The challenges of artificial intelligence from a sociotechnical, critical perspective

ROSER PUJADAS
University College London
r.pujadas@ucl.ac.uk

1. From the reification of artificial intelligence to the challenges of so-called *artificial intelligence technologies*

The term *artificial intelligence* (AI) represents an ambition which goes as far back as Ancient Greece, involving the development of a technology that could imitate or surpass human cognition. This seems a misguided tenet, given that human intelligences are multiple and social, but it has encouraged a range of innovations over the years. However, public debate around so-called *AI technologies* has only recently started, particularly since the launch in 2022 of OpenAI's ChatGPT with open access. This was followed by a range of alarmist discourses about the dangers posed by AI, expressed by AI experts and technology leaders, including Elon Musk, one of the founders of OpenAI. They even requested a moratorium to be able to prepare for the coming risks associated with AI. Ironically, those responsible for the development of AI were referring to existential risks in a future in which the ubiquity of AI is seen as unquestionable.

AI technologies present opportunities and challenges, but as suggested by critical studies of AI (Crawford, 2021; Suchman, 2023), it is worth considering the work that certain conceptions and discourses of AI do, who produces them and with what effects, and what questions are not addressed. AI seems to give coherence and stability to a phenomenon which is vague and distributed. It is not a monolithic technology but rather one that requires other diverse technologies and social practices, one that manifests itself in different ways. Technologies under the umbrella term *AI* have been used for a while in many contexts, and, knowingly or not, we interact with them regularly. For instance, AI is used for disease detection, weather forecasting, facial recognition, selection of candidates in recruitment processes, the production and detection of fake news, or the personalization and classification of content and offers in web portals. These technologies are diverse, and the contexts of use are varied; therefore, any generalisations about

the effects of AI, good or bad, hide more than they help answer, and tend to naturalise dominant conceptions of AI.

This reification tends to reduce AI and its potentials and risks to technological matters, and to imbue AI with intrinsic agency (Suchman, 2023). In addition, both utopian and dystopian visions of AI, expressed by leaders of the technological sector (mostly men, white, wealthy, and situated in Silicon Valley), may distract us from the real problems AI is already causing. Journalists, activists, and academics have provided empirical evidence, among others, of AI's role in the erosion of privacy, the perpetuation of inequalities, racial and gender bias and discrimination, disinformation and the production of fake news and fake images, behavioural and ideological manipulation, power concentration, and environmental harm (Broussard, 2023; Crawford, 2021; Eubanks, 2018). These are not only technical issues: they are structural and environmental, and disproportionally affect certain groups which are underrepresented in the technological sector.

2. The ethico-onto-epistemology of AI: When technology and knowledge configure realities and values

Science and technology studies (STS) evidence that technology does not have an inherent agency or intrinsic force that makes its evolution unavoidable. There are many possible versions of technologies and their implementations, as their design and use is situated in specific social contexts. Technological development results from negotiations and power relations, and from specific material conditions. They are not the outcome of single geniuses and they are reliant on tasks that are frequently invisibilised. At the same time, technology is an important social and political agent that embeds and can help reproduce certain values, significantly influencing social structures and practices. Therefore, it is important that researchers and society scrutinise its development and applications.

In the case of AI this is particularly true, given the significant performativity of these technologies. That is, their capacity to (re)configure reality is considerable. On the one hand, AI can be deployed to assist complex human task, but it is frequently used to automate decisions and actions. Particularly in this latter case, there is a delegation of agency towards a technology which is becoming increasingly opaque. On the other hand, even though AI tends to be presented as an objective technique, it involves a specific vision of "intelligence" and of how to predict accurately. As STS and feminism have revealed, knowledge is situated and partial. Furthermore, knowledge practices constitute sociotechnical assemblages (Latour, 2005) and are constitutive of certain configurations of reality and value systems, with political and ethical implications. Therefore, we need to consider what the ethico-onto-epistemologies (Barad, 2007) of AI actually are. That is, what are the epistemologies and knowledge practices associated with

AI and what realities and values do they help configure? And with what ethical and material consequences?

3. AI models based on big data and extractivism

Nowadays the term AI tends to refer to technologies incorporating machine learning (ML) techniques. In a nutshell, the development of an ML system involves using training datasets to produce algorithms and statistical models to represent an aspect of reality. The use of ML consists in applying this model to analyse automatically new data and make predictions. The assumption that a higher quantity of training data results in more accurate models, in addition to the ambition to produce AI technologies for general use (such as ChatGPT), has led to the massive and indiscriminate use of data. This is problematic for several reasons.

Despite the presumed objectivity and universal application of these models, the idea that a bigger volume of data guarantees a more accurate representation of reality is a fallacy. Like any kind of knowledge production, any model of AI is partial and has limitations, but it can be useful in specific contexts of use. For instance, the large language model ChatGPT makes statistical predictions of the likelihood of a certain string of words to answer a query. It might be useful to improve the style of a text, although when asked to answer a question, it will be able to provide a text that is grammatically correct but not necessarily factual.

In addition, it raises ethical issues and social justice concerns. On the one hand, certain practices of data extraction or acquisition, for instance from web portals, violate basic rights such as privacy or copyrights, while appropriating intellectual work and data. On the other, this massive collection of data is not necessarily representative, the data are decontextualized, and the quality of the data might not be good. Furthermore, datasets may contain the social biases that exist in the context of extraction. If so, these biases or incorrections will be embedded in the ML models and reproduced in its predictions or recommendations. If such technology is used in decision-making, those decisions will be unfair and may increase inequalities. There is plenty of evidence of AI technologies in use that reproduce biases, with damaging implications for vulnerable groups (Eubanks, 2018). This is aggravated when ML models require the labelling of training data (Crawford, 2021), such as those that identify images or videos. Labelling involves the imposition of categories, which are never neutral, and the subjective classification by workers, who are frequently low-paid and in some cases exposed to disturbing material, such as violent images (Gray & Suri, 2019).

Not only does such a paradigm of AI reproduce class differences, but it also amplifies structures of discrimination and unequal distribution of capital at a global level. The capacity to accumulate and process big data, which is necessary to generate these models of ML, demands considerable resources: natural, technological, economic, and of intellectual capital. These are only

available to a small number of technology companies, with monopolistic tendencies that are mainly located in USA and China, two countries which have adopted protectionist policies with the aim to win the race to lead AI innovation (Rikap & Lundvall, 2021). Finally, the environmental effects are noteworthy. The huge computation power needed to process big data and generate models consumes lots of energy, as well as water to cool data centres. In addition, technological components require a range of minerals which are frequently extracted in the Global South.

AI is politics by other means, but rather undemocratic, particularly in an economic system in which the rule of the strongest prevails. The inertia of an extractive version of AI – extractive of data, labour, and natural resources (Crawford, 2021) – seems difficult to stop. We will need bold public policies and determined social activism to face the complex challenges (economic, epistemic, ethic, social, politic, and environmental) posed by AI and we must rethink which versions of AI we wish to promote.

References

Barad, Karen M. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham, NC: Duke University Press.

Broussard, Meredith (2023). *More than a Glitch: Confronting Race, Gender, and Ability Bias in Tech*. Cambridge, Massachusetts: The MIT Press.

Crawford, Kate (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.

Eubanks, Virginia (2018). *Automating Inequality*. New York: St Martin's Press.

Gray, Mary L.; Suri, Siddharth (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston: Houghton Mifflin Harcourt.

Latour, Bruno (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford; New York: Oxford University Press.

Rikap, Cecilia; Lundvall, Bengt-Åke (2021). *The Digital Innovation Race: Conceptualizing the Emerging New World Order*. Cham: Palgrave Macmillan.

Suchman, Lucy (2023). "The Uncontroversial 'Thingness' of AI". *Big Data & Society*, 10 (2), pp. 1-5.