

Ethics and Redefining Notions of Bias: An Active, Intelligent Information Processing Goal in the Insight Engine

Quran Karriem

PhD Candidate, Computational Media Arts and Cultures, Computational Media, Arts and Cultures; Duke, 114 S Buchanan Boulevard, Bay 9, A289 Smith Warehouse, Duke Box 90766, Durham, NC 27708-0766, USA
quran.karriem@duke.edu

William Seaman

Professor, Computational Media, Arts and Cultures; Co-dir. Emergence Lab, Duke, 114 S Buchanan Boulevard, Bay 9, A289 Smith Warehouse, Duke Box 90766, Durham, NC 27708-0766, USA, bill.seaman@duke.edu

A large area of our research process in constructing the World Generator 2.0 is concerned with selection biases and the limits of intelligent systems to instantiate ethical solutions while operating within the constraints of an imperfect world. Our position is that a system that meaningfully “avoids biases” while using tools that come out of a biased society is a paradox and impossibility, and yet nevertheless one we must continually strive toward. We have to work in the world as it exists, with an ideal in mind for the one we want to build. Humans and human societies are unavoidably biased, and our machines inevitably produce recursions of those biases. Our position is that the best we can do is to instantiate shifts in our biases: to direct them toward explicitly and universally pro-human ends. Some part of that is being overt about inclusion and consent, and also openness to the notion of counterhistory as curators of the system. Much of AI, artificial life, etc. falls prey to a “great (white) man” theory of history. It is increasingly common knowledge that intelligent systems are subject to imperfections in data collection, rely on analysis of human artifacts that contain racist and misogynist assumptions, to name just a few of the issues.

But we maintain that the limits to the notion of an “objective” artificial intelligence run deeper than “bad data”: we can’t fix it just by fixing the data, or by including more data. *How* we collect and create the data is important, as is the data subject’s consent to inclusion within our project (invited researchers?). The computer’s very mode of operation is constrained by the course of cryptography and military science in Western epistemology. That is, data-based and computational thinking of course pre-date the computer itself, and therefore the very idea of computation as we know it, is a product of a specific time and belief structure in the history of cybernetic and pre-cybernetic thought, with all its promise, all its limitations and all its brutalities. The system’s intelligence can really only work with what we give it; we are in a relationship of co-becoming with our machines and cannot meaningfully separate the origins of today’s sociotechnical effects.

For us, the notion of an “autonomous system” is an oxymoron that has roots in outmoded ideas about (human) individuality. Any cursory inquiry into the actual functioning and maintenance of any technical assemblage will show that they all require human investment and bureaucracy to persist. We view the Insight Engine and World Generator as systems that will undoubtedly help us glean new insights about our corpus but will always need to be steered, and observe that it is no coincidence that the most sophisticated intelligent systems today belong to Google, Facebook, Apple, Amazon and Microsoft; corporations that command massive amounts of internal and external, direct and indirect human labor and capital. The engineers that have invented and implemented the contemporary forms of machine learning

and deep learning are largely sponsored by or work for these firms. At present, the outputs of the real instantiations of these algorithms are closely monitored and controlled by human will and toward ends consistent with what we've come to know as data capitalism. That is bias, though we tend to naturalize the operativity of such systems and view it as objective. We view as fallacious the widespread idea that a technical system might be unbiased. We need to explicitly bias them toward universalist pro-human rather than anti-human or post-human ends. In the most widely known accounts, posthuman agency is understood as *distributed* agency. The notion of distributed agency has long been associated with egalitarian or communitarian ideals, but we want to draw attention to the fact that the mere existence of technical systems doesn't inaugurate any such increase in freedom, and in fact has been instrumental to emergent forms of surveillance and protocological control. And, finally, we want to reject ideas of transcendence through technology. Too often, we simply forget that generative techniques, machine learning and intelligent systems in general, are contextualized and trained on corpuses. Systems can't include 'everything' and are therefore always already selections, which at some level of operational abstraction, require human decision. In our research into the World Generator, it is our goal to define a new set of questions that help us unveil new approaches to issues in the ethics of technosociality.