Rage Against the Algorithms

How can we know the biases of a piece of software? By reverse engineering it, of course.

NICHOLAS DIAKOPOULOS

OCT 3, 2013 | TECHNOLOGY

Shutterstock/agsandrew

When was the last time you read an online review about a local business or service on a platform like Yelp? Of course you want to make sure the local plumber you hire is honest, or that even if the date is dud, at least the restaurant isn't lousy. A recent survey found that 76 percent of consumers check online reviews before buying, so a lot can hinge on a good or bad review. Such sites have become so important to local businesses that it's not uncommon for scheming owners to hire shills to boost themselves or put down their rivals.

To protect users from getting duped by fake reviews Yelp employs an algorithmic review reviewer which constantly scans reviews and relegates suspicious ones to a "filtered reviews" page, effectively de-emphasizing them without deleting them entirely. But of course that algorithm is not perfect, and it sometimes de-emphasizes legitimate reviews and leaves actual fakes intact—oops. Some businesses have complained, alleging that the filter can incorrectly remove all of their most positive reviews, leaving them with a lowly one- or two-stars average.

This is just one example of how algorithms are becoming ever more important in society, for everything from search engine personalization, discrimination, defamation, and censorship online, to how teachers are evaluated, how markets work, how political campaigns are run, and even how something like immigration is policed. Algorithms, driven by vast troves of data, are the new power brokers in society, both in the corporate world as well as in government.

They have biases like the rest of us. And they make mistakes. But they're opaque, hiding their secrets behind layers of complexity. How can we deal with the power that algorithms may exert on us? How can we better understand where they might be wronging us?

Transparency is the vogue response to this problem right now. The big "open data" transparency-in-government push that started in 2009 was largely the result of an executive memo from President Obama. And of course corporations are on board too; Google publishes a biannual transparency report showing how often they remove or disclose information to governments. Transparency is an effective tool for inculcating public trust and is even the way journalists are now trained to deal with the hole where mighty Objectivity once stood.

But transparency knows some bounds. For example, though the Freedom of Information Act facilitates the public's right to relevant government data, it has no legal teeth for compelling the government to disclose how that data was algorithmically generated or used in publicly relevant decisions (extensions worth considering).

Moreover, corporations have self-imposed limits on how transparent they want to be, since exposing too many details of their proprietary systems may undermine a competitive advantage (trade secrets), or leave the system open to gaming and manipulation. Furthermore, whereas transparency of data can be achieved simply by publishing a spreadsheet or database, transparency of an algorithm can be much more complex, resulting in additional labor costs both in creation as well as consumption of that information—a cognitive overload that keeps all but the most determined at bay. Methods for usable transparency need to be developed so that the relevant aspects of an algorithm can be presented in an understandable way.

Given the challenges to employing transparency as a check on algorithmic power, a new and complementary alternative is emerging. I call it algorithmic accountability reporting. At its core it's really about reverse engineering—articulating the specifications of a system through a rigorous examination drawing on domain knowledge, observation, and deduction to unearth a model of how that system works.

As interest grows in understanding the broader impacts of algorithms, this kind of accountability reporting is already happening in some newsrooms, as well as in academic circles. At the *Wall Street Journal* a team of reporters probed e-commerce platforms to identify instances of potential price discrimination in dynamic and personalized online pricing. By polling different websites they were able to spot several, such as Staples.com, that were adjusting prices dynamically based on the location of the person visiting the site. At the *Daily Beast*, reporter Michael Keller dove into the iPhone spelling correction feature to help surface patterns of censorship and see which words, like "abortion," the phone wouldn't correct if they were misspelled. In my own investigation for *Slate*, I traced the contours of the editorial criteria embedded in search engine autocomplete algorithms. By collecting hundreds of autocompletions for queries relating to sex and violence I was able to ascertain which terms Google and Bing were blocking or censoring, uncovering mistakes in how these algorithms apply their editorial criteria.

All of these stories share a more or less common method. Algorithms are essentially black boxes, exposing an input and output without betraying any of their inner organs. You can't see what's going on inside directly, but if you vary the inputs in enough different ways and pay close attention to the outputs, you can start piecing together some likeness for how the algorithm transforms each input into an output. The black box starts to divulge some secrets.

Algorithmic accountability is also gaining traction in academia. At Harvard, Latanya Sweeney has looked at how online advertisements can be biased by the racial association of names used as queries. When you search for "black names" as opposed to "white names" ads using the word "arrest" appeared more often for online background check service Instant Checkmate. She thinks the disparity in the use of "arrest" suggests a discriminatory connection between race and crime. Her method, as with all of the other examples above, does point to a weakness though: Is the discrimination caused by Google, by Instant Checkmate, or simply by pre-existing societal biases? We don't know, and correlation does not equal intention. As much as algorithmic accountability can help us diagnose the existence of a problem, we have to go deeper and do more journalistic-style reporting to understand the motivations or intentions behind an algorithm. We still need to answer the question of why.

And this is why it's absolutely essential to have computational journalists not just engaging in the reverse engineering of algorithms, but also reporting and digging deeper into the motives and design intentions behind algorithms. Sure, it can be hard to convince companies running such algorithms to open up in detail about how their algorithms work, but interviews can still uncover details about larger goals and objectives built into an algorithm, better contextualizing a reverse-engineering analysis. Transparency is still important here too, as it adds to the information that can be used to characterize the technical system.

Despite the fact that forward thinkers like Larry Lessig have been writing for some time about how code is a lever on behavior, we're still in the early days of developing methods for holding that code and its influence accountable. "There's no conventional or obvious approach to it. It's a lot of testing or trial and error, and it's hard to teach in any uniform way," noted Jeremy Singer-Vine, a reporter and programmer who worked on the *WSJ* price discrimination story. It will always be a messy business with lots of room for creativity, but given the growing power that algorithms wield in society it's vital to continue to develop, codify, and teach more formalized methods of algorithmic accountability. In the absence of new legal measures, it may just provide a novel way to shed light on such systems, particularly in cases where transparency doesn't or can't offer much clarity.

ABOUT THE AUTHOR

NICHOLAS DIAKOPOULOS is a Tow Fellow at the Columbia University Graduate School of Journalism.