

#EqualAI

*Bulletin on Bias & Artificial Intelligence:
Issue 7, October 2020*

Spotlight of the Week: Aubrey Blanche



“Right now, we can take responsibility to consider the data we collect and how it can be changed to mitigate its inherent biases, build models that monitor their own outputs for inequitable and discriminatory outcomes, and change course. It's going to take all of us -- and that includes valuing and integrating the perspectives of people from impacted communities in the development process -- but we can do it if we choose to.”

- Aubrey Blanche, 'The Mathpath' &
Director of Equitable Design & Impact at Culture Amp

Hot Topics in Business and Technology 🔥

- **A GPT-3 bot posted comments on Reddit for a week and no one noticed:** Powered by OpenAI's powerful GPT-3 language model, a chatbot called '/u/thegentlemetre' interacted on Reddit, a forum with up to 30 million users, raising concerns around public operation of the AI technology (10/8).
- **Instagram Will Hide Some User Comments As Part Of New 'Anti-Bullying' Features:** Instagram announced the launch of two new anti-bullying features to hide potentially offensive comments and expand warning messages (10/6).
- **Parkland parents create AI video of slain son to spur voters:** Using data from Twitter to recreate their son's mannerisms and speech, parents urge young people to vote with an AI-powered video simulation of their late son, who was killed by gun violence (10/2).
- **New York University (NYU) researchers found that popular language models reinforce harmful stereotypes:** NYU researchers found that

language models including Google's BERT and ALBERT and Facebook's RoBERTa reinforce and amplify harmful stereotypes in the datasets (10/8).

- **Many Top AI Researchers Get Financial Backing From Big Tech:** A study showed that 58% of faculty at Stanford, MIT, UC Berkeley, and University of Toronto receive financial support from large tech firms like Google, raising concerns about influences in academic research (10/8).
- **Waymo opens driverless ride-hailing service to the public:** Alphabet's self-driving car service is expanding its fully driverless ride services to the public in the Phoenix area (10/8).
- **Tesla Autopilot scores low for driver engagement in European safety rating:** Autopilot, Tesla's AI driver assistance system, ranked 6/10 on a European safety assessment, scoring especially low on the tool's ability to aid driver focus, with Mercedes GLE topping the chart (10/1).
- **AI and Media: A Special Report in Variety:** Algorithms to detect audience preferences are being used to estimate potential success and failure in the entertainment and media industry. Algorithms analyze film scripts and generate surprisingly accurate models for how a film will perform in various territories or which song will be a hit. (10/1)
- **AI tool pens a \$5M seed round for its AI style guide that flags bias and tone:** Writer, an AI-powered language tool, aims to help find and redress language inconsistent with a company's style guide or values (10/2).
- **A.I. is playing a bigger role in music streaming than imagined:** In music services, AI is our DJ, powering music shuffling and adjusting sound volumes as listeners transition between songs. However, Fortune notes that music streaming services still remain reliant on human curators because of "[t]he truth is music is entirely subjective". (10/1)
- **A.I. gets down in the dirt as precision agriculture takes off:** Farmers are using high-resolution cameras and sensors linked to A.I. software to monitor grain collection for greater crop yields and minimized waste levels. (10/5)
- **Facebook bans QAnon across its platforms:** Facebook banned all QAnon content from its platforms, restating announced policy to remove accounts related to QAnon that insinuate violence. (10/6)
- **CSET Data Brief: Patent Landscape for Computer Vision: US and China:** CSET reviews patents filings and their source (private or public) in the U.S. and Chinese, highlighting that in China, the top 100 AI patenting organizations overall contained 65 universities. In the US, only 1 of the top 100 AI patenting organizations was a university. 10/6

AI & Public Health

- **AI tool aids in reducing hospital deaths:** Sepsis Watch, an AI monitoring system used to fight hospital mortality, has dramatically reduced sepsis-induced patient deaths and is now part of a federally-registered clinical trial expected to be published in 2021 (10/2).

US Federal and State Government News

- **Labor Department probes Microsoft's diversity policy:** DOL is investigating whether Microsoft's plan to increase Black representation in leadership positions constitutes illegal racial discrimination (10/6).

- **House Panel Urges Tech Giant Breakup in Plan Republicans Shunned:** The House antitrust subcommittee called for sweeping reforms to antitrust laws against Amazon.com Inc., Alphabet Inc., Facebook Inc. and Apple Inc., claiming the tech giants practice anti-competitive behavior (10/6).
- **JPMorgan Probe Revived by Regulators' Data Mining:** The U.S. Commodity Futures Trading Commission levied a \$920 million fine against JPMorgan Chase in a case bolstered by the use of data analytics to uncover market manipulations. "We could not have brought the JPMorgan case without the data analytics program we have now," (10/5).
- **Detroit Renews Facial Recognition Contract:** Detroit City Council voted to renew police contract with facial recognition providers (9/29).

AI Around the World 🌍

- **UK:** Robotics firm AutoStore sued British online supermarket company Ocado for allegedly infringing its patent for robotically operated warehouses (10/1).
- **Belgium:** In an effort to boost AI talent, the Belgian government is providing a free online course run by the University of Helsinki to teach AI fundamentals to students without prior related experience (9/30).
- **Norway:** Norway's fish farms are using AI models to cut costs and improve efficiency in salmon farming (9/30).

EqualAI in the News 📰

- **Michèle Flournoy joins advocates for DOD to work more collaboratively on AI testing:** EqualAI Senior Advisor Michèle Flournoy and other national security experts urge DoD to establish a more collaborative approach for testing and evaluation of AI, emphasizing the need for robust methods of testing, evaluation, validation and verification of AI (10/7).
- **Racist Algorithms: How Code Is Written Can Reinforce Systemic Racism:** EqualAI intern Calder Katyal published a piece in Teen Vogue explaining how algorithmic bias is impacting students and youth specifically, in addition to general societal harms (10/6):
 - *"What I did not realize until this summer was that my generation is already experiencing bias from our most trusted ally: the computer."*

Op-Eds: Top Picks of the Week 🤖

- **How GPT-3 Is Shaping Our AI Future:** On this Harvard Business Review 'Exponential Review' Podcast, OpenAI CEO Sam Altman reflects on the attention generated by GPT-3 and implications of true artificial general intelligence including its use both to reduce and exacerbate inequality. (10/7).
- **How much more abuse do female politicians face? A lot:** MIT Technology Review reports on study indicating exacerbated levels of online harassment towards women. Researchers found that women and people of color were "far more likely than men to be abused on Twitter." Between 5% and 10% of mentions of most male politicians were considered abusive, while mentions of female politicians on Twitter contained abuse between 15% and 39% of the time. (10/6):
 - *"The findings may not really be new, but they reinforce what women and people of color have been saying about Facebook, Twitter, and*

other major social-media companies for years: that while the sites may increasingly have policies banning abusive behavior, the enforcement of those policies often leaves its most common targets open to sustained, coordinated harassment.”

- **3 mantras for women in data:** According to MIT Sloan, women in the data industry should make diversity part of their data practice (10/6):
 - *“Being aware of how you might be training models based on assumptions that go into how you create applications is really important. To make it as equitable as possible, you really want diverse views baked into the use of that data, the collection of that data, from the very earliest stages” - Ashley Van Zeeland, VP at Illumina, to MIT*

Upcoming Events 🎉

- ‘How Will Quantum Computing Shape the Future of AI?’ hosted by The Center for Data Innovation: [October 14](#).
- ‘Civic Tech: The Path to Public Oversight of Surveillance Technology in Minneapolis,’ hosted by HAI: [October 14](#).
- Politico AI Summit: An AI-Powered World: [October 15](#).
- ‘100 Brilliant Women in AI Summit 2020,’ hosted by Women in AI Ethics: [October 19](#).
- ‘AI and Education,’ Oxford University Ethics in AI Seminar: [October 29](#).