An Interview with Lauren Maffeo: Understanding the risks of machine learning bias

INNOVATION LEADERS

by Bushra Anjum

Editor's Introduction

Lauren Maffeo is a research analyst who joined the global technology sector in 2012. She started her career as a freelance journalist covering tech news for The Next Web and The Guardian. She has also worked with CEOs of pre-seed to profitable SaaS startups on media strategy. Lauren joined [GetApp](#), a Gartner company, as a content editor in 2016. She covers the impact of emerging tech like AI on small and midsize business owners.

Lauren has been cited by sources including [Forbes](#), [Fox Business](#), [DevOps Digest](#), [The Atlantic](#), and [Inc.com](#). In 2017, Lauren was named to [The Drum's 50 Under 30](#) list of women worth watching in digital. She holds an M.Sc. from The London School of Economics and a certificate in Artificial Intelligence: Implications for Business Strategy from MIT's Sloan School of Management.

An Interview with Lauren Maffeo: Understanding the risks of machine learning bias

INNOVATION LEADERS

by Bushra Anjum

Bushra Anjum: What is your big concern about the future of computing to which you are dedicating yourself?

Lauren Maffeo: My main concern is the bias within datasets that are used to train machine learning (ML) algorithms. These algorithms underpin products that have far-reaching consequences.

One well-known example of machine bias is [a product called COMPAS](#), which is an ML algorithm that predicts how likely defendants are to recommit crimes. Research from ProPublica found that the algorithm has been two times more likely to correlate black skin with a higher likelihood to re-offend incorrectly. It has also been two times more likely to correlate white skin with a lower likelihood to re-offend incorrectly.

Another example of machine bias is voice-activated interfaces that only recognize a small number of voices. As a Ph.D. student, Dr. Carol Reiley used Microsoft's speech recognition API to build a voice-activated interface. But because the API was created mostly by men of a certain age, Dr. Reiley's interface couldn't recognize her own voice!

It is important to note that most machine bias isn't intentional. Instead, it appears as an indirect bias—a byproduct of non-sensitive attributes that correlate with sensitive attributes (race, religion, etc.). For example, even if a sensitive attribute like race is not directly accounted for within a dataset, other factors (like higher incarceration rates for people with black skin or a higher police presence in certain zip codes) might cause an incorrect correlation (like an algorithm incorrectly predicting that someone is more likely to re-offend based on their skin color). And since many of these algorithms are black-boxed, machine learning engineers can't always explain how they make decisions [[Dave Gershgorn, Quartz](#)].

Machine bias often occurs because the datasets that are used to train ML algorithms are not diverse enough. For example, voice-activated user interfaces fail when the datasets aren't exposed to enough voice variations, such as accents and inflections.

Without data scientists who consistently monitor the training data that ML algorithms receive in both deployment and production, the risk of bias creeps in and can go unchecked. Any business that builds and deploys ML algorithms should be responsible for assessing the health of their datasets throughout product lifecycles.

BA: How did you first become interested in ML bias? What inspired further research?

LM: Machine bias is a hot topic in the tech world right now. This isn't a new problem though, as [machine bias has existed](#) for as long as we've tried using tech to make "unbiased" decisions.

The difference today is that we have unprecedented amounts of data, and it's easier than ever to build raw predictive models and use these data sets as training sets. The more data an algorithm has, the more examples it has to "train" itself and learn what's right or wrong. But if the quality of data in a training set is poor and/or unmonitored, the algorithm will make incorrect associations. This is also where indirect bias comes in.

I first became interested in AI as an analyst at Gartner. Within recent years, we've seen widespread fear that AI technology will surpass humans' physical and cognitive capabilities. I found this fascinating and wanted to understand this technology to know why it has advanced in such a short timeframe.

Since Gartner is a treasure trove of research on AI, I read as much as I could about this subject so that I could start researching it myself. I initially pursued this research as a passion project before it became a stronger focal point of my work. To increase my knowledge in this area, I also earned a certificate in Artificial Intelligence: Implications for Business Strategy from MIT's Sloan School of Management in 2018.

I focus my current research on the ways that small and midsize businesses (SMBs) use new tech like AI to grow their businesses. This includes assessing various types of AI technologies (like ML versus. robotics) and concluding which technologies are the best fit to solve specific business problems. For example, some construction firms [use robotics for bricklaying](#)—a rules-based, repetitive task that's easily automated.

BA: What are some of the initiatives you are leading to spread awareness about machine bias?

PW: I've been passionate about spreading awareness of machine bias since I started researching AI back in 2016. Throughout 2018, I've presented on this subject at international conferences.

These include giving lightning talks at the 2018 Open Source Summit in Vancouver, Canada, and [All Things Open](#) in Raleigh, North Carolina making a case for why product teams should add bias testing to their development lifecycles. I also gave a longer version of my lightning talk on this subject at Google Dev Fest DC in September 2018 and [wrote an op-ed for OpenSource.com](#) making a case for why the classifiers in open source AI algorithms should be transparent.

In late 2017, I realized how ubiquitous AI already is within consumer and business-to-business (B2B) products that we use daily: It's just that we don't recognize it as "AI" per se. If you use products like Slack, Alexa, or an iPhone, you already use AI. But since there is so much misinformation about AI, discomfort and confusion can lead to not realizing how AI helps us and being too scared of this technology to avoid investing in it.

So, I recently [published research](#) for Gartner on how application leaders who report to CIOs and manage development teams can use AI technologies like deep learning and semantic parsing to solve a range of business problems. This research also explains how application leaders should assess the health of datasets that train the algorithms they'll use, for example working with software vendors who incorporate ML into their products to understand how these vendors manage their datasets.

It's a well-established fact that machine bias is unavoidable. It depends on priorities set by the algorithms' designers and which methods of fairness designers account for at the start. The problem is that as fairness increases by one measure, it may decrease by another.

That said, machine bias is not an unmanageable problem. There are steps that product and development teams can take to reduce machine bias throughout the lifecycle such as using a pick-step to select a representative set of predictions or conclusions to be explained and then approximating the model close to those predictions.

If readers have their own thoughts on how to achieve this, feel free to connect with me on [LinkedIn](#) and [Twitter](#). Many thanks to Bushra and the ACM Ubiquity team for letting me share my thoughts!

Biography

Bushra Anjum is a software technical lead at Amazon in San Luis Obispo, CA. She has expertise in Agile Software Development for large scale distributed services with special emphasis on scalability and fault tolerance. Originally a Fulbright scholar from Pakistan, Dr. Anjum has international teaching and mentoring experience and has served in academia for over five years before joining the industry. In 2016, she has been selected as an inaugural member of the [ACM Future of Computing Academy](#), a new initiative created by ACM to support and foster the next generation of computing professionals. Dr. Anjum is a keen enthusiast of promoting diversity in the STEM fields and is a mentor and a regular speaker for such. She received her Ph.D. in computer science at the North Carolina State University (NCSU) in 2012 for her doctoral thesis on Bandwidth Allocation under End-to-End Percentile Delay Bounds. She can be found on Twitter [@DrBushraAnjum](#).

DOI: 10.1145/3306159