The Bulletin

On Bias & Artificial Intelligence

Spotlight of the Week: Sean Perryman

*"Reducing bias in AI is critical in building the public's trust in technology that will shape our future. AI can either replicate past bias or help us eliminate them, and it's critical that we have a risk-based approach to policymaking that allows for continued responsible innovations in this field. **The way we shape this technology will determine whether we live up to the promise of a more equitable future or whether we simply replicate past discrimination in our new technologies**."*

- *Sean Perryman, Director of Social Impact Policy and Counsel at the Internet Association*

Hot Topics in Business and Technology 🔥

- **McKinsey finds that companies are behind in recognizing AI risk factors:** Recent survey results released by McKinsey revealed that less than half the firms identified AI risks due to regulatory compliance, privacy, fairness as "relevant" to their AI projects (12/3).
  - "It's difficult to understand why universal risks aren't recognized by a much higher proportion of respondents".
- **Columbia researchers find white men are the worst at reducing AI bias:** Researchers at Columbia University sought to identify the sources of algorithmic bias by tasking 400 AI engineers with creating algorithms that made over 8.2 million predictions about 20,000 people. They concluded that biased predictions are mostly caused by imbalanced data but that the demographics of engineers also play a role (12/9).
- **Google fires Timni Gebru, co-head of Google's Ethical Artificial Intelligence team:** Google terminated and retracted the termination of Timni Gebru,

which she claimed came in response to her refusal to retract a research paper and email sent complaining to her colleagues. CEO Sundar Pichai later apologized for her dismissal and pledged to investigate the event. (12/3, 12/9).

- **AI experts release CodeCarbon, a tool to track and reduce computing computing's CO2 emissions:** Mila, BCG GAMMA, Haverford College, and Comet.ml released CodeCarbon last Tuesday, an open source software package that integrates into Python to estimate the location-dependent CO_2 footprint of computing (12/1).
- **Microsoft releases new tool to help companies regulate data:** Microsoft dropped Azure Purview, a tool that uses AI to detect regulated data, to help data privacy and risk management officials ensure compliance with rules such as the European Union's General Data Protection Regulation (12/3).
- **Facebook to remove false claims and conspiracy theories about Covid-19 vaccines**: Facebook updated its misinformation policy around the coronavirus crisis to include vaccine-related content on Thursday, responding to claims that it wasn't doing enough to protect its users (12/3).
- **Facebook buys a chat box:** The social-media giant closed a deal to buy Kustomer, a startup that uses machine learning to analyze and classify inbound conversations on its customer-service platform, among other capabilities, a deal valued at more than $1 billion (11/30).
- **Google, Facebook and Amazon make gains as coronavirus affects ad spending:** For the first time, more than half of U.S. advertising spending is set to go to digital platforms such as Google and Facebook (12/1).

AI & Public Health 🏥

- **Korean company announces AI-based insurance screening system**: Kyobo Life Insurance says that its screening system uses big data, machine learning and automated algorithms. In the system, risks are assessed by type of claims and eligibility assessment is automatically carried out. It covers six types of claims, including indemnity, hospitalization and outpatient treatment (12/4).

US Federal & State Government News 🇺🇸

- **Trump signs executive order on principles for U.S. government AI use**: The EO involves nine principles for the design, development, acquisition and use of AI in government "to foster public trust and confidence in the use of AI, and ensure that the use of AI protects privacy, civil rights, and civil liberties."
  - "The order directs agencies to prepare inventories of AI-use cases throughout their departments and directs the White House to develop a road map for policy guidance for administrative use" (12/3).
- **IoT is the next step in government tracking:** The Air Force Research Lab is testing an AI-enabled software platform designed to parse data gathered from mobile phones connected to networks of hundreds of millions of computers, routers and other devices linked over the Internet of Things (11/27).
- **Facebook, Google to Face New Antitrust Suits in U.S.:** Federal and state antitrust authorities are preparing to file up to four new lawsuits against Facebook and Google by the end of January (11/30).

AI Around the World 🌍

- **EU:**

- EU ends hope for a new EU-U.S. data transfer pact, as concerns over a Biden administration's position on US surveillance remain (12/4).
- U.S. tech giants face tighter regulation in Europe as the European Commission completes regulatory plans outlining how online platforms should "remove illegal content quickly and refrain from using their power to quash rivals or push their own products on their sites." The commission plans sanctions for violators that include fines and separation of assets (12/3).

Op-Eds: Top Picks of the Week 🧐

- **Voice assistants have a gender bias problem. What can we do about it?** As voice-assisted smart devices are on the rise, partially thanks to the COVID-19 pandemic, so does concern over their perpetuation of gender stereotypes. For example: A 2016 Maxus study found that personal assistant bots were almost entirely female, while bots in roles characterized as "serious business"—such as law or finance—were almost entirely male. One solution according to this Brookings review: the pursuit of industry-wide standards (12/9).
- **Why business cannot afford to ignore tech ethics**: The *Financial Times* puts forth the following argument in this op-ed:
  - "This developing focus on ethics is no fad, and companies that treat it as such put themselves at considerable risk. Leaving aside legislative changes (a possibility, though a slow-moving one in most jurisdictions), public censure and potential loss of business is enough to require a close look at what systems are in place. It will not be enough simply to blame the creators of an algorithm that proves to be biased." Some tools they suggest to mitigate unethical AI include an ethics return-on-interest calculator, developed by tech ethics consultancy Hattusia and the National Health Service's assessment template for healthcare providers purchasing AI products. (12/6).
- **How our data encodes systematic racism:** Deborah Raji of Mozilla asks readers "what is the difference between overpolicing in minority neighborhoods and the bias of the algorithm that sent officers there? What is the difference between a segregated school system and a discriminatory grading algorithm? Between a doctor who doesn't listen and an algorithm that denies you a hospital bed? There is no systematic racism separate from our algorithmic contributions, from the hidden network of algorithmic deployments that regularly collapse on those who are already most vulnerable" (12/10).
- **We read the paper that forced Timnit Gebru out of Google. Here's what is says:** MIT Tech Review contributors examine the research paper that purportedly led to the termination of Google ethics researcher Timnit Gebru. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" lays out the risks of large language models. It presents the history of natural-language processing, an overview of four main risks of large language models, and suggestions for further research (12/4).

Upcoming Events 🥳

- Recorded event from December 10: The Business Case for AI Ethics

- Future of Speech Online virtual forum, hosted by the Center for Democracy and Technology's Future: <u>December 11.</u>

And a h/t to the <u>WSJ Pro's Artificial Intelligence Weekly</u> for concisely summarizing our statement of purpose here at EqualAI:

"Companies have been quick to embrace artificial intelligence in efforts to revamp corporate strategies, or entire business models. But in the race to get tech capabilities off the ground, researchers say, a surprising number of firms aren't taking the time to consider the broad range of risks involved in handling massive amounts of data—and using it to drive strategic decisions. For a game changer as sweeping as AI, potential hazards can include everything from regulatory compliance to privacy protections and climate change." - Angus Loten