



POSITION PAPER

“Dare to Think” Education for Democracy with AI

Vanessa Nurock^{1,2,*}

¹Centre de Recherches en Histoire des Idées, UniCa and ²UNESCO EVA Chair

* vanessa.nurock@univ-cotedazur.fr

Abstract

This paper suggests to take seriously the idea, proposed by Le Cun, that AI could be seen at “a new phase of the Enlightenment”. It proposes the idea that an education for democracy with AI is possible under two conditions which are a basic feature of an ethics by design of AI and the two faces of the same coin: the expansion and enrichment rather than restriction and impoverishment of humanity at an individual, generational and cultural level. First, AI should neither be designed, nor developed or used for deskilling humanity but rather designed, developed and used for the preservation and extension of human capacities. Second, AI should not promote any form of technological colonialism but rather aim at preserving and extending cultural and intellectual diversity. If we fulfill these two important criteria, we can thus think about AI not in terms of alarming risks but rather in terms of a constructive future promoting the motto of the Enlightenment: “dare to think” and thus enabling an education for democracy.



Figure 1. Social media post by Yann LeCun on the amplification of Human intelligence with AI.

Keywords: Ethics by Design, Artificial Intelligence, Democracy, Deskilling, Techno-colonialism

“A new phase of Enlightenment”?

“By amplifying human intelligence, AI may cause a new Renaissance, perhaps a new phase of the Enlightenment,” tweeted Yan Le Cun on April 1, 2023, making it clear that this was “not an April fool’s joke”. This idea was echoed in the main French media (Le Monde; 4/28/2023): “With the help of AI, everyone’s intelligence and creativity will be amplified. This could lead to a new Age of Enlightenment.” This formulation by Yan Le Cun may come as a surprise: not only does he strongly refer to the Enlightenment –I will develop this point

at length later—, but he also uses the term “amplification” rather than “enhancement” most often, if not always, used in the context of AI and particularly valued by transhumanists. Even if I do not wish to overinterpret Le Cun’s quote, in the context of the reference to the Enlightenment, it is probably safe to suggest that this idea of amplification implies a non-aristocratic form of enlargement which would not be limited to our becoming machines.

An insistent reference to the Enlightenment might seem obvious because the AI “revolution” covers the same scope as the Enlightenment revolution. The Encyclopedia, for example, a collective, reasoned work integrating both the sciences and the arts, with illustrations, is now called Wikipedia. If Rousseau were alive today, he would probably consider that his book *The Confessions* is quite similar to Facebook, tik tok or Instagram, and that his masterpiece on education, *Emile* is being replaced by e-learning or even by robotic baby-sitters. Echoing his *Social Contract*, some are even promoting a Social Contract 2.0 or even seriously suggesting that we should give up our political power to Artificial Intelligences, supposedly more impartial than humans and therefore better able to govern, without even wondering whether empathy would be a skill required to fulfill the delegation of political power.

This Artificial Intelligence revolution is very similar to the Enlightenment Revolution in at least two respects. First, it has a wide impact on the technological, scientific, social, ethical, and political fields. Second, it concerns certain key issues of the Enlightenment, such as identity, autonomy, moral meaning, and the Social Contract. In spite of this similarity, if we are not careful, there is a risk that the AI revolution could lead to a deformation or inversion of the Enlightenment. This risk is a good motivation for developing safeguards to ensure that AIs that foster democracy are built and thus make sure that Yann Le Cun’s wish to see AI usher in a new phase of the Enlightenment is not doomed, which would be a pity. The aim of this paper is to use the tools of philosophy in general, and moral and political philosophy in particular, to shed light on the conditions needed to make this wish a reality.

The risk of inversion lies in a rather simple fact, at least *prima facie*. The Enlightenment led to the dissemination of philosophical ideas advocating emancipatory autonomy in the true sense of the word, i.e., the ability to choose one’s own standards and obey freely agreed rules to build a more egalitarian or even democratic society –despite undeniable bloodshed and persistent inequalities and injustices. The current revolution, on the contrary, is based above all on a technicist ideology that reproduces or reconstructs forms of hierarchy, while at the same time relying on forms of connection that can eclipse or even exclude or annihilate relationships. As Henry Kissinger wrote in 2005 in an article for *The Atlantic* titled *How the Enlightenment ends*: “The Enlightenment started with essentially philosophical insights spread by a new technology. Our period is moving in the opposite direction. It has generated a potentially dominating technology in search of a guiding philosophy” (2018, 14). “In short,” Kissinger points out, “it is as if we were looking to AI for clues to build a narrative, a guiding vision, and as if AI were to provide us with philosophical, ethical, and political principles.” To put it another way, it is as if ethics and politics were soluble in engineering and computer science, as if we could dispense with philosophers and leave it solely to engineers as well as computer scientists to propose new guiding principles in the social, ethical, and political fields. Contrary to Kissinger’s pessimistic diagnosis, I would suggest that interdisciplinary collaboration is needed for this type of undertaking. Also, it seems most important that the spirit, if not the letter, of the Enlightenment be understood and applied as emphasized by Immanuel Kant’s phrase “Dare to think”. The leading concern should not only be to avoid the risk of an inversion of the Enlightenment, but first and foremost to ensure the advent and actualization of an Enlightenment.

Technological and social experimentation

These last years, the dilution of ethics and politics in technology has occurred through experimentation with a growing number of applications and tools that incorporate AI, without always reflecting on their social, ethical, and political implications. AI has spread within our societies through a form of experimentation that is not only technical but also, and perhaps above all, social.

This social experimentation is veiled by the widespread idea that, once a technology is available, all we can do is make use of it, as suggested by what some have called “Gabor’s Law” (most often unaware that Dennis Gabor criticized precisely this kind of technicist position). This “Gabor’s law” asserts that “what can be done technically will necessarily be done,” and echoes philosopher Jacques Ellul’s (critical) formulation (Ellul; 1954) that “everything which is a technique is necessarily used as soon as it is available, without distinction of good and evil. This is the principal law of our age.”

In political terms, this maxim is in line with the classic neo-capitalist formulation of “There really is no choice,” perfectly expressed by Margaret Thatcher’s famous “So in a sense we do have to do it. Because There Really Is No Alternative,” sometimes abbreviated to TINA. The very essence of TINA is to constrain the field of possibilities by posing a form of dilemma and selecting a solution because we do not really have a choice, rather than because we find it desirable.

And yet, as (Gabor; 1963), the father of the so-called “Gabor’s Law”, wrote: “The future cannot be predicted, but futures can be invented.” Interestingly, he criticizes man’s capacity for invention when it takes on a destructive stance. He argues that we must not adopt the blissful attitude of passive acceptance of technology as inexorable, but rather invent our possible futures. He also asserts that these futures must be resolutely constructive in the sense of a “nurturing”. The idea that new technologies should be seen as an experiment is quite widespread today, particularly in the Netherlands. (Van de Poel; 2017), for example, proposes that emerging technologies should be seen as social experimentation. By this term, he means three things: firstly, experimentation in society (in real life); secondly, experimentation on society, which requires us to rethink certain normative frameworks (e.g., legal); and thirdly, experimentation by society. It seems to me that AI is a social experimentation in all three senses, and that this triple dimension highlights the fact that AI is reshaping our societies in complex ways. However, as Van de Poel rightly points out, the crucial problem here is to determine under what conditions this triple social experimentation can (or cannot) be morally acceptable.

One way to determine this is to introduce ethical reflection not only at the end of this triple social experimentation, but throughout the AI development process, rather than focusing on principles or consequences alone. This type of approach has been proposed since the 2020s from various angles as “Ethics by Design” approaches (Nurock et al.; 2021; Boger; 2022; Brey and Dainow; 2023).

In the limited space available in this article, I would like to propose two related and complementary criteria already mentioned in (Nurock et al.; 2021), which seem essential to ensure the amplification of human intelligence (and creativity) by AI in an ethically acceptable way and the fulfillment of “a new phase of Enlightenment”: first, the need to prevent deskilling, and second, the need for AI pluralism. I wish

to emphasize the extent to which these two issues, usually raised separately, must also be thought of together to allow this “amplification” conceived as a consolidation of skills we consider crucial from a moral point of view.

AI and deskilling

The question of deskilling is not new to philosophy. In some respects, it is similar to certain processes linked to automation likely to lead to a loss of intellectual and manual skills and of know-how, which were analyzed at length in a neo-Marxist vein in the 1970s. This process has also been denounced more recently in the field of AI and has been called “deskilling”.

We can certainly suggest that this decapacitation is not necessarily a (great) loss for humanity when it concerns alienating, even debilitating tasks that can be automated precisely because they are repetitive or do not require reflection and are therefore devalued. However, we may well wonder, firstly, who decides whether certain tasks should or should not be valued, and secondly, whether automation through AI always targets tasks that are worthless, or even devaluing.

On the contrary, it has been argued that this deskilling sometimes involves the loss of a capacity for pedagogical choice and reflection, as (Apple; 1994) shows. In the field of healthcare, deskilling may also be synonymous with the mechanization of care work and the loss of attention to certain symptoms, or the automation of diagnosis, for example, by nurses or doctors (Rinard; 1996; Davis; 2000). For this reason, it is probably legitimate to suggest that we should consider with tremendous concern the fact that we might simply lose certain important skills that we no longer exercise.

For instance, AI is more and more frequently used today to make moral or political decisions. It sometimes serves as an aid to decision-making, by proposing a wide variety of scenarios from which citizens or elected representatives can choose a number of options for debate, for example, but sometimes also as a decision-maker, by programming the “moral machines” that driverless vehicles are likely to become, for instance. We can legitimately wonder whether this moral and political deskilling might not pose a serious problem not only for our democracies but also, and above all, for education in democracy.

Artistotelian philosophy suggests that some of our most important skills, such as ethics, are acquired dispositions or hexis. (Vallor; 2015) claims that it is the moral imagination that must lead the way to “techniques of the self,” understood in a Foucauldian sense as much as in the sense of contemporary technologies that are the medium of our current relationship to the world. She suggests that “we retain the ability to imagine and develop new and alternative forms of these relations, even as our present powers of moral imagination continue to be shaped by our existing technology relations. The future trajectory of human moral abilities is technologically conditioned, but not technologically determined.”

As (Jaffro; 2007) points out concerning moral education in an analysis relevant for education for democracy, “self-practices are not to be equated with techniques, but rather with dispositions and habits or, by extension, with techniques whose essential and natural use is to produce such dispositions and habits.” And he continues: “However, self-practices do not belong to a second type of techniques, but to an entirely different kind of thing –which I call ‘gymnastics.’” With his illuminating comparison between our moral capacity and jogging, (Jaffro; 2007) underlines the importance of wondering what kind of self-practice is likely to lead not just to any kind of transformation, but to the right kind of self-transformation –that is, self-improvement based on knowledge–. He suggests two essential elements of such gymnastics. The first is that it is an exercise, and the second is that it is an exercise of will supported by reason, itself guided by a moral conception. He therefore concludes that: “Gymnastics cannot be reduced to a mere technology.”

This kind of gymnastics, then, goes further than the idea of a simple acquired disposition, and, in any case, is quite distinct from the moral routine that is the basis of automated morality. It also presupposes regularity, care, and attention (sometimes almost unconscious) to “little things” as well as to the whole process, which is inevitably lost when we stop practicing it. Furthermore, it is useful to analyze deskilling not only on the ontogenetic level of the individual, but also on the phylogenetic level of the species. Just as certain languages are lost if they are no longer spoken and become “dead” languages –either because they are replaced by others, or because the populations who spoke these languages have disappeared– it is highly probable that, after a certain number of generations, the skills transferred to AIs would not be passed on anymore and would probably disappear in the course of evolution. The problem, then, would not simply be the loss of know-how, but rather the loss of skills that could be important not only from a technical or cognitive perspective, but also from an ethical and political perspective. When deskilling occurs, it becomes more challenging to educate for democracy.

Should we then, in a technophobic stance, as Yann Le Cun fears, promote a “new obscurantism” that would fearfully block the development of AI? In the opposition between, on the one hand, this new phase of Enlightenment and, on the other, the denunciation of this obscurantism, we may find ourselves stuck in a binary dilemma, between Scylla and Charybdis. We might be able to get away from this alternative by proposing a genuine amplification of our intelligence, not by AI but with AI. In the words of Daniel Andler, perhaps it would be appropriate to promote a form of “bilingualism” that would enable us to combine two skills: the ability to think and act with AI, and the ability to think and act without AI, or the ability to use AI in a variety of ways so as not to lose our skills. In this way, the amplification of our intelligence and creativity promoted by Yan Le Cun would not be associated with deskilling. This “bilingualism” exercised in social, ethical, and political fields could thus combine AI and education for democracy.

AI and education for pluralism

We can then focus on the question of skill loss and gain, deskilling and upskilling, not only from the point of view of individuals, of the species, as we have just seen, but also from the point of view of populations, states, and cultures. If we wish to examine under what conditions AI and education for democracy can and must go hand in hand, then we need to consider the rights of people, whatever their culture or generation, to decide for themselves. To put it another way, the issue of education for democracy brings to the fore the question of AI’s inclusivity.

Even though AI is often presented as a “view from nowhere”, scientific, objective, and neutral, these characteristics are far from self-evident, as confirmed by the debates on AI biases that grew after the publication of (O’Neil; 2016). Moreover, it is now commonly suggested that these technologies reproduce relations of economic, social, and cultural domination, with the risk of a return to colonial

empires or the development of new colonial empires, linked to the development of Artificial Intelligence. In this perspective (Miaillhe; 2018) bluntly asserts that the USA and China will dominate the world thanks to AI, that Europe is threatened with “cyber-vassalization” and Africa with “cyber-colonization”. More recently, in an interview at the beginning of January 2024, Seydina Moussa Ndiaye, lecturer at the Université Numérique Cheikh Hamidou Kane in Senegal and member of the UN’s AI Advisory Committee, considered, in the same vein, that the greatest risk currently posed by AI to the African continent is that of colonization (UN News; 4.1.2024). This danger is not only one of the causes why certain communities and countries mistrust the development of AI, but it also poses a fundamental threat to education for democracy, which cannot legitimately be based on a techno-colonialist foundation but should on the contrary develop in an inclusive way with the participation of all stakeholders of different cultures. Since the turn of the 2020s, the idea of techno-colonialism has emerged with respect to four main issues: “click workers”; data; the extraction of essential resources and the environmental consequences of AI (which extend beyond the techno-colonial dimension); and, last but not least, the worldviews underlying the development of AI, which represent a form of cultural, social, ethical, and political techno-colonialism. While the imagination and vocabulary attached to AI give the impression of an immaterial reality, “online” or in the cloud, the materiality of AI production reveals practices that (Mejias and Couldry; 2024) sum up with a shocking formula: “Explore, expand, exploit, exterminate.”

The first three points have been widely developed in recent years (see (Mejias and Couldry; 2024) for a recent review). In the limited framework of this article, I will therefore confine myself to the last point, on which philosophy can, I hope, provide food for thought.

If we really want AI to help us expand our intelligence and creativity, to use Yan Le Cun’s terms, one of the central difficulties is to ensure that it is part of an open cultural repertoire that does not deskill languages and cultures, to use the expression developed in the previous section. In other words, in this matter too, we need to adopt an approach that is not only bilingual or plurilingual, but also pluricultural. One of the difficulties for doing so lies in the way in which both the development of AI and the critique of this development are based on culturally localized and dominant concepts and values, which makes it difficult not to “speak” in the terms in which this worldview is embedded, much less to criticize it. The notion of the “coded gaze”, proposed by Buolamwini in reference to both the “colonial gaze” and the “male gaze”, reflects a worldview and power relations rooted in the domination exercised by the code or AI. The unique feature of this coded gaze is that it passes through an artifact, which is a machine that serves both for seeing and for prescribing, whether explicitly (for example, when AI is used in the legal field) or implicitly (when search engines show only masculine figures when searching for doctors or CEOs).

However, as suggested in (Nurock; 2024, in press), this coded gaze is, most of the time, at least in Europe where I stand, a coded US gaze. The risk of exporting the “US gaze” is added to the risk of the coded gaze, whether in the biases themselves or in the solutions proposed to avoid them. For example, the notion of “sharing”, which is valued to the extreme in American culture, as (John; 2017) has shown, is posited as a value per se. This cultural tendency is probably reinforced by the “open curtains” culture, which holds that if you have nothing to hide, then you should show it. It is also reinforced by the post-2001 surveillance culture, which holds that if you have nothing to hide, then you have nothing to fear, as (Solove; 2011) points out. However, this “sharing” is not the same as open source, but rather giving up your data, our data. (Varon and Peña; 2021) have clearly shown that the question of consent involves a balance of power that is very tricky because it is binary and tantamount to an all-or-nothing choice. Refusal to consent means to exclude oneself from a globalized society where data “sharing” is posited as the norm.

Conversely, the analyses and theoretical tools used to analyze and criticize AI most often come from American culture or are disseminated via American culture. This is particularly true when considering biases, particularly racial ones, which are part of a specific cultural and historical framework. So, when we discuss “human-centric” AI, as opposed to “technology-centric” AI, we often tend to forget that the “human” we are talking about is ethnocentric and focused on American culture –speaking from where I stand in Europe–, and doubtless based on Chinese culture for those living elsewhere. It is therefore urgent, as some authors, and even organizations such as UNESCO in its 2021 recommendation on the ethics of AI (UNESCO; 2021), have already suggested, to open up the cultural repertoire of AI and introduce genuine pluralism, which is an essential ingredient of education for democracy. This open approach is also supported by the manifesto “Indigenous Data Sovereignty and Governance” (2022) of the Global Indigenous Data Alliance formalized by Carroll, Cummins and Martinez, and summarized by the acronym CARE (Collective benefit, Authority to control, Responsibility, and Ethics). It explicitly complements the acronym FAIR (Findable, Accessible, Interoperable, Reusable) and is based on principles that promote open access and not just “sharing”. However, the main point here is not only to promote certain principles of inclusion even if it is important. As (Roche et al.; 1996) make clear, the main challenge is not only the necessity to include cultural elements to diversify the data offered as input to AI, but rather to structurally rethink how to reshape the development of (Gwagwa et al.; 2022), for example, suggest the integration into AI thinking of the notion of Ubuntu, which expresses the interrelation between individuals and a sense of community more profound than the “sharing” approach, and could renew the way we consider connections.

“Sapere aude”, dare to think. This is how Immanuel Kant summed up the thinking of the Enlightenment. If we want AI development to include education for democracy as an essential feature, we cannot afford to deskill individuals, generations, and cultures. Of course, this is not the only criterion required, but it does help us think of AI not only in terms of restrictive risks, but also in terms of constructive openness to new possibilities.

Acknowledgements

Vanessa Nurock is grateful to Yvonne van der Does of the Université Côte d’Azur Office of International Scientific Visibility (IdEx) for English language editing of the manuscript and to John Shawe-Taylor for some suggestions to help me clarify the English version of this paper. This work was supported by the French government through the France 2030 investment plan managed by the National Research Agency (ANR), as part of the Initiative of Excellence of Université Côte d’Azur under reference number ANR-15-IDEX-0.

References

- Apple, M. W. (1994). Computers and the deskilling of teaching, *Computer Professionals for Social Responsibility Newsletter* 12(2).
URL: <http://cpsr.org/prevsite/publications/newsletters/issues/1994/Spring1994/apple.html/>

- Boger, J. (2022). Culture change, human-centered design, and ethical by design as transactional cornerstone concepts in the development of technology for supporting aging, *Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments* pp. 556–561.
- Brey, P. and Dainow, B. (2023). Ethics by design for artificial intelligence, *AI Ethics* .
URL: (advance online publication) <https://doi.org/10.1007/s43681-023-00330-4>
- Davis, A. (2000). Outsourced radiology: will doctors be deskilled?, *British Medical Journal* 337: 542.
- Ellul, J. (1954). *La technique ou l'enjeu du siècle*, Paris, Librairie Armand /The Technological Society, (J. Wilkinson, trans.), Vintage Books.
- Gabor, D. (1963). 1964, *Inventing the Future*, New York: Knopf.
- Gwagwa, A., Kazim, E. and Hilliard, A. (2022). The role of the african value of ubuntu in global AI inclusion discourse: A normative ethics perspective, *Patterns* 3(4): 100462.
- Jaffro, L. (2007). L'exercice moral est-il assimilable à une technique ? une critique de michel foucault, *Revue d'Auvergne* 585: 73–96.
- John, N. (2017). *The age of sharing*, Cambridge: Polity Press.
- Le Monde (4/28/2023). Yann Le Cun, directeur à Meta : « L'idée même de vouloir ralentir la recherche sur l'IA s'apparente à un nouvel obscurantisme ».
URL: https://www.lemonde.fr/idees/article/2023/04/28/yann-le-cun-directeur-a-meta-l-idee-meme-de-vouloir-ralentir-la-recherche-sur-l-ia-s-apparente-a-un-nouvel-obscurantisme_6171338_3232.html
- Mejias, U. A. and Couldry, N. (2024). *Data Grab: The new Colonialism of Big Tech and how to fight back*, Penguin Random House, WH Allen.
- Miaillhe, N. (2018). Géopolitique de l'intelligence artificielle, le retour des empires., *Politique Etrangère* 3: 105–117.
- Nurock, V. (2024, in press). *What Does 'Ethical by Design' Mean?*, Enjeux éthiques et métaéthiques de l'Intelligence Artificielle « Où va la philosophie française » ed. I. Alfandary, S. Laugier, R. Zagury-Orly, Paris, Vrin.
- Nurock, V., Chatila, R. and Parizeau, M.-H. (2021). *What Does 'Ethical by Design' Mean?*, In: Reflections on Artificial Intelligence for Humanity ed. By B. Braunschweig and M. Khallab, Springer.
- O'Neil, C. (2016). *Weapons of Mass Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Crown.
- Rinard, R. (1996). Technology, deskilling and nurses: the impact of the technologically changing environment, *Advances in Nursing Science* 18 (4).
- Roche, C., Wall, P. and Lewis, D. (1996). Ethics and diversity in artificial intelligence policies, strategies and initiatives, *AI Ethics* pp. 1–21.
- Solove, D. J. (2011). *Nothing to hide: the false tradeoff between privacy and security*, New Haven, Yale University Press.
- UN News (4.1.2024). Interview: Ai expert warns of 'digital colonization' in Africa.
URL: <https://news.un.org/en/story/2024/01/1144342>
- UNESCO (2021). Recommandation sur l'éthique de l'intelligence artificielle.
URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455_fre
- Vallor, S. (2015). Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character, *Philos. Technol.* 28: 107–124.
- Van de Poel, I. (2017). Society as a laboratory to experiment with new technologies, In: Bowman D. M., Stokes A. R., and Rip A. *Embedding New Technologies into Society: A Regulatory, Ethical and Societal Perspective*, Pan pp. 61–87.
- Varon, J. and Peña, P. (2021). Artificial intelligence and consent: a feminist anti-colonial critique, *Internet Policy Review* 10(4).