Don't Believe the Algorithm

Blind faith in machines (and machine learning) has left us vulnerable to biased and incoherent AI. The solution? A healthy dose of skepticism and human oversight.

PHOTO: SOPHY HOLLINGTON

By Hannah Fry

Sept. 5, 2018 10:27 am ET

The Notting Hill Carnival is Europe's largest street party. A celebration of black British culture, it attracts up to two million revelers, and thousands of police. At last year's event, the Metropolitan Police Service of London deployed a new type of detective: a facial-recognition algorithm that searched the crowd for more than 500 people wanted for arrest or barred from attending.

Driving around in a van rigged with closed-circuit TVs, the police hoped to catch potentially dangerous criminals and prevent future crimes.

It didn't go well. Of the 96 people flagged by the algorithm, only one was a correct match. Some errors were obvious, such as the young woman identified as a bald male suspect. In those cases, the police dismissed the match and the carnival-goers never knew they had been flagged. But many were stopped and questioned before being released. And the one "correct" match? At the time of the carnival, the person had already been arrested and questioned, and was no longer wanted.

Given the paltry success rate, you might expect the Metropolitan Police Service to be sheepish about its experiment. On the contrary, Cressida Dick, the highest-ranking police officer in Britain, said she was "completely comfortable" with deploying such technology, arguing that the public expects law enforcement to use cutting-edge systems. For Dick, the appeal of the algorithm overshadowed its lack of efficacy.

She's not alone. A similar system tested in Wales was correct only 7% of the time: Of 2,470 soccer fans flagged by the algorithm, only 173 were actual matches. The Welsh police defended the technology in a blog post, saying, "Of course no facial recognition system is 100% accurate under all conditions." Britain's police force is expanding the use of the technology in the coming months, and other police departments are following suit. The NYPD is said to be seeking access to the full database of drivers' licenses to assist with its facial-recognition program.

Law enforcement's eagerness to use an immature technology underscores a worrisome trend you may have noticed elsewhere: Humans have a habit of trusting the output of an algorithm without troubling themselves to think about the consequences. Take the errors we blame on spell check, or the tales of people who follow their GPS over a cliff. We assume that the facial-recognition booths at passport control must be accurate simply because they're installed at our borders.

In my years of working as a mathematician with data and algorithms, I've come to believe that analyzing how an algorithm works is the only way to objectively judge whether it is trustworthy. Algorithms are a lot like magical illusions. At first they appear to be nothing short of wizardry, but as soon as you know how the trick is done, the mystery evaporates. There's often something laughably simple (or reckless) hiding behind the facade.

There's no doubting the profound positive impact that algorithms have had on our lives. The ones we've built to date boast a bewilderingly impressive list of accomplishments. They can help us diagnose breast cancer, catch serial killers and avoid plane crashes. But in our hurry to automate, we seem to have swapped one problem for another. Algorithms—useful and impressive as they are—have already left us with a tangle of complications.

USE 'MAGIC' TO SPOT BOGUS ALGORITHMS

The helpfulness of algorithms varies drastically. So how can you tell the bad from the good? There's a quick trick I use to weed out suspicious examples. I call it the Magic Test. Whenever you see a story about an algorithm, replace buzzwords like "machine learning," "artificial intelligence" and "neural network" with the word "magic." Does everything still make grammatical sense? Is any of the meaning lost? If not, I'd be worried that something smells like bull—. Because I'm afraid—long into the foreseeable future—we're not going to "solve world hunger with magic" or "use magic to write the perfect screenplay" any more than we are with AI.

Our reluctance to question the power of a machine has handed junk algorithms the power to make life-changing decisions, and unleashed a modern snake-oil salesman willing to trade on myths and profit from gullibility. Despite a lack of scientific evidence to support such claims, companies are selling algorithms to police forces and governments that can supposedly "predict" whether someone is a terrorist or a pedophile based on his or her facial characteristics alone. Others insist their algorithms can suggest a change to a single line of a screenplay that will make the movie more profitable at the box office. Matchmaking services insist their algorithm will locate your one true love.

Even the algorithms that (mostly) fulfill their promises have issues. The facial-recognition algorithm at the Manchester Airport failed to notice when a husband and wife accidentally presented each other's passports to the scanners. Recidivism algorithms used in courtrooms overestimated black defendants' likelihood to be repeat offenders, and underestimated the same likelihood for white defendants. Algorithms used by retailers to pinpoint pregnant women and serve them ads can't be turned off, even after a miscarriage or a stillbirth.

In addition to faulty or biased algorithms, there are countless examples of humans using algorithms for unintended purposes. An algorithm designed to identify people who might be involved in a future gun crime, either as a perpetrator or as a victim, was misused by the Chicago Police Department to drum up a list of suspects whenever a homicide had occurred.

The inherent problems of algorithms are magnified when they are paired with humans and our ready acceptance of artificial authority.

But maybe that's precisely the point. Perhaps thinking of algorithms as some kind of authority is where we went wrong.

Even when algorithms aren't involved, there are few examples of perfectly fair, accurate systems. Wherever you look, in whatever sphere you examine, you'll find some kind of bias if you delve deep enough.

Smart Cities: Safer Living or Cyber Attacks?

Urban areas around the world are increasingly equipped with sensors to track-and control-everything from traffic patterns to wastewater. But this increased connectivity and data collection opens us up to devastating hacks and unwanted surveillance.

00:00 / 17:56   SUBSCRIBE

Imagine if we accepted that perfection doesn't exist. Algorithms will make mistakes. Algorithms will be unfair. In time, they will improve. But admitting that algorithms, like humans, have flaws should diminish our blind trust of their authority and lead to fewer mistakes. In my view, the best algorithms take their makers into account at every stage. They recognize our tendency to overtrust machines. They embrace their own uncertainty.

IBM's Watson did this during its "Jeopardy!"-winning run by presenting a confidence score indicating how sure it was of an answer, as well as a list of other guesses it had considered. Perhaps if recidivism algorithms presented something similar, judges might find it easier to question them. If facial-recognition algorithms used by police presented a number of possible matches—instead of homing in on a single face—misidentification might be less of an issue.

The best results occur when humans and algorithms work together. Neural networks that screen breast cancer slides aren't designed to diagnose tumors; they are designed to narrow down a vast array of cells to a handful of suspicious areas for the pathologist to check. The algorithm performs the lion's share of the work, and the human comes in at the end to provide expertise. Machine and human work together in concert, exploiting each other's strengths and embracing each other's flaws.

This is the future I'm hoping for, one where arrogant, dictatorial algorithms are a thing of the past, and we stop seeing machines as objective masters and start treating them as we would any other source of power. We need to question algorithms' decisions, scrutinize their motives, acknowledge our emotions, demand to know who stands to benefit, hold the machines accountable for their mistakes, and refuse to accept underperforming systems. This is the key to a future in which the net effect of algorithms is a positive force in society. The job rests squarely on our shoulders. Because one thing is for sure: In the age of the algorithm, humans have never been more important.

Hannah Fry is an associate professor in the mathematics of cities at University College London. This essay is adapted from her book "Hello World: Being Human in the Age of Algorithms," to be published Sept. 18 in the U.S. by W.W. Norton & Company.