

SOME REMARKS ON THE ETHICAL RELEVANCE OF LEAKY ABSTRACTIONS

Marco Almada

European University Institute (Italy)

Marco.Almada@eui.eu

Algorithms occupy a central place in the current debates on the ethics of computing. As computers are used to replace human labour in various tasks, the greater level of scrutiny over automated decision-making has led to increased attention to the computational procedures that lead to an automated decision. Consider, for example, how “Ditch the algorithm” — and its less polite variants — has become a slogan for protesting against unjust algorithmic decisions, both in the case of British students who had their grades determined by predictions based on historical data and not on their in-class performance (Amoore, 2020) and of American medical professionals who were deprioritised for vaccination despite working in the frontlines of the response to the pandemic (Wamsley, 2020). Nevertheless, while the focus on algorithms can be useful both as a slogan and as an instrument for revealing ethical issues on partially or fully automated decision-making processes, this paper argues that reducing the discussion over computer systems to a debate about algorithms misses relevant questions about the ethical premises and consequences of automation.

The term “algorithm” is often used in one of two meanings. In the narrow sense, used among computer scientists, an algorithm is a formalised description of a procedure that starts from some input values and produces outputs (Louridas, 2020, pp. 21–22). This definition highlights three elements of the computing system that follows such steps: the data that the system receives as an input, the computing steps that are applied to these inputs, and the resulting data that will be returned as an output. By attending to these steps, a growing body of analysis of computer systems has managed to not just identify undesirable outcomes of computation but also to show how such outcomes are produced by biased data collection methods or the application of discriminatory or otherwise inadequate procedures for arriving at a decision from the input data (see, e.g., Washington, 2018). This efficacy, in turn, has contributed to a broader, *metonymical* use of “algorithm” as a description of the entire sociotechnical system that is responsible for the total or partial automation of a decision-making process (Dourish, 2016, p. 3).

Both senses of “algorithm” are relevant for understanding how decision-making processes are automated and the social consequences of such automation. However, adopting an algorithm-centric perspective of automated decision-making systems may mislead analyses of the ethical implications of automation. This paper analyses two such modes of failure — *algorithm reductionism* and *algorithm fetishism* —, describing their impact on how we understand decision-making systems and their common origin: the idea that there is a single privileged perspective that must be used for the analysis of a computer system.

Algorithm reductionism consists of the conflation of the narrow and broad meanings of “algorithm”. Since an algorithm is a mathematical description of a procedure, it will only produce results if *something* follows this procedure — and, in the case of automated decision-making, this something is a computer system. Referring to computer systems as algorithms is a *pars pro toto* substitution that obscures other elements of the system, such as the code that implements an algorithm into an executable program, the tools used in this development process, and the material support in which the program is executed. These elements provide the backdrop for an algorithm’s functioning but might introduce ethical issues of their own. Background elements of a system become relevant whenever they shape the operating premises or the outputs of an algorithm: for example, decision-making systems validated against biased benchmarks have been shown to reproduce these biases (Buolamwini & Gebru, 2018; Caliskan et al., 2017). However, the technological infrastructure behind an algorithm might raise ethical issues not directly related to that algorithm’s main function, as shown by the substantial footprint left by large-scale AI systems (Crawford & Joler, 2018). Consequently, any ethical analysis that looks solely at the algorithm as a decision-making procedure might fail to identify and address the reasons why an AI system generates certain effects.

Algorithm-centric analysis may also contribute to algorithm fetishism, that is, the treatment of algorithms as entities that independently produce social effects (Chun, 2008). Fetishism is not wrong *per se*, as the social meaning ascribed to computer systems often goes beyond the technical properties displayed by those systems (Thomas et al., 2018). Still, tracing broader social, economic, and cultural outcomes to an algorithm can be misleading if one fails to distinguish between the term's broad and narrow senses. This can be seen in the two cases described above, in which the algorithm became a scapegoat for the harm generated by the application of criteria established by human stakeholders (educational and medical authorities, respectively). An algorithm-centric discussion of computer ethics may thus be weaponised by human actors that seek to avoid responsibility for their decisions by placing the blame upon a computational system that cannot deal with the ethical and legal consequences of its actions (Bryson et al., 2017).

At first glance, fetishism and reductionism may seem to pull ethical inquiry toward opposite directions: while reductionism disregards aspects of computing that are not described in the algorithm, fetishism can overestimate that algorithm's reach by blaming (or praising) it for things that are at best mediated — but not caused — by computation. However, these issues can be traced to the same source: the idea that ethical analysis of computer systems can proceed by adopting a single, privileged point of view: that of the algorithm that drives its operation.

Ethical and legal analyses centred on algorithms rely on the computer science method of *abstraction*, which moves to the background information not related to the elements of interest for the question at hand (Colburn & Shute, 2007). In this sense, both meanings of “algorithm” provide different abstractions: the narrow sense of algorithm moves to the background information about the material and social context in which the procedure is followed, while the broad sense hides many of the technical contingencies internal to the sociotechnical system. However, these abstractions are often *leaky* (Spolsky, 2004) in the sense that the technical and human factors that were abstracted away might themselves have consequences for the ethical and legal consequences associated with the system.

In this context, algorithm reductionism and fetishism provide signs that an abstraction is leaking and, consequently, obscuring relevant elements for ethical analysis. This paper argues that leaky abstractions should not be addressed through a change of unit of analysis. Instead, these systems should be understood as assemblages (Nail, 2017) formed by a multiplicity of elements and their relations, which are subject to changes as time passes and social contexts change. Recognising that abstractions leak and that such leakages can have ethical and legal consequences, the paper thus proposes a framework for multi-layered analyses, which look at AI systems from various levels of abstraction at the same time. By combining insights from these various perspectives on a system, multi-layered analyses can contribute to the design of effective ethical and legal interventions in technology, ensuring that they target elements in which change can be effected and reducing the risks of misleading fetishism and reductionism that might follow from the choice of a single analytical vantage point.

KEYWORDS: sociotechnical systems, assemblage theory, algorithms, abstraction, levels of abstraction

REFERENCES

- Amoore, L. (2020, August 19). Why ‘Ditch the algorithm’ is the future of political protest. *The Guardian*. Retrieved from <http://www.theguardian.com/commentisfree/2020/aug/19/ditch-the-algorithm-generation-students-a-levels-politics>
- Bryson, J. J., Diamantis, M. E., & Grant, T. D. (2017). Of, for, and by the people: The legal lacuna of synthetic persons. *Artificial Intelligence and Law*, 25(3), 273-291.
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81, 1-15.
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183-186.
- Chun, W. H. K. (2008). On ‘Sourcery,’ or Code as Fetish. *Configurations*, 16(3), 299-324.
- Colburn, T., & Shute, G. (2007). Abstraction in Computer Science. *Minds and Machines*, 17(2), 169-184.
- Crawford, K., & Joler, V. (2018). *Anatomy of an AI System*. Retrieved from <http://www.anatomyof.ai>

- Dourish, P. (2016). Algorithms and their others: Algorithmic culture in context. *Big Data & Society*, 3(2).
- Louridas, P. (2020). *Algorithms*. Cambridge, MA: The MIT Press.
- Nail, T. (2017). What is an Assemblage? *SubStance*, 46(1), 21-37.
- Spolsky, J. (2004). The Law of Leaky Abstractions. In J. Spolsky (Ed.), *Joel on Software: And on Diverse and Occasionally Related Matters That Will Prove of Interest to Software Developers, Designers, and Managers, and to Those Who, Whether by Good Fortune or Ill Luck, Work with Them in Some Capacity* (pp. 197-202). New York, NY: Apress.
- Thomas, S. L., Nafus, D., & Sherman, J. (2018). Algorithms as fetish: Faith and possibility in algorithmic work. *Big Data & Society*, 5(1).
- Wamsley, L. (2020). Stanford Apologizes After Vaccine Allocation Leaves Out Nearly All Medical Residents. *NPR.Org*. Retrieved from <https://www.npr.org/sections/coronavirus-live-updates/2020/12/18/948176807/stanford-apologizes-after-vaccine-allocation-leaves-out-nearly-all-medical-resid>
- Washington, A. L. (2018). How to argue with an algorithm: Lessons from the Compas-ProPublica debate. *Colorado Technology Law Journal*, 17(1), 131-160.