledge.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Preethi, S. 2020. "A Survey on Artificial Intelligence." *International Journal of Intelligent Computing and Technology (IJICT)* 3 (2): 39–42.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 OJ L, 2024/1689, 12.7.2024, ELI: http://data.europa.eu/eli/reg/2024/1689/oj

Reynolds, Emily. 2018. *The Agony of Sophia, the World's First Robot Citizen Condemned to a Lifeless Career in Marketing*. Wired. June 1, 2018. https://www.wired.com/story/sophia-robot-citizen-womens-rights-detriot-become-human-hanson-robotics/

Sandberg, Anders. 2013. "An Overview of Models of Technological Singularity." In *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, eds. Max More and Natasha Vita-More, 376–395. Malden, MA: Wiley-Blackwell.

Sartori, Giovanni. 1987. *The Theory of Democracy Revisited*. Chatam, NJ: Chatham House Publishers.

Schmidt, Andreas T. 2017. "The Power to Nudge." *American Political Science Review* 111 (2): 404–417. DOI: 10.1017/S0003055417000028

Schumpeter, Joseph A. 2018. *Capitalism, Socialism and Democracy*. Floyd, VA: Wilder Publications.

Sheikh, Haroon, Corien Prins, and Erik Schrijvers [Sheikh *et al*.]. 2023. "Artificial Intelligence: Definition and Background." In *Mission AI: The New System Technology*, eds. Haroon Sheikh, Corien Prins, and Erik Schrijvers, 3–28. Cham: Springer.

Thaler, Richard H., and Cass R. Sunstein. 2021. *Nudge: The Final Edition*. New York: Penguin.

Theocharis, Yannis, and Will Lowe. 2016. "Does Facebook Increase Political Participation? Evidence from a Field Experiment." *Information, Communication & Society* 19 (10): 1465–1486. DOI: 10.1080/1369118X.2015.1119871

Tirosh-Samuelson, Hava. 2012. "Transhumanism as a Secularist Faith." *Zygon*® 47 (4): 710–734. DOI: 10.1111/j.1467-9744.2012.01288.x

Trappl, Robert. 1992. "AI: Introduction, Paradigms, Applications (Including CBR), Impacts, Visions." In *Advanced Topics in Artificial Intelligence*, eds. Vladimir Marik, Olga Štepankova, and Robert Trappl, 1–24. Berlin: Springer-Verlag.

Turing, Alan M. 1950. "Computing Machinery and Intelligence." *Mind* 59 (236): 433–460.

Varoufakis, Yanis. 2023. *Technofeudalism: What Killed Capitalism*. London: Bodley Head.

Vimeo. 2015. "Does Democracy Work With AI?" August 5, 2015. https://vimeo.com/135493343

Warwick, Kevin. 2013. *Artificial Intelligence: The Basics*. London: Routledge.

Weizenbaum, Joseph. 1966. "ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine." *Communications of the ACM* 9 (1): 36–45.

Whitby, Blay. 1996. "The Turing Test: AI's Biggest Blind Alley?'." In *Machines and Thought: The Legacy of Alan Turing*, vol. 1, eds. Peter Millican and Andy Clark, 53–65. Oxford: Oxford University Press.

Winner, Langdon. 1980. "Do Artifacts Have Politics?" *Daedalus* 109 (1): 121–136.

Winner, Langdon. 2010. *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. Chicago: University of Chicago Press.

Wolkenstein, Fabio. 2022. "What Is Democratic Backsliding." *Constellations* 30 (3): 261–275. DOI: 10.1111/1467-8675.12627

Yudkowsky, Eliezer. 2008. "Artificial Intelligence as a Positive and Negative Factor in Global Risk." In *Global Catastrophic Risks*, eds. Nick Bostrom and Milan Ćirković, 308–345. Oxford: Oxford University Press.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

Ивана Дамњановић*

Факултет политичких наука,Универзитет у Београду

ВИ ПРОТИВ ДЕМОКРАТИЈЕ: САДАШЊОСТ И МОГУЋЕ БУДУЋНОСТИ**

Резиме

Предмет рада је сложен и динамичан однос између вештачке интелигенције (ВИ) и демократије, са тежиштем на тренутно постојећим применама ВИ и њиховим импликацијама како по само друштво тако и по друштвену, а пре свега политичку, теорију. Полазећи од чињенице да се ВИ све више интегрише у свакодневни живот и да расте интересовање академске заједнице за њене политичке последице, у раду се првенствено разматрају начини на које ВИ може да учврсти постојеће структуре моћи и поткопа демократске принципе. Користећи теорију демократије Роберта Дала као оквир, рад идентификује три основне демократске вредности које су угрожене развојем ВИ: једнакост, аутономију и јавну сферу. Системи ВИ, посебно они који се користе за предиктивне анализе и алгоритамско одлучивање, често репродукују друштвене и политичке неједнакости. Међу примерима су пристрасни алгоритми за изрицање затворских казни и микроциљано политичко оглашавање, који непропорционално погађају маргинализоване групе и негатино утичу на суштинску једнакост као претпоставку демократије и изборног процеса. Вештачка интелигенција угрожава персоналну аутономију и аутономију избора већ својим применама у технологијама надзора и алгоритмима за препоруке. Технологије препознавања лица и обраде великих скупова података омогућавају свеприсутан надзор, потенцијално обесхрабрујући политичко ангажовање и слободу изражавања. Алгоритми за препоруке и технике усмеравања суптилно манипулишу понашањем корисника, ограничавајући истински избор, што може да утиче на резултате

* Имејл: ivana.damnjanovic@fpn.bg.ac.rs; ORCID: 0000-0002-8553-492X

** Овај рад је подржан средствима Министарства науке, технолошког развоја и иновација Републике Србије на основу Уговора о реализацији и финансирању научноистраживачког рада НИО у 2025. години, број: 451-03-136/2025-03 од 27.01.2025. године.

избора. Овакве примене ВИ изазивају забринутост због ерозије личне слободе схваћене као аутономије, неопходног темеља грађанства у демократским државама. Јавна сфера, неопходна као простор за размену мишљења обавештених грађана, такође је угрожена. Алгоритми друштвених мрежа које покреће, или ће у најближој будућности покретати, ВИ фрагментишу јавни дискурс у ехо-коморе, подстичући поларизацију и смањујући могућности за делиберацију. Иако се на платформама друштвених мрежа ВИ већ користи за модерацију садржаја, није се показала ефикасном у сузбијању ових трендова. Један сегмент рада посвећен је и спекулацији о могућим будућностима које укључују општу вештачку интелигенцију (ОВИ) и вештачку суперинтелигенцију (ВСИ). ОВИ, дефинисана као ВИ способна да обавља било који задатак на истом нивоу као човек, отвара питања о укључивању у демократски процес. Ако ОВИ достигне ниво људске интелигенције, да ли би требало да добије право гласа или држављанство? У раду су размотрени неки од изазова дефинисања „демоса" у таквим сценаријима, укључујући питања телесности, репликације и аутономије. ВСИ, потенцијално последица самоусавршавања ОВИ, могла би надмашити људску интелигенцију и постати глобални суверен. Оваква верзија будућности подстиче на размишљања о апокалиптичним сценаријима краја људске врсте, политике и демократије. Чак и када би ВСИ не би угрожавала демократске системе, огромна когнитивна разлика између људи и суперинтелигентних ентитета могла би учинити традиционалне про-демократске аргументе застарелим. У закључку рада констатује се да негативни утицаји ВИ по демократију нису неизбежни и да се могу ублажити кроз демократску дебату и партиципативни дизајн, односно веће учешће демократске јавности у одлучивању о правцима развоја ВИ и њеним применама. Спекулативне расправе о ОВИ и ВСИ, иако корисне за тестирање неких поставки демократске теорије, не би требало да скрену пажњу са решавања стварних и актуелних проблема које изазивају већ постојећи ВИ системи.

Кључне речи: вештачка интелигенција, демократија, једнакост, аутономија, јавна сфера, општа вештачка интелигенција, вештачка суперинтелигенција

* This manuscript was submitted on September 14, 2025, and accepted by the Editorial Board for publishing on November 10, 2025.