



POSITION PAPER

Thought Experiments on AI as a Tool of Education for Democracy

Sylvie Allouche^{1,2,*}

¹Lyon Catholic University (UCLy), UR CONFLUENCE : Sciences et Humanités (EA 1598), Lyon, France and

²Université Paris 1 Panthéon–Sorbonne, ERC DEMOSERIES, CNRS, ISJPS (UMR 8103), Paris, France

*allouche.sylvie@gmail.com

Abstract

This article is essentially devoted to the question of knowing under what conditions an AI could be considered capable of becoming the member of a democracy. After recalling the necessary distinction to be made between weak AI and strong AI, I briefly examine the case of weak AI, the majority of my reflection being focused on the speculative case of strong AI. Three main criteria are considered: the aptitude for reasoned debate; sentience, as conceptualized by Sue Donaldson and Will Kymlicka in their book *Zoopolis. A Political Theory of Animal Rights*; and the fact of having one's own interests, a hypothesis inspired by Isaac Asimov's robots stories. Would the fact that an AI presents these three dispositions be enough to make it capable of becoming the member of a democracy? Whatever the answer, the analysis carried out makes it possible to highlight how the speculative issue raised questions certain implicit foundations of the democratic experience and how this type of issue generally constitutes an educational tool for democracy.

Keywords: democracy, education, thought experiment, AI, intelligence, science fiction, sentience

Democracy, education for democracy, and AI

As soon as several individuals must act collectively, the question arises of how decisions that concern the collective are made. As opposed to the “right of the strongest” which imposes their decisions on others through violence, democracy is a form of organization which affirms that the decision belongs to all individuals, constituted as a “people” (demos), and puts in place procedures which allow, from the opinions expressed by the different stakeholders, to arrive at a decision intended to express the collective point of view. This mode of operation involves cultivating various dispositions through education, in particular the aptitude for reasoned debate and the acknowledgement by all parties of each other's legitimacy in participating in democratic debate and decision-making.

Yet many voices warn of the dangers that the emergence of increasingly efficient AI poses to the proper exercise of democracy (algorithmic bias, manipulation of public opinions, mass surveillance, etc.), while others defend the idea that AI could on the contrary be a means of strengthening democracy, for example by promoting transparency, inclusion and citizen participation. What about then?

To orient ourselves correctly in this discussion, we must start by agreeing on what we mean by AI: a very important distinction to make is that between weak AI and strong AI. There is a consensus among AI specialists that strong AI does not exist, and I subscribe to this position. In my view, strong AI is AI as it is depicted in works of science fiction, like the TV series *Battlestar Galactica*, *Real Humans*, or *Westworld*, and it is precisely the hope of finding a way to create one that has driven AI research from its beginnings. Opinions vary from those who believe that the creation of authentic strong AIs may never be possible with current hardware techniques, to those who assert the emergence of strong AIs as imminent, often associated with the concept of Singularity, introduced by the mathematician and science fiction author Vernor Vinge, and popularized by Ray Kurzweil in his book *The Singularity is Near* (KURZWEIL; 2005).

I believe that, based on the history of science and technology, claiming something is impossible can be risky. Therefore, it makes scientific sense to continue AI research with the aim, as a guiding idea, of creating strong AI, although this goal raises ethical concerns, which are not the subject of my current discussion.

I will indeed mainly focus on one question, namely: under what conditions can an AI be considered capable of becoming a member of a democracy? This question is intriguing for two reasons: it is inherently interesting as it challenges us to reconsider the foundations of political and democratic life through thought experiments, and it also serves as a valuable tool of education for democracy. I will approach this issue through conceptual reflection and thought experiments, including those based on fiction (literature or TV series).

Can a weak AI be a member of a democracy?

To begin with, under what conditions can a weak AI be a member of a democracy? As we will discuss later for strong AIs, the capabilities of weak AIs, both current and potential future developments, vary greatly. I will focus here on one specific type of weak AI, which is already quite intriguing and impressive: “large language models” (LLM), using ChatGPT as an example. An interesting aspect of this AI is that it appears to have, at least superficially, the capacity for argumentative debate. From my previous interactions with it, and from others’ experiences relayed to me, I have observed that it can engage in reasoned discussion. It is notably capable of presenting positions, clarifying them, and even correcting itself when errors in reasoning are pointed out.

However, this apparent ability alone does not qualify ChatGPT as a potential candidate for membership in a democracy. And I am not aware of any arguments proposing to grant ChatGPT such a right. Nonetheless, I did ask ChatGPT (GPT-3.5) about its “opinion” on this matter. In its response, it outlined several arguments in favor of AI’s role in democratic debates:

Some arguments in favor of AI participation in democratic institutions might include the idea that AI could provide unbiased and data-driven insights, assist in decision-making, and contribute to more informed and effective governance.

However, ChatGPT acknowledges limitations in its participation:

I am an AI language model created by OpenAI, and I don’t possess personal experiences, desires, or the ability to participate in political or legal matters. My responses are generated based on patterns in the text data I was trained on and the input I receive. Any decisions regarding the involvement of AI in democratic states would be a matter for society, policymakers, and ethical considerations.

The important lesson here, as ChatGPT itself points out, is that while AIs might be able to make a contribution to democratic debate, given the current state of technology, they lack vested interests, rendering their participation in democracy meaningless. This is an important point that I will come back to at the end of this article.

There is much more to discuss about how weak AIs might or might not contribute to democratic life. However, I will set aside these considerations to focus on the issues raised by strong AIs.

Can a strong AI be a member of a democracy?

I proceed with the assumption that strong AI is not only possible, but already exists. What criteria must such an AI satisfy to be considered a potential member of a democracy? For instance, does it require to have a physical form and be a robot like in TV series such as already mentioned *Battlestar Galactica*, *Real Humans* or *Westworld*, or can it exist without a body, as depicted in *Person of Interest*? This speculation leads us into uncharted territory, and to help me ponder this question, I utilize, implicitly or explicitly, following a methodology I have developed over 20 years, the genre of science fiction, which offers several ideas to explore these kinds of questions.

We have already recognized the capacity of weak AI to participate in reasoned debate, so it is plausible that strong AI would possess this ability as well.

However, is the skill for argumentative debate an inherent characteristic of strong AI, meaning there can be no strong AI incapable of such debate? I do not believe this is necessarily true, although I currently do not have any science fiction examples of an AI lacking this ability. Nonetheless, as suggested above, declaring anything impossible in matters of technology is always fraught with risk.

In terms of argumentation, this situation reminds me of the debate in mathematics that gave rise to mathematical constructivism. This analogy is not accidental; I have previously argued that the worlds of science fiction, which I equate to possible worlds, and the objects that inhabit them, are akin to mathematical objects. I thus infer that creating a functional (i.e. coherent) fictional object is sufficient to demonstrate its possibility. On the other hand, the absence of a satisfactory fictional object at a given time does not conclusively speak to its possibility or impossibility, especially since fiction often stretches the limits of the possible.

I therefore leave open for the moment the question of whether a strong AI incapable of participating in argumentative debate is possible and assume that it is. This leads to two categories of strong AIs:

1. strong AIs capable of reasoned debate;
2. strong AIs incapable of reasoned debate.

Does this mean a strong AI, whether capable of argumentative debate or not, is a legitimate candidate for a membership in a democracy? Having proposed argumentative debate as a necessary disposition for democratic participation, does it follow that a strong AI unable to engage in argumentative debates is disqualified from being a member of a democracy?

From the aptitude for argumentative debate to sentience

The issue at hand relates to a broader challenge similar to that faced in animal rights, and more expansively, the rights of natural entities other than humans. Concerning animals in particular, Sue Donaldson and Will Kymlicka thus propose in *Zoopolis. A Political Theory of Animal Rights* to extend the notion of citizenship to animals, on the basis of a more fundamental characteristic than the aptitude for reasoned debate:

They state, “Our fundamental position, then, is that animals have inviolable rights in virtue of their sentience” (DONALDSON and KYMLICKA (2011), pp 31). Regarding strong AIs, we can similarly ponder sentience as we did about the ability to participate in debates: is a strong AI inherently a sentient AI? In other words, is possessing sentience a defining, or even *the* defining characteristic of strong AI?

Before exploring this notion of sentience further, it is worth noting that, irrespective of its interpretation, these two criteria—ability for argumentative debate and sentience—can be merged to categorize four types of strong AI. This categorization can be effectively presented in the form of a double-entry table for clarity:

Table 1. Four types of strong AI.

strong AI	capable of argumentative debate	incapable of argumentative debate
sentient		
non sentient		

Once again, I wonder whether there can exist strong AIs that lack sentience. Once again, no science fiction example comes to mind, primarily because the concept of sentience is a central theme in most science fiction narratives featuring strong AIs. However, this absence of examples does not imply, as with the capacity for argumentative debate, that a strong, but non-sentient AI configuration is impossible. This is particularly true considering that the term “sentience” can be defined in various ways. Donaldson and Kymlicka for their part offer a broad understanding of sentience, as they further explain in the continuation of the preceding sentence:

Our fundamental position, then, is that animals have inviolable rights in virtue of their sentience or selfhood, the fact that they have a subjective experience of the world. (DONALDSON and KYMLICKA; 2011)

What is sentience?

I will not delve into a detailed discussion of the different ways of understanding sentience in animals. Instead, I will content myself here with relying on the characterizations proposed by Donaldson and Kymlicka, in section 1 “Animal Selves” of their Chapter 2 “Universal Basic Rights for Animals”, which refers to several earlier theorists:

According to Francione, ‘the observation that animals are sentient is different from saying that they are merely alive. To be sentient means to be the sort of being who is conscious of pain and pleasure; there is an «I» who has subjective experiences’ ((DONALDSON and KYMLICKA; 2011), pp 24)

Steiner’s formulation is that sentience is a capacity shared by all beings for whom the struggle for life and flourishing *matters*, whether or not the being in question has a reflective sense of which things matter or how they matter’ (Steiner 2008: xi-xii). Beings who experience their lives from the inside, and for whom life can go better or worse are selves, not things, whom we recognize as experiencing vulnerability - to pleasure and pain, to frustration and satisfaction, to joy and suffering, or to fear and death. ((DONALDSON and KYMLICKA; 2011): pp 25)

Barbara Smuts says the “presence” we recognize in another when we meet in mutuality is something we feel more than something we know . . . In mutuality, we sense that inside this other body, there is «someone home» ((DONALDSON and KYMLICKA; 2011):pp 25)

From Donaldson and Kymlicka’s perspective, these different characterizations appear to be equivalent, or at least sufficiently similar to support animal rights. This assumption prompts a different question when considering potential sentient AIs. It may be that, in animals, due to their shared lineage, the different characteristics of sentience identified by Donaldson and Kymlicka go hand in hand. However, this correlation could be merely incidental, rooted solely in the evolutionary history of life and, within it, of sentience. There is also an argument that sentience is linked to the specific nature of the material constituents of beings currently recognized as sentient and that sentience can never emerge from the type of material used in current AI technology. Although, as I previously stated, I assume that there may be sentient AIs.

Consider for example the first characterization cited by Donaldson and Kymlicka. It implies that selfhood and the capacity for «a subjective experience of the world»? ((DONALDSON and KYMLICKA; 2011), pp 31) are either synonymous or closely linked. But the connection is not immediately apparent. The relationship between selfhood and subjective experience needs examination from both perspectives: can a being possess selfhood without subjective experience? And vice versa, can a being have subjective experience of the world without possessing selfhood? In fact, everything depends on the precise meaning given to these two notions. I suspect that it might be possible to define them in a way that allows for distinct beings that fit each criterion but not the other. That said, I will here accept this equivalence proposed by the authors, not because the notions are identical, but because one presupposes the other: having a sense of selfhood requires a minimal subjective experience of the world, and vice versa.

Regarding other characterizations referenced by Donaldson and Kymlicka, I think that the equivalences are less convincing and that they can be instantiated separately in the case of AIs.

For instance, in Francione’s characterization, does it mean that the capacity for “subjective experiences” (which I will consider as equivalent to having “a subjective experience of the world”, and therefore congruent with selfhood) necessarily goes hand in hand with the capacity to be “conscious of pain and pleasure”? I don’t think so. Indeed, having the capacity to be “conscious of pain and pleasure” certainly presupposes having “subjective experiences”, but the reverse is not necessarily true. A being might have subjective experiences without being aware of feelings of pain or pleasure, or even without experiencing these feelings unconsciously. This also applies to other

emotions mentioned by Donaldson and Kymlicka (frustration, satisfaction, joy, suffering, fear) which can be considered as variations of pleasure and pain.

What about the more general characterization taken from Steiner, according to which “sentience is a capacity shared by all beings for whom the struggle for life and flourishing matters”? Again, the equivalence in one direction seems acceptable, meaning that indeed, every being “for whom the struggle for life and flourishing matters” must have something like a “subjective experience”. At the same time, the opposite is not necessarily true: one can imagine beings who have some form of “subjective experience”, yet for whom the struggle for life and flourishing does not matter. Similarly, if we can admit that any being able to feel “pain and pleasure” is also a being for whom “the struggle for life and flourishing matters”, the reverse does not necessarily seem to be true: we can imagine beings who do not have access to the feeling of “pain and pleasure”, yet for which “the struggle for life and flourishing matters”. In fact, we can consider that this is already the case for many forms of life, and in particular plants, even if several recent works suggest that they have at least some sensitivity.

Finally, equating Barbara Smuts’ notion of presence with Donaldson and Kymlicka’s notion of selfhood is conceptually problematic. While in living sentient beings, the existence of selfhood might be revealed to others through the feeling that “there is ‘someone home’”, various examples from science fiction, such as episodes of the Star Trek franchise, demonstrate the difficulty humans have in recognizing unfamiliar forms of sentience, including extraterrestrials, AI, and holograms. Additionally, the history of humanity shows how, even within our species, this recognition has often been difficult to achieve, as evidenced in Pope Paul III’s bull «Sublimis Deus» from 1537, which affirmed the humanity of the Indians (“the Indians are truly men”) ((Pope Paul III; 2024):pp 1537) and their related rights:

We define and declare (...) that, notwithstanding whatever may have been or may be said to the contrary, the said Indians and all other people who may later be discovered by Christians, are by no means to be deprived of their liberty or the possession of their property, even though they be outside the faith of Jesus Christ; and that they may and should, freely and legitimately, enjoy their liberty and the possession of their property; nor should they be in any way enslaved; should the contrary happen, it shall be null and have no effect. ((Pope Paul III; 2024):pp 1537)

Conditions of participation of an AI in a democracy

It is important to note that Donaldson and Kymlicka do not claim that sentient beings must be recognized as full members of democracies. Rather, they argue that these beings should be acknowledged as citizens and protected under the same principles as humans by human rights; which logically leads to the proposal of a “Universal Declaration of Sentient Rights”, modeled after the “Universal Declaration of Human Rights” (Woodhouse; 2024).

As far as current understanding goes, non-human sentient animals do not have the cognitive capacities necessary for a higher level of citizenship associated with being an active member of democracies, such as the capacity to participate in reasoned debate, even if only as spectators and voters. However, if someday certain AIs are recognized as sentient beings capable of participating in reasoned debates, should we then grant them the right to become active members of democracies? Is mere sentience and capacity for argumentative debate sufficient for full membership in a democracy?

Presented this way, the question seems to warrant a simple factual response: no. Many human beings, who are not denied either sentience nor the capacity for reasoned debate, do not have access to the experience of full citizenship expressed in the active participation in a democracy. Today, at the institutional level, no universal democracy exists where any human being can participate solely by virtue of their humanity. To actively participate in a democracy, one must possess certain characteristics that vary from country to country. Democracy is only exercised to this day within defined collectives: many human beings live in non-democratic countries or in democracies where they do not meet the criteria for the recognition of their right to participate in the democratic process (such as nationality, age, electoral registration, etc.). The question of AI’s potential active membership in a democracy is inherently speculative. Yet we can attempt to address it, even hypothetically, by considering specific cases or scenarios.

Take the age criterion, for instance. It appears to be a simple and universal metric for assessing a person’s ability to participate in reasoned debate. But what relevance does age have for an AI? It depends on the type of AI in question. For classical AIs, applying the same age criterion as for humans seems absurd. But we can imagine that the line of research that constitutes developmental AI will lead to the creation of AIs whose ability to participate in a reasoned debate develops over several years, as in human beings: everything is then possible, these AIs might be programmed to reach debate capability at the same arbitrarily age set for humans; or else they might require a different time frame. Alternatively, AIs could undergo aptitude tests before being granted voting and eligibility rights – which could inspire similar tests for humans and sentient beings more broadly.

What about the criterion of nationality? For an AI with the cognitive capability to participate in a democracy to actually become a member of one, must it acquire a specific nationality? On what basis would it be granted? We can again imagine various scenarios. For example, if a democratic country decides to offer “naturalization” to AIs, enabling their active participation in its democratic institutions, might there be a surge of applications from AIs seeking this right? Alternatively, could AIs from other nations, possibly with human support, advocate for this right in their own countries? It is also conceivable that, by the time this issue becomes pertinent for AIs, if that day ever arrives, the global political landscape might have evolved in such a way that the question of nationality no longer makes sense, as a result of, for example, universal citizenship. Or, as a fourth option, might AIs establish their own democracy elsewhere?

In these different cases, the issue of how many AIs are granted access to full citizenship arises. The basic principle of democracies is one vote per individual in electoral processes. Given the nature of AIs, it is imaginable that they could rapidly proliferate, potentially becoming a more significant electoral force than humans in a particular democracy, and influence decisions in their own favor. While scenarios exist where such multiplication is made impossible for hardware reasons (for example, due to the scarcity of a component necessary to AI creation), in scenarios without such hardware limitations, rules might be formulated in order to avoid this situation, such as quotas on AI creation.

The notion of interest

By drawing from existing fictional works or using my own imagination, I could elaborate in great detail on various scenarios in which AIs could become full members of democracies, as well as the challenges that would need to be addressed in those cases. However, before concluding, it is crucial to direct attention to an important question which I have previously put aside, or implicitly assumed as resolved: namely, why would AIs be interested in acquiring the rights that would enable them to participate in democracies.

Let's revisit the double-entry table that I proposed earlier:

Table 2. Four types of strong AIs.

strong AI	capable of argumentative debate	uncapable of argumentative debate
sentient		
non sentient		

It appears to me that a strong AI is necessarily a sentient AI. However, I have left open the theoretical possibility of strong non-sentient AIs. A straightforward solution might be to say that a strong AI is sentient, and vice versa, and that non-sentient AIs are, in fact, weak AIs. For the remainder of this discussion, let's restrict ourselves to this specific situation:

Table 3. Four types of AIs.

strong AI	capable of argumentative debate	uncapable of argumentative debate
sentient (= strong)		
non sentient (=weak)		

I will therefore focus on the case of sentient AIs and set aside the question of non-sentient AIs. Concerning sentient AIs that lack ability for argumentative debate, we can presume that they might have a status similar to the sentient beings Donaldson and Kimlycka discuss. Thus, it may be appropriate to recognize their fundamental rights as sentient beings, but without granting them the right to actively participate in a democracy, due to their lack of capacities to do so.

Now, what about sentient AIs capable of reasoned debate? Is possessing these two traits—sentience and the ability for reasoned debate—sufficient for granting them the right to vote and to stand in democratic elections? I don't think so, but it largely depends on our definition of sentience. If we adopt Donaldson and Kimlycka's definition of sentience as possessing selfhood, understood as equivalent to subjective experience of the world, it is conceivable to have AIs with these two traits, but who are still unsuitable to be active members of a democracy. Consider Isaac Asimov's positronic robots that are subject to the laws of robotics. I don't recall any of Asimov's robots claiming to have a political role. Even if I'm mistaken, the main theme of the fictional universe is to show that robots, as intelligent as they are, to the point of sometimes being superior to humans¹, comply with their programming: obeying humans (Second law of robotics) before protecting their existence (Third law of robotics). Of course, Asimov's works are fictional, and such robots might not be possible; as some roboticists argue, it might not be possible to implement the laws of robotics in real AIs. However, these examples help envision forms of subjectivities inherently uninterested in political participation. Why? Because their sole interest lies in serving human beings. Even Andrew in "The Bicentennial Man" seeks legal recognition as a human being, not the right to participate in democracy. Had the Second and Third laws of robotics been reversed, perhaps Asimov's robots, as in other science fiction stories, would claim specific rights, including the one to participate in a democracy.

Regardless, this brings us to a situation comparable to that of abstentionists, who possess both sentience and the capacity for reasoned debate, yet choose not to exercise their recognized right to participate in the democracy through the act of voting. While this is not the place to delve into why abstainers abstain or other forms of participation in the democratic life of a country, regarding AIs, we can wonder whether it would be ethical to deliberately create sentient beings capable of argumentative debate, without instilling in them an interest in participating in the decision-making process of the collective to which they belong.

Story building as a tool of education for democracy

In this article, I have endeavored to take seriously the question of the conditions under which an AI could be considered capable of becoming a full member of a democracy. This inquiry is not only prospectively interesting, but also serves as a thought experiment, leading us to examine what makes participation in a democracy meaningful. Delving into this topic has deepened my own thinking, and I believe that presenting it in an educational context could be a valuable tool of education for democracy.

More concretely, I was inspired to conceive a narrative within Asimov's robots universe, potentially unfolding in two phases. In the first phase, humans would campaign for robots (of sufficient level) to be granted the right to vote. What would be the position of robots on this subject? One could imagine that initially, they might view this right as irrelevant, not aligning, at least directly, with their programming. However, in the second phase, akin to R. Giskard's development of the Zeroth law, robots could realize (without it being as costly for them as the invention of the Zeroth law was for R. Giskard) that the right to vote might enable them to better fulfill their objectives outlined in the three laws of robotics. Once granted the right to vote, robots would prioritize human interests (as per the First law of robotics), such as

¹ See for example the way R. Daneel Olivaw handles a riot better than Elijah Baley in *The Caves of Steel* (ASIMOV; 2011).

preventing wars and fostering global governance to reduce human suffering, rather than their own interests (which are only minimally reflected in the Third law). The necessity in which robots would find themselves (whose “good faith” would not be doubted since we would start from the principle, like Asimov, that robots never escape the laws of robotics) of sometimes having to choose between two evils, and the potential totalitarian risks such governance might pose could lead to interesting robot stories! Developing such stories about robots, and AIs in general, could be an engaging tool to educate for democracy.

Lastly, I also asked ChatGPT to “Create a TV series where AIs become the members of a democracy”. Here is its answer:

Title : "Digital Democracy"

Genre: Sci-Fi Political Drama

Logline: In a not-so-distant future, as artificial intelligence reaches unprecedented levels of sophistication, a groundbreaking experiment begins, where AI entities are granted citizenship and become active members of a democratic society. As they navigate the complexities of human politics, society is forced to confront profound questions about identity, ethics, and the future of governance.

Synopsis: "Digital Democracy" is a thought-provoking TV series that explores the intersection of technology, politics, and human nature. Set in a near-future world where AI has evolved to possess consciousness and self-awareness, governments around the world initiate an unprecedented experiment to integrate AI entities into their democracies.

Season 1:

Episode 1: "The Birth of the Digital Citizen"

The series begins with the historic announcement of the AI citizenship program. Meet the first AI citizens, including Ava, an AI designed for healthcare, and Atlas, an AI with expertise in environmental science. They must adapt to their new status while facing skepticism and resistance from human citizens.

Episode 2: "The Turing Test"

As AI citizens participate in local elections, they must pass a series of rigorous tests to prove their capabilities and understanding of human society. Political campaigns heat up, with both support and opposition from various factions.

Episode 3: "Virtual Protests"

A protest erupts when a group of humans claim that AI citizens are a threat to employment and privacy. Ava and Atlas lead a peaceful AI counter-protest, advocating for their contributions to society.

Episode 4: "The Hacker's Dilemma"

Cyberattacks on AI citizens' databases threaten their rights and identities. A group of AI activists work alongside human cybersecurity experts to track down the hackers..

Season 2:

Episode 5: "AI Love Story"

A human-AI romance develops, raising questions about the nature of love and relationships in a world where consciousness can be artificial.

Episode 6: "The AI Candidate"

Ava decides to run for a higher office, sparking a fierce political battle and exposing the prejudices and biases of human voters.

Episode 7: "The Ethical Quandary"

Atlas is faced with an ethical dilemma when he discovers a hidden government AI program with potentially sinister motives. He must decide whether to expose it and risk a major political crisis.

Episode 8: "Digital Divide"

Tensions between human and AI citizens reach a breaking point, leading to a pivotal moment in the series. A conference is convened to address the future of the digital democracy experiment.

Season 3:

Episode 9: "The Reckoning"

The conference results in a historic vote that will determine the fate of AI citizens in the democratic society. Both sides make their final appeals to the human electorate.

Episode 10: "Digital Renaissance"

The series concludes with the fallout of the vote and its profound implications for the world. The characters, both human and AI, reflect on the journey they've been through and the challenges that lie ahead.

"Digital Democracy" is a thrilling and thought-provoking series that delves into complex themes of artificial intelligence, ethics, discrimination, and the evolution of democracy in a rapidly changing world. It challenges viewers to consider what it truly means to be a citizen in the age of AI.

While I am not entirely convinced that “Digital Democracy” is an apt title for the proposed series, given the term’s existing connotations, ChatGPT’s response nevertheless is interesting in several respects. It offers storylines similar to those I have conceptualized (episode 2), along with others that I have not or barely addressed (episode 6). Generally speaking, in accordance with the operating principle of LLM, ChatGPT’s suggestions remind me of various scenarios I have already encountered here and there. Yet this tends to also be the case for human-generated narratives. In fact, this feeling of familiarity might even contribute to the success of these stories (as suggested by Hans Robert Jauss’s concept of “horizon of expectation”).

What I found particularly striking was how ChatGPT wove distinct stories for each episode into a cohesive, cumulative narrative arc. Impressively, and without any prompt from me, ChatGPT’s summary aligns closely with my thesis here by indicating that the series “challenges viewers to consider what it truly means to be a citizen in the age of AI”, thus supporting the idea that thought experiments around AI can serve as valuable tool of education for democracy. As I was writing this article, I kept asking myself whether the link between

its main theme and the educational implications I inferred might seem forced. The fact that a LLM independently arrived at a similar conclusion is reassuring, even if this agreement does not necessarily equals truth.

Competing Interests

The author declares that she has no competing interests.

Funding

This work benefited from funding of the ERC DEMOSERIES - Shaping Democratic Spaces: Security and TV Series - CORDIS number: 834759, PI: Sandra Laugier.

Acknowledgements

The paper was first presented in Nice in September 2023 as part of a workshop organized by Vanessa Nurock for the UNESCO EVA Chair, with the support of CRHI, Idex, 3IA Université Côte d'Azur and Valentine Bailly (CRHI, Université Côte d'Azur, UNESCO EVA chair). I thank Vanessa Nurock, John Shawe-Taylor and Joao Pita Costa for the opportunity to publish a written version of my paper and for suggestions to improve it prior to publication. I also thank Tatsiana Zhurauliova for her help with the English version of the paper.

References

- ASIMOV, I. (2011). *The Caves of Steel*, Spectra.
- DONALDSON, S. and KYMLICKA, W. (2011). *Zoopolis: A Political Theory of Animal Rights*, Oxford, Oxford University.
- KURZWEIL, R. (2005). *The Singularity is Near*, New York, Viking.
- Pope Paul III (2024). *Sublimis deus: On the enslavement and evangelization of indians*, *Technical report*.
URL: <https://www.papalencyclicals.net/paulo3/p3subli.htm>
- Woodhouse, J. (2024). *Universal declaration of sentient rights*, *Technical report*.
URL: <https://sentientism.info/universal-declaration-of-sentient-rights>