Bittersweet Mysteries of Machine Learning (A Provocation)

Frank Pasquale, professor of law at the University of Maryland, reflects on the roles of machines and machine learning in today's society, and to what extent 'opaque' algorithmic systems should be subject to human oversight.

In the film *2001: A Space Odyssey,* the mission-controlling computer HAL acts mysteriously, and ultimately malevolently. The theme of "technics out of control" animated several other mid-20th-century films. Common themes emerged, variations on the Frankenstein or "sorcerer's apprentice" myths. For example, in *Colossus: The Forbin Project,* a machine assures perpetual peace, but appears capable of controlling much more than nuclear arsenals. It sternly warns one of its creators that "freedom is just an illusion" and "In time you will come to regard me not only with respect and awe, but with love."

The concerns raised by *2001* and *Colossus* have reemerged in global campaigns against Lethal Autonomous Weapon Systems (LAWS). There is broad moral agreement that opaque computational systems should not be allowed to make life-and-death decisions on battlefields. Similar queasiness greets the rise of proprietary "threat scores" and robotic police and prison guards.

However, in commercial contexts, there appears to be a growing enthusiasm for (or at least resignation to) algorithmic decision making inscrutable to humans. Many argue that in the case of the high technology and finance firms focused on in *The Black Box Society,* transparency is a quixotic goal. The tech press sombrely pronounces that the algorithms behind, say, Google's search engine, are so complex that nobody understands them. As one reviewer, David Auerbach, put it, "even those on the *inside* can't control the effects of their algorithms. . . .Who has the time to validate hundreds of millions of classifications?"

Some take the argument further, assuming that the computation involved now amounts to a form of cognition as hard to explain as that of a human decision-maker. Genetic algorithms may, for instance, themselves spawn, each second, dozens of ways of processing information, which are then evaluated on some metric, and Darwinianly given a chance to persist based on their performance. Iterative machine learning processes may be similarly complex and opaque. Just as we can't map all the brain's neurons to connect a person's decision to eat a slice of cake to some set of synapses, we can't map or unravel the sequence of events that leads to a given algorithmic score or sorting.

A Step Back

US Supreme Court Justice Anthony Kennedy has declared that liberty is "the right to define one's own concept of existence, of meaning, of the universe, and of the mystery of human life." Dissenting, Justice Antonin Scalia called those words a "sweet-mystery-of-life passage," and concluded that "if the passage calls into question the government's power to regulate *actions based on* one's self-defined 'concept of existence, etc.,' it is the passage that ate the rule of law." I believe that we should be just as suspicious of the deregulatory impulse behind characterisations of machine learning as "infinitely complex," beyond the scope of human understanding. The artificial intelligence that commercial entities celebrate can just as easily evince artificial imbecility, or worse.

Moreover, there are several practical steps we can take even if we agree with the "mystery beyond human understanding" characterisation of machine learning processes.

For example, we may still want to know what data was fed into the computational process. Presume as complex a credit scoring system as you want. I still want to know the data sets fed into it, and I don't want health data in that set—and I believe the vast majority agree with me on that. An account of the data fed into the system is not too complex for a person to understand, or for their own software to inspect. A relatively simple set of reforms could greatly increase transparency here, even if big data proxies can frustrate accountability.

Policymakers are also free to restrict the scope of computational reasoning too complex to be understood in a conventional narrative or equations intelligible to humans. They may decide: if a bank can't give customers a narrative account of how it made a decision on their loan application, including the data consulted and algorithms used, then the bank can't be eligible for (some of) the array of governmental perquisites or licences so common in the financial field. They may even demand the use of public credit scoring models. (This is also a concern at the core of campaigns regarding lethal autonomous weapons: maybe countries shouldn't develop killing machines powered by algorithms that evolve in unpredictable ways in response to unforeseeable stimuli).

Finally, the results of algorithmic processes can and should be subject to regulation. Consider the new lending startup Affirm. Questioned about why it made a particular loan, its CEO stated "I wouldn't know. Our math model says [the borrower's] OK. Probabilistically, he's good for the money." However, its lawyer added that the firm wanted to avoid a situation where "you seem to have a neutral tool, but its impact is not." In other words, the lender is aiming to avoid disparate impact on certain groups. Restrictions on such disparate impacts need to be entrenched and expanded as complex algorithms and machine learning dominate more business decisions.

For the most *laissez-faire* commentators in the debate on algorithmic accountability, each step in the process of algorithmic ordering is immune from legal contestation or inspection: 1) the data gathered for processing are protected as trade secrets, 2) the processing itself is too complex for any human to understand, and 3) its outputs are "free expression," exempt from ordinary legal restrictions. I have disputed 1) and 3) in other work, and in this provocation I deem 2) the "sweet mystery of machine learning" approach to deflecting regulation.

This mystery is in fact a "bittersweet one," in the sense of a failed line of US Supreme Court decisions that attempted to restrict claimants' due process rights. In those cases, conservative justices stated that the government can create property entitlements with no due process rights attached. In other words, a county might grant someone benefits with the explicit understanding that they could be terminated at any time without explanation: the "sweet" of the benefits could include the "bitter" of sudden, unreasoned denial of them. In *Cleveland Board of Education v. Loudermill* (1985), the Court finally discarded this line of reasoning, forcing some modicum of reasoned explanation and process for termination of property rights.

In the realm of algorithmic accountability, we should be similarly suspicious of the bittersweet mystery of machine learning: firms' insistence that we must take the "bitter" of unregulability with the "sweet" of instant, computed decisionmaking. Even if algorithms at the heart of these processes "transcend all understanding," we can inspect the inputs (data) that go into them, restrict the contexts in which they are used, and demand outputs that avoid disparate impacts.

This article gives the views of the author and does not represent the position of the LSE Media Policy Project blog, nor of the London School of Economics.

This post was published to coincide with a workshop held in January 2016 by the Media Policy Project, 'Algorithmic Power and Accountability in Black Box Platforms'. This was the second of a series of workshops organised throughout 2015 and 2016 by the Media Policy Project as part of a grant from the LSE's Higher Education Innovation Fund (HEIF5). To read a summary of the workshop, please click here.

On 26 January 2016, Professor Pasquale gave a public lecture at LSE on 'The Promise (and Threat) of Algorithmic Accountability'. To see more information about this, and to download a *podcast and video of the event, please click here.*

February 5th, 2016 | Algorithmic Accountability, Featured | 4 Comments