September 2023

AI as complex sociotechnical systems: Problems, approaches and reflections

Valerie Hafez | University of Vienna, Department of Science and Technology Studies; Women in AI Austria[1]

Rania Wazir | leiwand.ai

Fariba Karimi | Technical University Graz; Complexity Science Hub Vienna

Fariba Karimi and Rania Wazir acknowledge the funding from WWTF roadmap grant number RO22-002.

[1] Valerie Hafez works for the Austrian telecommunications regulator RTR. Her contribution as an independent researcher and member of Women in AI Austria remains her own work and does not reflect or prejudge the opinions or activities of RTR or other related independent regulatory bodies.

Introduction

AI systems are complex sociotechnical systems – that is, they consist of material and social components which, by being put into particular kinds of relations, work together in specific ways. Consequently, it is not sufficient to understand AI systems as isolated lines of code – instead, AI systems should be understood as intertwined with data, computational power, storage, market relations, organizational ontologies, societal practices, and epistemic capabilities. In effect, AI systems are not isolated from the various 'other' issues often thought of as tangential, such as “those pesky humans that get in the way of AI systems functioning properly.”

In this paper, we first outline some of the limitations of AI systems from a data science perspective. While many of these issues have been discussed before, they provide a fundamental lens to understanding the principles upon which these technologies are built, within the confines of 'the code itself'. This includes questioning the techniques used for developing AI systems, common confusions around the interpretation or application of data and methods, and concerns about the use of particular parameters for making decisions (i.e. about how to design a system).

Building on this discussion, in the second section we detail how AI systems cannot be contained as merely technical issues, instead brimming with complexity. The issues of 'the code itself' cannot be regarded merely as technical faults with technological fixes - in fact, such an understanding is inherently problematic. However, understanding the wider issues only becomes possible with a grasp of the data and computer science fundamentals, both their strengths and limitations, because it is precisely these mechanisms which lead to particular kinds of issues in contexts of different scales and scope. In this section, we explore matters of power, scale and structure, as well as the value(s) we enact with AI systems.

With this basis, we move on in the third section to raise some questions and offer a series of indications for dealing with AI systems. We seek to draw in wider questions and contemplations to round off our reflection on the implications of complex, sociotechnical AI systems in our world. Understanding AI systems as complex sociotechnical systems unfolds pathways to address existing issues, and we argue that different ways of thinking will be key to handling the challenges to governance posed by AI systems in very particular contexts. In the boxes rounding off this paper, we display a small selection of further concerns which informed our approach and arguments.

To conclude, we briefly touch upon governance - briefly only, for we hope that the thoughts and ideas that we have provided will feed into collaborative efforts to make AI systems (with) care for our world.

1.Diving into the technical problems with AI systems

We have become used to considering quantitative reasoning and technology-based decisions as objective and therefore superior to alternative approaches. Such a view is dangerous because it simultaneously overestimates and idealizes what AI systems are capable of. Hence, we would like to address some of the issues with AI systems from a technical point of view, while maintaining sensitivity to the qualities of AI systems that cannot be reduced to 'purely' technical issues.

1. What is the basis for making a decision?

When the criteria for making a decision are completely opaque, it is impossible to determine whether there is some logical process taking place, or whether the decision-making is, in fact, completely arbitrary – based on the equivalent of a coin-toss. This actually has nothing to do with automation, or autonomy, or AI. It could be a simple set of rules that someone wrote down. Or a formula that someone pulled out of a hat (This kind of pseudo-math is particularly insidious: just because there is a formula, it is assumed that everything is mathematical and hence scientific/objective. Very few people bother to ask how that formula was derived – in particular, what assumptions lie behind it). So lack of transparency about the decision-making process, and pseudo-objectivity, is a more general problem and not restricted to AI systems. What is, however, an issue for AI (or actually, any software-encoding of decision-making processes), is that the decision-making criteria can be hidden behind trade secrets; and, where machine learning is concerned, the data which is part of the "criteria-creation" process can be protected by copyrights or even GDPR. All of this, while making the decisions *seem objective* to the non-expert – because, after all, it's all code and math.

2. Correlation is not causation.

Machine learning is simply a matter of finding *correlations* in data. So under what conditions is a decision based on correlation admissible? Is it legitimate to make an important decision affecting a person (such as employment, parole, education opportunity, credit or insurance, asylum), based on predictions that derive from statistical correlations – especially correlations with protected characteristics?[2] Shouldn't such decisions be based on the principle of ***causation***? This is not to say that all decisions based on correlation are bad (if trying to decide whether or not to smoke, the high correlation between smoking and various forms of cancer could be a good reason for refraining). But we *should* be thinking carefully about when (as in, for which types of decisions, and involving which features) correlations are admissible.

3. The intended characteristic and its measurable proxy: is the operationalization scientifically sound?

For example: much emotion detection software is based on physical markers (such as a person's facial expression). But the connection between the outwardly visible and the internal state is highly

[2] It is important to note that, even if protected characteristics are not directly used by a machine learning system, they can often directly or indirectly affect other, non-protected characteristics which *are* used as input to the system. This leads to system decisions that nonetheless correlate with the protected characteristics.

controversial, and lacks a sound scientific basis. Should decisions be based on the 21st Century equivalent of Phrenology?

4. Does the AI system even work?

Even if the AI system is based on scientifically sound principles – without transparency, we lack the proof that the system *actually works.* For an in-depth critique of this blind faith approach, see The Fallacy of AI Functionality.[3] Furthermore, we need to come to terms with what exactly we mean by "*the system works*". Again, this is an issue that plagues not just AI systems, but processes and methods in general: do we have quality control measures in place? But in the case of AI systems, such considerations are often neglected; if at all, the easy-to-measure option of "accuracy" is often resorted to – without regard to fundamental questions such as "accuracy based on which test", and "is accuracy even the correct measure to use".[4]

5. Is predicted behavior the correct basis for decision-making?

When should we instead be trying to understand how to change conditions in order to improve the future outcome? For example: Many teachers will have faced this conundrum - is the test supposed to be used to measure the student's performance (and in some way, satisfy the need to classify them as good or bad students) – or is the test supposed to be used as a tool to understand gaps in understanding, and concepts that students have grasped particularly well, in order to tailor future learning activities better? Another example: the Austrian Employment Service (AMS) algorithm. It used predictions of future employability in order to decide on training opportunities for the unemployed. In particular, it used those predictions to deny opportunities for those that the algorithm classified as "hopeless". The use of such a system (even assuming that the predictions are correct), accepts the employment landscape as an accomplished fact – and perpetuates it. Is this what we expect from our social institutions? Or would we rather try to better understand what were the obstacles to employment (and yes, absolutely, use algorithms as a tool to assist us in studying the labor market), in order to try to remove or mitigate them? In the words of Arvind Narayanan, speaking about a similar algorithm used to predict recidivism, "in the criminal-risk prediction scenario, the decision that we make based on predictions is to deny bail or parole, but if we move out of the predictive setting, we might ask, "What is the best way to rehabilitate this person into society and decrease the chance that they will commit another crime?" It opens up the possibility of a much wider set of interventions.

[3] See also: https://predictive-optimization.cs.princeton.edu/

[4] For a discussion on accuracy, its meaning in various communities, and examples of when it's the wrong measure to use, see: https://www.leiwand.ai/blog/ai-comedy-of-errors

2.Handling AI systems as complex and sociotechnical phenomena

As we have previously argued, AI systems should be more correctly understood as sociotechnical systems spanning across a range of spaces, contexts, issues and practices. Our concern in this section is to alight on how these sociotechnical aspects matter under specific circumstances, and how thinking of AI systems as a complex socio-technical entity invites us to take into consideration factors beyond their programming and design as part and parcel of what AI systems are, and what they do.

6. The codification and scaling up of discriminatory structures.

The great risk with AI lies in its great power: it encodes a procedure, a decision-making process, and then scales it up – applying it to the masses. An interesting case is currently being debated in the US courts: an AI system that is being used in hiring decisions by many companies.[5] Once it has rejected a particular candidate for a position at one company, it is not clear (because the decision-making criteria and the construction of the system are protected by trade secrets – although now that the case has landed in court, hopefully some disclosure will be required) whether that means it will reject this candidate for a similar position at all other companies. If this is the case, this person has been condemned to unemployment by software – because the software is the same at all the companies. If instead a human were making the decisions, it would be a different human at each company, and the chances would be higher that these humans would not all react in the same way to the candidate (this procedure refers to the wisdom of crowd concept). The great weakness of humans is, in this case, also a strength: we are idiosyncratic, and can make unexpected decisions. When it comes to systemic bias and discrimination, the fact that not everyone behaves accordingly all the time is what can give people from at-risk groups a chance.

7. AI systems could do better.

Building an AI system forces us to encode our decision criteria (to the extent that this is possible). If system designers know that they have to be transparent, this might encourage them to be more careful and deliberate in choosing the decision criteria. If transparency and testing were the rule, we would actually know more about decisions and have more data (input/output for testing and validation), than would often be the case with human decision-making. We could then determine if the decision-making criteria were acceptable (as in points 2-4 above). In those cases where the answer is yes, the decision-making criteria would be transparent and validated, and logging could be put in place for constant monitoring. This process of making weights and decision criteria part of a deliberative process offers a pathway for practicing power-sensitive tactics in AI development. For example, Shakir Mohamed, Marie-Therese Png and William Isaac suggest imbuing critical technical practice with insights and theoretical approaches from decolonial theory to engage in reverse and reciprocal tutelage, or to develop AI systems in solidarity with negatively affected communities. From this perspective, the development of AI

[5] "In February 2023, a class action lawsuit was filed against Workday, Inc. in the Northern District of California, alleging that the company had offered its customers biased applicant screening tools which resulted in racial, age, and disability discrimination." https://www.law.umaryland.edu/content/articles/name-660254-en.html

systems can become a site for practicing the insights of decolonial, feminist, queer and otherwise othered theory.

8. AI systems are built on and also structure power relations.

From the humans labeling the data[6] to the CXOs reaping gains, AI technologies require labour and yield labour that can be harnessed for profit (see also Karen Hao, Heidi Swart, Andrea Paola Hernández, and Nadine Freischlad's series on AI colonialism for MIT Technology Review). For this reason, Claudia Aradau and Mercedez Bunz argue that "Rather than seeing AI as a high-tech autonomous weapons system that is a killer robot, or an automated facial recognition system – i.e. as a coherent 'thing' Suchman cautioned against – AI is a distributed socio-technical system that is always already produced, circulated, maintained and repaired through dispersed, intensive and underpaid labour" (Aradau & Bunz, 2022). In fact, AI systems reconfigure questions of labour because the very proposition advanced by AI providers – to offer services – shapes the labour relations in the organisation to which these AI systems are provided. Organisations using AI systems may need to hire employees with skills in the field, or set up liaisons with external providers if they rely solely on the support of AI providers. (Organisational) path dependence (which can also be framed in terms of market power), value extraction from labour and the distribution of wealth stemming from the use of AI systems are related to the ways in which these systems are built and deployed, as Meredith Whittaker powerfully demonstrates in her analysis of the adjoint histories of plantation labour and AI systems.

9. Scale matters.

AI systems have the potential to circulate (their outputs) very quickly due to a high degree of automation and high potential for scale. This means that AI systems can achieve, relatively quickly, a state in which they become infrastructural to other processes, practices, fields and materials. Consider the case mentioned above of unemployment by software. The key problem is not only that an automated system makes a decision which affects an applicant's life negatively – the problem emerges from the widespread use of the same system, from the difficulty of changing systems once it has been implemented, from the decision-making power of the company owning the AI system and the affected person's lack of capacity to contest the effects of the intricately complex sociotechnical system. At the same time, the potential harms at any of these levels will be different precisely because of the effects of scale: AI systems simultaneously affect individuals and (different kinds of) groups in distinct ways. Unless we account for scale, we risk reducing issues with AI to merely 'technical' problems, thus ignoring the sociotechnical complexity of AI systems. Scale, therefore, requires us to be aware of different effects at different levels and develop instruments capable of handling such complexity, e.g. the multi-scale ethics framework proposed by Melanie Smallman.

10. AI systems contain, create and extract value.

As we have tried to show, AI systems raise questions about value which are not yet adequately or even equitably resolved. Take generative AI, for instance. Generative AI works well when the data it is based on is *human*-generated. As recently shown by Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolar Papernot and Ross Anderson, generative AI systems experience model collapse when too much AI-generated data is introduced to them (reported on by Carl Franzen). Effectively, this means that in

[6] See, for example, https://time.com/6247678/openai-chatgpt-kenya-workers/

order for generative AI to create value for its owners, it depends on a steady stream of human labour. Through being put into use, AI systems not only embed flows of information, create categories or execute functions – they also embed flows of value. In this sense, we need to gain a deeper understanding of how the value that AI systems accrue, generate or extract, is distributed. And crucially, we need to reach just decisions as a society about who gets to participate and how in determining these allocations.

3. Implications and pathways for further reflection

The heterogeneous environment we have described above calls for diverse approaches and expertise. How sociotechnical systems are set up and put into use affects a wider public, and hence governance should not be restricted to elite expert groups and powerful lobbies.

11. Diversity of disciplines and epistemologies is key to managing AI systems

Instead of leaving decisions solely to engineers, data scientists, marketing managers, legal advisers or CEOs, the questions of how to monitor AI systems, develop adequate responses to risks, or even discontinue the use of AI systems in case of harm require multiple and very diverse forms of expertise. Usually, claims such as this are followed by a call for more ethics specialists, but we suggest going beyond this solution. Aside from involving a variety of disciplines and laypersons, we propose that dealing with AI systems also requires epistemological diversity (see Box 1).

12. We need to be cautious of how we think about AI systems

Just like other sociotechnical systems, AI systems can mean different things in different contexts. For this reason, we need to be very careful when it comes to defining AI systems as belonging to a particular category: the matter of definition is not innocent because it has particular effects on the world. As Kate Crawford has argued, defining AI systems in a particular matter does work – it changes how we frame the questions raised by AI and which kinds of responses we develop. For instance, defining AI systems as *products*, like in the AI Act, comes with a host of implications about what can be done about AI systems, which frameworks are applicable to AI systems, and which approaches can be considered a model for how to deal with AI systems. It is a pragmatic solution to label AI systems as long as we consider them products; were we to consider them *infrastructures* (see Box 2), the proposal to provide a packaging note informing of risks and side effects would probably be considered more controversial. Klaus Hoeyer takes a similar approach when arguing that more attention should be paid to the analogies we use for describing and explaining the role of data in our societies: when describing data as the new oil, we make a set of assumptions about what data is (=naturally occurring resource that requires extraction), what data does (=power our devices and societies), and what we can or should do about data (=extract it!). If we were instead to think of data in the same terms we think of pharmaceuticals, we would arrive at completely different conclusions: data would be thought of as always produced through complex processes, causing harm and benefit equally depending on the circumstances of the affected persons, and therefore necessarily subjected to stringent controls. All this goes to say that how we define AI systems is tied to which kinds of regimes of governance we develop.

13. Some AI systems are more than products.

Right above, we argued that although AI systems are currently framed as products through the approach taken by the AI Act, other framings are possible and even necessary. In part, we could see this necessity play out in the AI Act, where both the European Council and the European Parliament amended rules for, respectively, general purpose AI or foundation models (see Box 3). Both of these legislative bodies felt the need to appropriately account for AI systems, which, although potentially of an immense scale, can be integrated into a variety of applications for any number of purposes. They did so by loosening the

regulatory requirements for AI systems of this type because the product framework, hinging on intended purposes, areas of usage and manufacturer-user relations, simply does not make sense for AI systems like large language models, generative AI systems, or search algorithms. Yet other aspects remain unaddressed. How, for instance, do we address the enclosure of epistemic capacity - truly our ability to create knowledge about the world - which proprietary AI systems enable? OpenAI's discontinuation of Codex, a proprietary large language model used by researchers, brought awareness to the crisis of reproducibility that ensues in academia.[7] But lack of stability affects business as well, when integrated versions of AI systems are no longer available or stable versions are offered only for short periods of time. The tying of cloud and AI services, which is already common, continues to proliferate, posing challenges to the ability of smaller companies to retain their independence from larger providers and overcome the prohibitive challenge of developing their own open-source AI models. At the same time, the dependence of researchers and other companies on the outputs of AI systems indicates that it will be very difficult to switch to alternative providers and also raises issues of service quality, availability and access alongside competition concerns. In this context, how do we ensure that AI systems with such large-scale infrastructural qualities such as foundation models remain open for everyone, usable and sufficiently understandable? Foundation models in particular are notorious for their environmental impact: even though information is scarce and claims are not always comparable, the GHG emissions and materials consumption (including water) associated with AI systems are significant and appear to be increasing. In a world of finite resources where we can no longer rely on simply creating more alternatives, how do we ensure that innovative technologies are accessible to and benefit all?

Focusing on the 'merely technical' product (e.g. image generation) obscures how the AI system becomes an obligatory passage point for various flows of people, things, information, value and others. These infrastructural considerations call for appropriate regulatory responses which should put democratic considerations and participatory processes at the heart of their efforts.

14. Rethinking human oversight.

We are well aware of the cognitive biases that effectively limit the efficacy of individualised human oversight processes, such as the classic human-in-the-loop, human-on-the-loop or human-in-control set-ups. Considering the issues we have raised above, we would like to advance a daring proposal: perhaps what we need to develop to make human oversight more than a mere item on a checklist are forms of collective oversight. Strengthening collective, collaborative and commons-inspired approaches to AI governance could be a measure to distribute power more equitably amongst those who, in one way or another, become part of the sociotechnical assemblage formed by AI systems. Such an approach could also offer a clearer analytical and practical lens for the question of participation in AI systems. If we take seriously the fact that every user and every affected person participate in the AI system in a way which leads to its improvement, increase in scale or scope, we need to find ways to acknowledge this participation in a socially equitable manner and design governance structures which take due account of this relationship.

[7] https://www.aisnakeoil.com/p/openais-policies-hinder-reproducible

4.Conclusion

Taking our observations in this section together, we believe it is necessary to pay close attention to efforts to govern AI systems in particular ways. This concerns overtly political interventions such as laws as well as less obvious initiatives such as standards (both those agreed in technical committees and those set unilaterally by companies through interoperability or other specifications). Which kind of approach these governance efforts take is significant: depending on the approach, certain effects of AI systems will fall through the cracks. It is both a matter of research to find meaningful ways of ensuring collective decision-making when the need arises – that is, ways for people who are affected by, contribute to or are otherwise entangled with AI systems to participate in shaping what these systems do and the conditions under which they are put to use –, but also a matter of policy for implementing and enforcing these meaningful ways of making decisions about automated decision-making systems. Necessarily, proposals for governance need to take questions of scale, power, discrimination, and distribution as well as the criteria for our decision-making into account; in effect, both local specificities as well as large-scale system-level effects need to be addressed through adequate tools and instruments.

Harms can come in many shapes, sizes and scopes. Like in other fields, what is considered a harm in the field of AI is hotly contested, although the contesting sides are anything but equal. We can see this regularly in controversies about the effects of technological applications, like in the case of the effects of neonicotinoids on honey bee hive health described by Sainath Suryanarayanan and Daniel Lee Kleinman. Key to this controversy is the question how harms are defined: do you detect harm in experiments, i.e. controlled settings, with defined risk and harm levels? This is what industrial toxicologists do, and this is why the Environmental Protection Agency of the USA focuses on false-negative standards. But then the honey bee colony collapse disorder became a frequent problem, and commercial beekepers as well as more complexity-oriented researchers outside the USA, observe harm over time, in uncontrolled settings, began to call for limiting the use of neonicotinoids to protect honey bee health.

When we do not know which effects will come from a technology, we need to find ways to deal with the unknown. There are different approaches to doing so, as Stefan Böschen, Karin Kastenhofer, Ina Rust, Jens Soentgen and Peter Wehlig observed. One approach relies on historical experience and predefined risk categories, often formed on the basis of positivist testing. This approach tends to quantify risks and currently provides the dominant framework for thinking about risk. Yet this *control-oriented approach* struggles with scenarios in which multiple failures occur simultaneously, where we do not have sufficient experience to define risks or where causes of failure are difficult to identify in advance – and in this sense, it is more accurately a framework for handling known unknowns. This is the approach chosen by the AI Act: risks are defined in advance, procedures are put in place to mitigate these risks, and anything not foreseen in advance falls through the cracks.

However, there are at least two other epistemological frameworks for dealing with unknown unknowns – and thereby also for defining risks and harms. The *complexity-oriented framework* is highly attuned to the interrelations that shape developments at system level in important, unpredictable and usually unforeseen ways. As such, complexity-oriented approaches are by default concerned with system-level interactions as well as identifying and understanding patterns emerging from these interrelations. This epistemological framework recognises the potential of harms emerging from long-term processes or in a cumulative manner, for instance in the field of epigenetics: instead of looking at short-term harm in isolated settings, harms can emerge through complex interactions within a system. *Case-focused epistemologies*, on the other hand, share the same attention to emergence, interrelation and complexity as complexity-oriented approaches. However, their level of observation starts from the specifics of one single or very few cases to explore these effects at a local, circumscribed level. One example of this would be clinical medicine, which is concerned with formulating wider observations based on in-depth observations of individual case histories.

These and other epistemological frameworks can, and should, complement each other if we strive for robust and effective policy frameworks. This may mean developing complexity-oriented monitoring mechanisms to record subtle changes over time after AI systems have been put to use in a particular context. It could also mean taking evidence from single cases into account when the harm evidenced by one case is widespread or resistant to change on its own account.

Box 1: Why we need diverse approaches to knowing harms and unknown unknowns

In the social sciences, infrastructures are understood both as "things but also the relation between things" (Brian Larkin, 2013, 329): that is, you can observe a road as built matter, but this built matter alone tells you little about what this road does and means. Instead, we need to look at how the road is used, which places it connects, why and how it was built, who maintains and/or controls it, and which kinds of effects it produces locally and at system-level. All of these matters are significant because they affect which kind of infrastructure we are dealing with.

Susan Leigh Star and Karen Ruhleder describe the most salient characteristics of infrastructures as:

- embeddedness into other (social) assemblages,
- capable of supporting all further actions without the need to reinvent it,
- reaching beyond a single site or event,
- learned as part of membership,
- linked with conventions of practice,
- embodying standards,
- built on an installed base, and
- becoming visible when it breaks down.

They develop their framework by studying the development and set-up of a distributed digital research infrastructure. Which is an important point to make, because when we think of infrastructures, we usually think of road, energy, telecommunications and other networks, but less often of (less tangible) sea roads, trade networks, or software like operating systems. Looking at their characterisation, many of the themes we raised begin to come together. Thinking of AI systems as infrastructures means remaining attentive to the ongoing flows of information, labour and value which are enabled and disabled by these systems – much like Kate Crawford has done in her seminal book, An Atlas of AI.

This does not mean that all AI systems are infrastructures: determining whether and which kind of infrastructure AI systems have become is rather an empirical matter. To return to Brian Larkin, "what distinguishes infrastructures from technologies is that they are objects that create the grounds on which other objects operate, and when they do so they operate as systems" (Larkin, 2013, 329) – that is, we need to take into account the entire set of relations that belong to the system, instead of simply the technology 'itself'. But when AI systems become infrastructural, we need to find ways to contest the "significant political, ethical and social choices [that] have without doubt been folded into its [the infrastructure's] development" (Susan Leigh Star & Geoffrey Bowker, 2006, 233).

Box 2: Infrastructures revisited

Large language models are often referred to as foundation models because they can be adapted to a variety of different purposes. It is fairly certain that the speed of development and uptake of AI systems will lead to more AI systems becoming defined as foundation models, but currently, they are an illustrative example of this class of AI systems.

For our purposes, foundation models are interesting because:

- their scale is difficult to replicate (not only in terms of data but also computing infrastructure and resource consumption for training)
- their versatility allows for integration into a variety of fields (bringing the strengths and weaknesses of the foundation model as an installed base with it!)
- their workings are not transparent to users (individuals, businesses or researchers), which means that users cannot rely on the procedure to produce outputs accurately relating to their requests, instead need to verify whether the outputs are plausible.

This list is not meant to be conclusive, but indicative of some of the questions we might face - and Ian Brown offers an in-depth review of related issues in his discussion of AI supply chains for the Ada Lovelace Institute. By themselves, these characteristics may be addressed through specific instruments. However, it remains to be seen whether the mix of these characteristics will necessitate other instruments. For instance, the difficulty to replicate a model increases the chance that a large number of users will depend on a certain AI system. Changes to the AI system may therefore lead to a large number of affected users, who are not aware of which changes were made and how they interact with their own prompts, data and linked systems. How the AI system has been integrated by users in different contexts will not be apparent to its developers, which means it can be difficult to anticipate how users are affected by various changes.

Box 3: AI systems of a particular scope and scale

References

Angwin, J. (2023, January 28). *Decoding the Hype About AI. A conversation with Arvind Narayanan*. https://themarkup.org/hello-world/2023/01/28/decoding-the-hype-about-ai

Aradau, C., & Bunz, M. (2022). Dismantling the apparatus of domination?: Left critiques of AI. *Radical Philosophy*, *212*, 10–18. https://www.radicalphilosophy.com/article/dismantling-the-apparatus-of-domination

Böschen, S., Kastenhofer, K., Rust, I., Soentgen, J., & Wehling, P. (2010). Scientific Nonknowledge and Its Political Dynamics: The Cases of Agri-Biotechnology and Mobile Phoning. *Science, Technology, & Human Values*, *35*(6), 783–811. https://doi.org/10.1177/0162243909357911

Brown, I. (2023). *Allocating accountability in AI supply chains: A UK-centred regulatory perspective* [Expert explainer]. Ada Lovelace Institute. https://www.adalovelaceinstitute.org/wp-content/uploads/2023/06/Allocating-accountability-in-AI-supply-chains-June-2023.pdf

Casusi, J. (2023, May 10). *What is a Foundation Model? An Explainer for Non-Experts*. Stanford University: Human-Centered Artificial Intelligence. https://hai.stanford.edu/news/what-foundation-model-explainer-non-experts

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press. https://doi.org/10.12987/9780300252392

Franzen, C. (2023, June 12). The AI feedback loop: Researchers warn of 'model collapse' as AI trains on AI-generated content. *VentureBeat*. https://venturebeat.com/ai/the-ai-feedback-loop-researchers-warn-of-model-collapse-as-ai-trains-on-ai-generated-content/

Hao, K., Swart, H., Hernández, A. P., & Freischlad, N. (n.d.). AI Colonialism. *MIT Technology Review*. https://www.technologyreview.com/supertopic/ai-colonialism-supertopic

Hoeyer, K. (2023). *Data paradoxes: The politics of intensified data sourcing in contemporary healthcare*. The MIT Press.

Larkin, B. (2013). The Politics and Poetics of Infrastructure. *Annual Review of Anthropology*, *42*(1), 327–343. https://doi.org/10.1146/annurev-anthro-092412-155522

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, *33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Raji, I. D., Kumar, I. E., Horowitz, A., & Selbst, A. D. (2022). The Fallacy of AI Functionality. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 959–972. https://doi.org/10.1145/3531146.3533158

Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., & Anderson, R. (2023). *The Curse of Recursion: Training on Generated Data Makes Models Forget* (arXiv:2305.17493). arXiv. http://arxiv.org/abs/2305.17493

Smallman, M. (2022). Multi Scale Ethics—Why We Need to Consider the Ethics of AI in Healthcare at Different Scales. *Science and Engineering Ethics*, *28*(6), 63. https://doi.org/10.1007/s11948-022-00396-z

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111

Suryanarayanan, S., & Kleinman, D. L. (2013). Be(e)coming experts: The controversy over insecticides in the honey bee colony collapse disorder. *Social Studies of Science*, *43*(2), 215–240. https://doi.org/10.1177/0306312712466186

Wang, A., Kapoor, S., Baracas, S., & Narayanan, A. (2022, October 4). *Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy*. FAccT 2023. https://predictive-optimization.cs.princeton.edu/

Wazir, R. (2023, May 11). *AI Comedy of Errors, Or: The Importance of Choosing Your AI Performance Metric With Care*. https://www.leiwand.ai/blog/ai-comedy-of-errors

Whittaker, M. (n.d.). Origin Stories: Plantations, Computers, and Industrial Control. *Logic(s)*, *19*. https://logicmag.io/supa-dupa-skies/origin-stories-plantations-computers-and-industrial-control/