The Inescapability of Uncertainty

Jennifer Wortman Vaughan
Microsoft Research
New York, NY 10011
jenn@microsoft.com

Hanna Wallach
Microsoft Research
New York, NY 10011
hanna@dirichlet.net

Abstract
As big data becomes a larger presence in our lives, we increasingly rely on artificial intelligence to help us make decisions. And why not? There is a certain appeal to the idea that clever algorithms and troves of data will let us tame the uncertainty in our lives. In this position paper, we examine three properties of AI systems: 1) AI is built on assumptions about data and desired system behavior; 2) algorithms can have unfair consequences; and 3) algorithmic predictions can be hard to interpret. We argue that no matter how clever our algorithms are or how much data we collect, it will never be possible to fully escape uncertainty. Therefore, rather than treating AI systems as a way to automatically make our decisions for us, freeing us from dealing with the uncomfortable reality of uncertainty, it is smarter to treat AI systems as a way to leverage the complementary strengths of people and machines.

Author Keywords
Uncertainty; artificial intelligence; machine learning.

ACM Classification Keywords
I.2.0 [Artificial Intelligence]: General; K.4.0 [Computers and Society]: General
Introduction
Uncertainty makes people uncomfortable. We expect our politicians and leaders to project an air of confidence about their decisions,1 and we view admissions of uncertainty as a sign of weakness. It is comforting to imagine that there is a single right choice we could make if only we had access to the right information.

As big data becomes a larger presence in our lives, we increasingly rely on artificial intelligence—a very broad term that covers machine learning and other algorithms that generate predictions based on observations and experience—to help us make intelligent decisions. Every time you buy a suggested product on Amazon or watch a movie that Netflix recommends, you are relying on AI [1].

If you believe the hype, AI will soon be able to handle the hard problems in life for us, freeing up our time for friends, family, and the things that matter most. And why not? After all, AI has already made it possible for computers to solve tasks that traditionally required human input, such as detecting fraudulent credit card transactions, digitizing our handwritten scribbles, or routing customer support calls to the right department. There is a certain appeal to the idea that clever algorithms and troves of data will let us tame the uncertainty in our lives.

But this way of thinking is dangerous. As AI researchers, we are confident that no matter how clever our algorithms are or how much data we collect, it will never be possible to fully escape uncertainty. Here’s why.

AI is Built on Assumptions
Designers of AI systems deal with uncertainty in part by relying on assumptions about their data or the way their systems should work. A particular prediction algorithm might perform well when conditions in the future are more or less the same as conditions in the past or when a system has only well-intentioned users; the same algorithm might fail when conditions change over time, or when a rogue user deliberately tries to subvert the system.

In practice, it is rare for assumptions to perfectly reflect reality. Sometimes algorithms provide useful predictions anyway. But, as we learned from big banks during the financial collapse, mathematical models with flawed assumptions—say, that the past behavior of an asset is a good indicator of how it will perform in the future, or that changes in the values of two assets are unrelated—can be dangerous when we place too much trust in their predictions without carefully considering what could go wrong.

Human Behavior is Complex
Some phenomena are particularly hard to capture using assumptions, especially those that require us to understand people’s collective behavior. Collective behavior is complex. It involves multiple different factors that interact in ways that change all the time. For example, think of the difficulty of predicting fluctuations in the stock market. Or consider the widespread shock that followed the 2016 United States election, despite the money and brainpower spent on polls and analytics.

When making predictions about an election, prospective voters—the very people whose collective actions we would like to predict—are influenced by the predictions themselves. The outcome of the election could change if enough voters read that a particular candidate is favored to win in
their state, causing them to believe their votes don’t matter. This phenomenon perhaps explains the choices some British voters made regarding Brexit; anecdotes suggest some voters were so confident Brexit would lose that they thought they could vote against it as a consequence-free act of protest.

These kinds of unpredictable feedback loops, with their inherent uncertainty, can lead AI to make very bad calls.

**Algorithms Can Have Unfair Consequences**

Sometimes assumptions have serious consequences. Consider the recent ProPublica investigation into the proprietary system widely used within the legal system to predict how likely a defendant is to reoffend. Analyses showed that, on average, black defendants who did not go on to reoffend were twice as likely as corresponding white defendants to be labeled as medium or high risk. The designers of the system argued that they had taken fairness into account; their goal was to ensure that black defendants at any given risk level had the same probability of reoffending as white defendants at the same level. However, satisfying this notion of fairness means that the system will necessarily suffer from the type of apparent bias discovered by ProPublica. It is mathematically implied. In other words, the designers' apparently reasonable assumption of what it means to be fair led to behavior that is arguably unfair under a different measure.

If we want to understand when and why AI algorithms fail, we need to examine the assumptions that lie behind the algorithms. But here’s the challenge. AI technology is becoming much cheaper and more widely available, making it both appealing and easy for people to make decisions based on AI without understanding the assumptions behind it, or even being aware that these assumptions exist. This may not be too much of a problem when we’re talking about annoying online store recommending products you don’t want. But it is a very big problem when AI algorithms are used to make decisions with major consequences for people’s lives.

**Algorithmic Predictions Can Be Hard to Interpret**

Even apart from assumptions, there is the question of how to interpret algorithmic predictions. What did we learn by observing that Clinton’s predicted chance of victory moved from 78% to 80%? Did such a shift reflect a meaningful change in Clinton’s odds of winning, or was it only a temporary blip? What does an 80% chance of victory even mean? Loosely speaking, this prediction says that if our assumptions are correct and we held an election like this one five times, we’d expect Clinton to win four of them. But this particular election only happened once. Trump’s win does not necessarily mean that this prediction was wrong.

People want to believe that algorithmic predictions capture some notion of truth, especially when these predictions are stated precisely and backed up by rigorous math. But it is important to realize that precision and certainty are not the same thing. You can buy a scale that measures weight to the nearest ounce, but it will not tell you whether an observed day-to-day loss of a pound is meaningful or simply a random fluctuation. The seeming precision of algorithmic predictions can make them appear more authoritative than they in fact are.

**AI Can Still Be Useful**

Despite all of this, when used with care, AI can and should be a powerful tool to augment and inform human decision-making. Computers excel at discovering patterns in large amounts of data, while people have the life experience, empathy, and creativity required to ensure that these dis-

---

2http://bit.ly/1XMKh5R
coveries are put to use in ways that are effective and fair. Rather than treating AI systems as a way to automatically make our decisions for us, freeing us from dealing with the uncomfortable reality of uncertainty, it is smarter to treat AI systems as a way to leverage these complementary strengths of people and machines.

A good first step is to bake more transparency into AI—especially transparency surrounding uncertainty. Rather than conceal the assumptions and uncertainty in their predictions, AI systems should enable users to understand the roots of this uncertainty and provide them with ways to reason about it more effectively. Longer term, this is an education issue. We need to acknowledge that uncertainty is here to stay and equip future generations to embrace it.

In the mean time, though, remember that regardless of how precisely stated, there are hidden assumptions behind every prediction.

Acknowledgments
An earlier version of this essay appeared on Medium in October, 2016 as part of Data & Society: Points. We thank Henry Farrell for providing helpful advice and suggesting several detailed edits that made this essay stronger, Jeff Vaughan for countless discussions and debates about these topics, and Solon Barocas and Dave Pennock for valuable feedback. Photo credit: Abigail Jacobs.

REFERENCES