BROOKINGS

Report

A guide to healthy skepticism of artificial intelligence and coronavirus

Alex Engler Thursday, April 2, 2020

Editor's Note:

This report from The Brookings Institution's Artificial Intelligence and Emerging Technology (AIET) Initiative is part of "AI Governance," a series that identifies key governance and norm issues related to AI and proposes policy remedies to address the complex challenges associated with emerging technologies.

The COVID-19 outbreak has spurred considerable news coverage about the ways artificial intelligence (AI) can combat the pandemic's spread. Unfortunately, much of it has failed to be appropriately skeptical about the claims of AI's value. Like many tools, AI has a role to play, but its effect on the outbreak is probably small. While this may change in the future, technologies like data reporting, telemedicine, and conventional diagnostic tools are currently far more impactful than AI.

Still, various news articles have dramatized the role AI is playing in the pandemic by overstating what tasks it can perform, inflating its effectiveness and scale, neglecting the level of human involvement, and being careless in consideration of related risks. In fact, the COVID-19 AI-hype has been diverse enough to cover the greatest hits of exaggerated claims around AI. And so, framed around examples from the COVID-19 outbreak, here are eight considerations for a skeptic's approach to AI claims.

1. Look to the subject-matter experts

No matter what the topic, AI is only helpful when applied judiciously by subject-matter experts—people with long-standing experience with the problem that they are trying to solve. Despite all the talk of algorithms and big data, deciding what to predict and how to frame those predictions is frequently the most challenging aspect of applying AI. Effectively predicting a badly defined problem is worse than doing nothing at all. Likewise, it *always* requires subject matter expertise to know if models will continue to work in the future, be accurate on different populations, and enable meaningful interventions.

In the case of predicting the spread of COVID-19, look to the epidemiologists, who have been using statistical models to examine pandemics for a long time. Simple mathematical models of smallpox mortality date all the way back to 1766, and modern mathematical epidemiology started in the early 1900s. The field has developed extensive knowledge of its particular problems, such as how to consider community factors in the rate of disease transmission, that most computer scientists, statisticians, and machine learning engineers will not have.

> **“There is no value in AI without subject-matter expertise.”**

It is certainly the case that some of the epidemiological models employ AI. However, this should not be confused for AI predicting the spread of COVID-19 on its own. In contrast to AI models that only learn patterns from historical data, epidemiologists are building statistical models that explicitly incorporate a century of scientific discovery. These approaches are very, very different. Journalists that breathlessly cover the “AI that predicted coronavirus” and the quants on Twitter creating their first-ever models of pandemics should take heed: There is no value in AI without subject-matter expertise.

2. AI needs lots of data

The set of algorithms that conquered Go, a strategy board game, and “Jeopardy!” have accomplishing impressive feats, but they are still just (very complex) pattern recognition. To learn how to do anything, AI needs tons of prior data with known outcomes. For instance, this might be the database of historical "Jeopardy!" questions, as well as the correct answers. Alternatively, a comprehensive computational simulation can be used to train the model, as is the case for Go and chess. Without one of these two approaches, AI cannot do much of anything. This explains why AI alone can't predict the spread of new pandemics: There is no database of prior COVID-19 outbreaks (as there is for the flu).

So, in taking a skeptic's approach to AI, it is critical to consider whether a company spent the time and money to build an extensive dataset to effectively learn the task in question. Sadly, not everyone is taking the skeptical path. VentureBeat has regurgitated claims from Baidu that AI can be used with infrared thermal imaging to "see" the fever that is a symptom of COVID-19. Athena Security, which sells video analysis software, has also claimed it adapted its AI system to detect fever from thermal imagery data. Vice, Fast Company, and Forbes rewarded the company's claims, which included a fake software demonstration, with free press.

To even attempt this, companies would need to collect extensive thermal imaging data from people while simultaneously taking their temperature with a conventional thermometer. In addition to attaining a sample diverse in age, gender, size, and other factors, this would also require that many of these people *actually have fevers*—the outcome they are trying to predict. It stretches credibility that, amid a global pandemic, companies are collecting data from significant populations of fevered persons. While there are other potential ways to attain pre-existing datasets, questioning the data sources is always a meaningful way to assess the viability of an AI system.

3. Don't trust AI's accuracy

The company Alibaba claims it can use AI on CT imagery to diagnose COVID-19, and now Bloomberg is reporting that the company is offering this diagnostic software to European countries for free. There is some appeal to the idea. Currently, COVID-19 diagnosis is done through a process called polymerase chain reaction (PCR), which requires specialized equipment. Including shipping time, it can easily take several days, whereas Alibaba says its model is much faster and is 96% accurate.

However, it is not clear that this accuracy number is trustworthy. A poorly kept secret of AI practitioners is that 96% accuracy is suspiciously high for *any* machine learning problem. If not carefully managed, an AI algorithm will go to extraordinary lengths to find patterns in data that are associated with the outcome it is trying to predict. However, these patterns may be totally nonsensical and only appear to work during development. In fact, an inflated accuracy number can actually be an important sign that an AI model is not going to be effective out in the world. That Alibaba claims its model works that well without caveat or self-criticism is suspicious on its face.

"[A]n inflated accuracy number can actually be an important sign that an AI model is not going to be effective out in the world."

In addition, accuracy alone does not indicate enough to evaluate the quality of predictions. Imagine if 90% of the people in the training data were healthy, and the remaining 10% had COVID-19. If the model was correctly predicting all of the healthy people, a 96% accuracy could still be true—but the model would still be missing 40% of the infected people. This is why it's important to also know the model's "sensitivity," which is the percent of correct predictions for individuals *who have COVID-19* (rather than for everyone). This is especially important when one type of mistaken prediction is worse than the other, which is the case now. It is far worse to mistakenly suggest that a person with COVID-19 is not sick (which might allow them to continue infecting others) than it is to suggest a healthy person has COVID-19.

Broadly, this is a task that seems like it could be done by AI, and it might be. Emerging research suggests that there is promise in this approach, but the debate is unsettled. For now, the American College of Radiology says that "the findings on chest imaging in

COVID-19 are not specific, and overlap with other infections," and that it should not be used as a "first-line test to diagnose COVID-19." Until stronger evidence is presented and AI models are externally validated, medical providers should not consider changing their diagnostic workflows—especially not during a pandemic.

4. Real-world deployment degrades AI performance

The circumstances in which an AI system is deployed can also have huge implications for how valuable it really is. When AI models leave development and start making real-world predictions, they nearly always degrade in performance. In evaluating CT scans, a model that can differentiate between healthy people and those with COVID-19 might start to fail when it encounters patients who are sick with the regular flu (and it is still flu season in the United States, after all). A drop of 10% accuracy or more during deployment would not be unusual.

In a recent paper about the diagnosis of malignant moles with AI, researchers noticed that their models had learned that rulers were frequently present in images of moles known to be malignant. So, of course, the model learned that images without rulers were more likely to be benign. This is a learning pattern that leads to the appearance of high accuracy during model development, but it causes a steep drop in performance during the actual application in a health-care setting. This is why independent validation is absolutely essential before using new and high-impact AI systems.

"When AI models leave development and start making real-world predictions, they nearly always degrade in performance."

This should engender even more skepticism of claims that AI can be used to measure body temperature. Even if a company did invest in creating this dataset, as previously discussed, reality is far more complicated than a lab. While measuring core temperature from thermal body measurements is imperfect even in lab conditions, environmental factors make the problem much harder. The approach requires an infrared camera to get a clear and precise view of the inner face, and it is affected by humidity and the ambient temperature of the target. While it is becoming more effective, the Centers for Disease Control and Prevention still maintain that thermal imaging cannot be used on its own—a second confirmatory test with an accurate thermometer is required.

5. Most predictions must enable an intervention to really matter

In high-stakes applications of AI, it typically requires a prediction that isn't just accurate, but also one that meaningfully enables an intervention by a human. This means sufficient trust in the AI system is necessary to take action, which could mean prioritizing health-care based on the CT scans or allocating emergency funding to areas where modeling shows COVID-19 spread.

With thermal imaging for fever-detection, an intervention might imply using these systems to block entry into airports, supermarkets, pharmacies, and public spaces. But evidence shows that as many as 90% of people flagged by thermal imaging can be false positives. In an environment where febrile people know that they are supposed to stay home, this ratio could be much higher. So, while preventing people with fever (and potentially COVID-19) from enabling community transmission is a meaningful goal, there must be a willingness to establish checkpoints and a confirmatory test, or risk constraining significant chunks of the population.

This should be a constant consideration for implementing AI systems, especially those used in governance. For instance, the AI fraud-detection systems used by the IRS and the Centers for Medicare and Medicaid Services do not determine wrongdoing on their own; rather, they prioritize returns and claims for auditing by investigators. Similarly, the celebrated AI model that identifies Chicago homes with lead paint does not itself make the final call, but instead flags the residence for lead paint inspectors.

6. AI is far better at minute details than big, rare events

Wired ran a piece in January titled "An AI Epidemiologist Sent the First Warnings of the Wuhan Virus" about a warning issued on Dec. 31 by infectious disease surveillance company, BlueDot. One blog post even said the company predicted the outbreak "before it happened." However, this isn't really true. There is reporting that suggests Chinese officials knew about the coronavirus from lab testing as early as Dec. 26. Further, doctors in Wuhan were spreading concerns online (despite Chinese government censorship) and the Program for Monitoring Emerging Diseases, run by human volunteers, put out a notification on Dec. 30.

That said, the approach taken by BlueDot and similar endeavors like HealthMap at Boston Children's Hospital aren't unreasonable. Both teams are a mix of data scientists and epidemiologists, and they look across health-care analyses and news articles around the world and in many languages in order to find potential new infectious disease outbreaks. This is a plausible use case for machine learning and natural language processing and is a useful tool to assist human observers. So, the hype, in this case, doesn't come from skepticism about the feasibility of the application, but rather the specific type of value it brings.

"AI is unlikely to build the contextual understanding to distinguish between a new but manageable outbreak and an emerging pandemic of global proportions."

Even as these systems improve, AI is unlikely to build the contextual understanding to distinguish between a new but manageable outbreak and an emerging pandemic of global proportions. AI can hardly be blamed. Predicting rare events is just very hard, and AI's reliance on historical data does it no favors here. However, AI does offer quite a bit of value at the opposite end of the spectrum—providing minute detail.

For example, just last week, California Gov. Gavin Newsom explicitly praised BlueDot's work to model the spread of the coronavirus to specific zip codes, incorporating flight-pattern data. This enables relatively precise provisioning of funding, supplies, and medical staff based on the level of exposure in each zip code. This reveals one of the great strengths of AI: its ability to quickly make individualized predictions when it would be much harder to do so individually. Of course, individualized predictions require individualized data, which can lead to unintended consequences.

7. There will be unintended consequences

AI implementations tend to have troubling second-order consequences outside of their exact purview. For instance, consolidation of market power, insecure data accumulation, and surveillance concerns are very common byproducts of AI use. In the case of AI for fighting COVID-19, the surveillance issues are pervasive. In South Korea, the neighbors of confirmed COVID-19 patients were given details of that person's travel and commute history. Taiwan, which in many ways had a proactive response to the coronavirus, used cell phone data to monitor individuals who had been assigned to stay in their homes. Israel and Italy are moving in the same direction. Of exceptional concern is the deployed social control technology in China, which nebulously uses AI to individually approve or deny access to public space.

Government action that curtails civil liberties during an emergency (and likely afterwards) is only part of the problem. The incentives that markets create can also lead to long-term undermining of privacy. At this moment, Clearview AI and Palantir are among the companies pitching mass-scale surveillance tools to the federal government. This is the same Clearview AI that scraped the web to make an enormous (and unethical) database of faces—and it was doing so as a reaction to an existing demand in police departments for identifying suspects with AI-driven facial recognition. If governments and companies continue to signal that they would use invasive systems, ambitious and unscrupulous start-ups will find inventive new ways to collect more data than ever before to meet that demand.

8. Don’t forget: AI will be biased

In new approaches to using AI in high-stakes circumstances, bias should be a serious concern. Bias in AI models results in skewed estimates across different subgroups, such as women, racial minorities, or people with disabilities. In turn, this frequently leads to discriminatory outcomes, as AI models are often seen as objective and neutral.

While investigative reporting and scientific research has raised awareness about many instances of AI bias, it is important to realize that AI bias is more systemic than anecdotal. An informed AI skeptic should hold the default assumption that AI models are biased, unless proven otherwise.

> **“An informed AI skeptic should hold the default assumption that AI models are biased, unless proven otherwise.”**

For example, a preprint paper suggests it is possible to use biomarkers to predict mortality risk of Wuhan COVID-19 patients. This might then be used to prioritize care for those most at risk—a noble goal. However, there are myriad sources of potential bias in this type of prediction. Biological associations between race, gender, age, and these biomarkers could lead to biased estimates that don’t represent mortality risk. Unmeasured behavioral characteristics can lead to biases, too. It is reasonable to suspect that smoking history, more common among Chinese men and a risk factor for death by COVID-19, could bias the model into broadly overestimating male risk of death.

Especially for models involving humans, there are so many potential sources of bias that they cannot be dismissed without investigation. If an AI model has no documented and evaluated biases, it should increase a skeptic’s certainty that they remain hidden, unresolved, and pernicious.

The future of AI systems is more promising

While this article takes a deliberately skeptical perspective, the future impact of AI on many of these applications is bright. For instance, while diagnosis of COVID-19 with CT scans is of questionable value right now, the impact that AI is having on medical imaging is substantial. Emerging applications can evaluate the malignancy of tissue abnormalities, study skeletal structures, and reduce the need for invasive biopsies.

Other applications show great promise, though it is too soon to tell if they will meaningfully impact this pandemic. For instance, AI-designed drugs are just now starting human trials. The use of AI to summarize thousands of research papers may also quicken medical discoveries relevant to COVID-19.

AI is a widely applicable technology, but its advantages need to be hedged in a realistic understanding of its limitations. To that end, the goal of this paper is not to broadly disparage the contributions that AI can make, but instead to encourage a critical and discerning eye for the specific circumstances in which AI can be meaningful.

The Brookings Institution is a nonprofit organization devoted to independent research and policy solutions. Its mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations for policymakers and the public. The conclusions and recommendations of any Brookings publication are solely those of its author(s), and do not reflect the views of the Institution, its management, or its other scholars.

Microsoft provides support to The Brookings Institution's Artificial Intelligence and Emerging Technology (AIET) Initiative. The findings, interpretations, and conclusions in this report are not influenced by any donation. Brookings recognizes that the value it provides is in its absolute commitment to quality, independence, and impact. Activities supported by its donors reflect this commitment.

Report Produced by **Center for Technology Innovation**