The Role of Culture in the Intelligence of AI

Mercedes Bunz

Artificial Intelligence (AI) clearly suffers from its name, which easily leads to misunderstandings. The name suggests that its intelligence is like human intelligence, only 'artificial'; it would have been far better and more precise to call it 'machine intelligence'. Neural networks, the technology that is in part the foundation of the current boom and facilitated deep learning approaches, adds to this, as it is a term that further confuses the discussion. The term suggests that machine learning systems are built on 'neurons' that operate just like those in the brain. If you ask experts in the field, however, they will quickly explain that biological neurons function very differently. They are much more complex than the mathematical functions we find in the nodes of a human neural network; this is the case internally (they seem to transmit signals that are chemical, but can also be electromagnetic or operate on ion channels), as well as externally with respect to their architecture.

So why do we hold on to these misleading descriptions? Unfortunately, pretending that the technology is 'inspired' by human biology seemed to be an easy way to persuade everyone to believe that the technology will work at some point. Of course, it makes sense to a human that something inorganic is becoming intelligent when it imitates 'the master'. Only that machine intelligence functions very differently. Machine learning (ML) models are composed of multiple processing layers that analyse and learn representations of data with multiple levels of abstraction (LeCun/Benigo/Hinton 2015). The model starts on the lowest layer by looking at, for example, pixel formations, while subsequent layers configure more complex features, that is, forms typical for the dataset from those pixel formations. This bottom-up approach starting with the smallest entity enables the models to discover 'intricate structures in large datasets' (LeCun/Benigo/Hinton 2015, 436).

The intelligence we find here is therefore rather particular: the model learns stochastically by adding up very small elements to calculate 'a bigger picture' or (in the case of large language models) to calculate the meaning of a sentence from analysing the context of thousands of tokens (entities similar to words), taking note of which other tokens are in nearby vectors. This means that ML models operate on a very different level than human intelligence. They have very particular abilities, but also make very specific mistakes due to the way they 'look' at our digitized world—images and texts, for instance, or videos.

Possible Roles of Archives and Museums in Machine Learning Development

Cultural heritage institutions like archives and museums have important roles to play with respect to the question of how to counter the developments in machine intelligence. First of all, there is a lot of knowledge held in and around cultural archives from which machine learning could benefit, and these two areas should thus become more interconnected. For example, cultural archive studies are well informed about approaches to decolonizing the archive, which is relevant when it comes to training machine learning systems. Here, the development of ML systems can thus benefit to a great extent from established debates around cultural archives.

Then, of course, there is a material overlap between archives and the training data for ML systems. The cycle of hype around AI currently focusses mainly on the general learning of very large ML systems. But that is not necessarily always the best approach. Smaller, very specific datasets can be used to train and specialize machine learning systems, and archives can and should have a big role to play in this. For it is true: ML systems are a different approach to finding knowledge, and the basis for their knowledge is the big data that is the archive, which means that their fates are intertwined.

We are also only starting to learn what kind of knowledge we can find in archives by using the stochastic antennas of ML systems. Our cultural understanding of what we might want from the new approaches to knowing is still in development. In literature, we have debated 'distant reading', which Franco Moretti (2013) described in excellent detail at a quite early point in time. The term describes a method in literary studies that applies computational analysis to a large collection of texts with the aim of identifying patterns within the text collection. As it looks at multiple texts, it is often conceived as the opposite of a 'close reading' of one particular work. At the moment, we are still working on ways to understand what 'distant seeing' might mean for art history, though Leonardo Impett (2020; also Azar/Cox/Impett 2021) has made a start here by trying to measure, compare, or analyse gesture across large sets of images.

Another important point is the role of museums as well as contemporary art institutions in educating people about our new technologies of knowledge, which we unfortunately call 'artificial intelligence'. There are currently no places in Western societies where we allow people to test, reflect on, and playfully understand those technologies. Citizens encounter AI technologies as users and AI is presented as a functioning service. This is problematic since these technologies are going to be integrated on quite a large scale in analysing, operating, working, and categorizing, side by side with us humans. Only that we rarely get to look under the hood and learn about how they function in acquiring AI literacy. There are not many places in our societies where citizens can meet and encounter this technology now, while the so-called 'black box' is being opened and experimented with. Museums and cultural institutions are such places where inquiring artworks and cultural experiments facilitate such encounters.

Critique of AI

One of the problems we face when it comes to critiquing AI is that it is needed on so many levels—Kate Crawford's book *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* from 2021 is a very good example here. Critiques of AI are confronted with the confusing need to be everywhere at once: the computational material that AI runs on requires the extraction of rare minerals, which often come from developing countries, where miners work under horrible conditions. AI is trained on labelled images, a categorization that is again outsourced to developing countries or organized through crowdsourcing using platforms such as Amazon's Mechanical Turk, and, as such, is based on precarious work. Even if this work is organized with care, datasets are often biased, and models amplify biases further (Chun 2021). The training of machine learning models is energy-intensive and problematic for the environment. The workplaces involved in developing AI systems struggle with diversity issues. Any participatory moment, any integration of citizens into the development of AI technology early on is thus lacking, even in a regulation such as the European Union's AI Act, which was conceived to tackle AI critically (Bunz/Vrikki 2022). All these points force critiques to look everywhere all at once, which therefore makes critique difficult. Even if we take a step back, the critique nonetheless does not come to an end: the critiques listed above also must be critiqued.

Because when you read through the list above, it becomes clear that no one with a critical mind would be interested in starting to work in this field—and that is a problem. The messy situation made visible by the critiques above, drives people away from interacting and engaging with the technology. What could attract anyone to confront such a big mess? This is the biggest problem at the moment, because the result is a real danger: the danger that the people developing this technology are generally dull and non-reflective individuals, who simply want to get rich (Aradau/Bunz 2022).

It is already possible to see the effects of this: The main question we seem to be asking in Western societies is how technology can assist the future of businesses. Not much attention is, however, given to the debate on how it might foster public infrastructures, and there is no alternative version of a public AI, that is, the development of an AI funded by public resources. The main attention is given to commercial players, even though there are exciting projects such as LAION, which provides large datasets to democratize the ability to train models; or even Stability AI, a company that embraces the idea of opensource for generative AI and has worked in the past with the Ludwig Maximilian University Munich. What are also missing are approaches to how citizens might participate in the development and implementation of the technology that will categorize them in the future.

Machine learning models have some amazing skills. But the questions we need to ask are: How can we start to implement that infrastructure in the best way possible? And how can we make it a technology that serves our society? A critical theory of AI therefore needs to go further than merely pointing out the flaws; we need to participate in its development actively, and not simply criticize it. Here, I am very attracted to what Phil Agre (1997) called a 'critical technical practice', in which different views and interests as well as practices regarding AI come together—computer scientists, cultural critics, citizens, sociologists, government officials, et cetera—with the goal of producing technologies together. This is important, despite the knowledge that this journey will not be easy.

The most productive and effective attitude is to be aware of all the downsides, but to retain your curiosity and interest, and your will to play a part in shaping AI. I wrote about this different approach of having a critical practice with respect to technology in an open access text (Bunz 2022), because this also means that we need to develop a different attitude towards technology in general. A productive critical attitude means leaving behind the idea of technology as an 'instrument' that needs to function and is supposed to serve us. It asks us if we are willing to adopt a more collaborative approach that includes a more profound engagement with it through asking 'uncomfortable' questions such as: How is this AI configured? What are the technical reasons and challenges behind configuring it like this? What situations are produced by this configuration? Which situations have been forgotten and ignored, and which ones cannot be addressed by it? But also: what labour was needed to configure it like this and how can we make some of this technical labour better and fairer?

Different Logics

My own main interest is understanding the different logics of AI through machine learning models. I think it is important to know that machine learning models are not just able to process data, and also to get something of an idea of *how* they process that data. The systems are not just fed with thousands of text documents or images; they also break each of those documents or images up further into smaller and smaller entities. They approach our world by calculating those very small entities or elements with their stochastic logic. This is the case for both image recognition and language modelling. When AI looks at images, it examines pixel constellations; when it looks at language it analyses tokens, the elements that make up words. Putting these smallest elements in relation to each other and finding a typical pattern is the way AI models learn—in connection with images, they use pixel edges, lines, colour changes, and shadings to find similar patterns, which on a higher level can then be identified as leaves, fur, or a right angle, in order to construct a motif such as a bush, cat, or door.

Getting these elements right is where its particular skill lies, but also where it makes mistakes—I am fascinated by adversarials (Buckner 2020), which are data constellations that induce machine learning models to make erroneous predictions and categorizations, but are imperceptible to us humans. At the beginning, we thought that adversarials are created by people to attack the machine—say, to hide the pixel constellation of a dog in an image of weapons in order to confuse the machine, which then cannot identify the weapons. By now, however, computer scientists have found thousands of natural images that machine learning systems constantly get wrong.

I think it is important to be aware of this logic or approach to our world, to know where AI is extremely helpful and where it is bound to make mistakes. AI is, unfortunately, a general-purpose technology, and this means that the changes and transformations in each field are different and require their own form of assessment. What worries me most is that the knowledge and infrastructure is mainly in the hands of big corporations and that governments are not doing enough to ensure that public infrastructures keep pace with current developments. We have already started employing AI in healthcare, city administration, policing, and education. This is not just problematic because we rent AI from big businesses. What is more problematic is that this denotes a transfer of knowledge, knowledge that was previously linked to public infrastructures is now being handled by AI services. This means that this knowledge is being transferred from being public to being private and commercial, and this is worrying. Besides the need to regulate AI systems with respect to risks, there is also the need to ensure that technical knowledge remains public. I hope we will have the intelligence to do so!

Potentials and Perspectives

Personally, I see potential for AI in helping us deal with an information-intense world—whether images, texts, or data. AI systems are very effective in analysing information and spotting certain trends. We read and write more than any other generation before us: there are more publications, but there is also more work or personal communication on the multiple channels and platforms available to us on a daily basis in both our social and work lives, continually notifying us of messages that have been received and should be answered. My colleague Matthew Kirschenbaum (2023) has warned that AI might lead to a 'textpocalypse', an ever-growing stream of generated content. But I think that this was already happening before AI models such as ChatGPT started to generate writing. AI could thus help to sort and summarize information that is relevant for us. One trend I would like to see when it comes to generating AI is more research on watermarking texts and images generated by AI; I think that would be helpful. Overall, I see that we have started to understand AI as a collaborator, in other words, as a system that collaborates with humans in a loop. This is a step forward from understanding AI as a system of automation replacing the human.

References

Agre, Philipp E. (1997). Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. In: Geoffrey Bowker/Susan Leigh Star/Les Gasser et al. (Eds.). Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide. New York, Psychology Press, 131–57. https://doi.org/10.4324/9781315805849 (all URLs here accessed in August 2023).

Aradau, Claudia/Bunz, Mercedes (2022). Dismantling the Apparatus of Domination? Left Critiques of AI. Radical Philosophy 212, 10–18. Available online: https://www.radicalphilosophy.com/article/dismantling-the-apparatus-of-domination.

Azar, Mitra/Cox, Geoff/Impett, Leonardo (2021). Introduction: Ways of Machine Seeing. AI & SOCIETY 36 (4), 1093–104. https://doi.org/10.1007/s00146-020-01124-6.

Buckner, Cameron (2020). Understanding Adversarial Examples Requires a Theory of Artefacts for Deep Learning. Nature Machine Intelligence 2 (12), 731–36. https://doi.org/10.1038/s42256-020-00266-y.

Bunz, Mercedes (2022). How Not to Be Governed Like That by Our Digital Technologies. In: Kathrin Thiele/Birgit Mara Kaiser/Timothy O'Leary (Eds.). The Ends of Critique. Methods, Institutions, Politics. Lanham, Rowman & Littlefield, 179–200. Available online: https://rowman.com/webdocs/theendsofcritiquepdf.pdf.

Bunz, Mercedes/Vrikki, Photini (2022). From Big to Democratic Data: Why the Rise of AI Needs Data Solidarity. In: Michael Filimowicz (Ed.). Democratic Frontiers. Taylor & Francis. https://library.oapen.org/handle/20.500.12657/57277.

Chun, Wendy (2021). Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition. Cambridge, MA, The MIT Press. https://doi.org/10.7551/mitpress/14050.001.0001.

Crawford, Kate (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. New Haven/London, Yale University Press. https://doi.org/10.1007/s00146-022-01488-x.

Impett, Leonardo (2020). Analyzing Gesture in Digital Art History. In: Kathryn Brown (Ed.). The Routledge Companion to Digital Humanities and Art History. New York/London, Routledge, 386–407. https://doi.org/10.4324/9780429505188-33.

Kirschenbaum, Matthew (2023). Prepare for the Textpocalypse. The Atlantic, 8 March 2023. https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318/.

LeCun, Yann/Bengio, Yoshua /Geoffrey Hinton (2015). Deep Learning. Nature 521.7553, 436–44. https://doi.org/10.1038/nature14539.

Moretti, Franco (2013). Distant Reading. London/New York, Verso. https://doi.org/10.3366/ccs.2013.0105.