

# The Bulletin

On Bias & Artificial Intelligence

Issue 14, December 2020

## Spotlight of the Week: Our Forum Panelists



*"Bias can enter AI through each human touch point throughout a product's lifecycle-- from a lack of diverse perspectives in the design to the bias baked into the data from which it learns and iterates.*

***But each touch point is also an opportunity to identify & eliminate bias."***

**Join the [EqualAI Forum](#) this Friday to hear from Heather Cox of *Humana*, Malcolm Frank of *Cognizant* and Alex Spinelli of *LivePerson* on how to strategically and effectively address bias in your AI systems.**

### Hot Topics in Business and Technology 🔥

- **Bots and voice assistants embed gender bias:** A Brookings report examines how AI, particularly customer-facing service robots, reinforce harmful gender portrayals with traditionally female attributes (11/23).
- **Smiling faces cause bias in algorithms:** MIT-affiliated researchers identified a new challenge for facial recognition systems: bias toward facial expressions, particularly smiles, the expression in the photos most often shared in databases. (11/23).

- **Law360 discusses AI ethics and law:** Law360 released an analysis on how the principles informing AI ethics can be applied to law firms looking to appropriately deploy AI applications in their business. One approach they recommend to avoid discriminatory AI is to assemble AI committees within firms to merge technology and legal teams (11/23).
- **Facial recognition raises gender bias concerns:** A new study on image recognition offered by Google, Microsoft, and Amazon characterized women more often by their appearance. Looking at women lawmakers, the systems often tagged them with labels such as “girl” and “beauty.” The programs also failed to detect women more often than they failed to see men (11/19).
- **New research points to fundamental flaw in AI training:** A group of researchers across Google have identified a new cause for the common failure of machine-learning models: “underspecification.” The tests can miss arbitrary differences when applied in real world settings for which it hasn't been trained. (11/18).
- **ICE reignites controversy as it looks to hire Microsoft and Amazon:** U.S. Immigrations and Customs Enforcement is planning to hire Microsoft and Amazon’s cloud computing services for 2021, and internal scrutiny from employees is likely (11/17).
- **Twitter will relaunch account verification in 2021:** Following ethical concerns in 2017 after verifying the account of the person who organized the deadly white supremacist rally in Charlottesville, Twitter announced its plans to release new verification policies next year (11/24).
- **Facebook and hate speech:** The social media platform claimed that its AI systems have improved in their detection and removal of hate speech, estimating that out of the hate speech content the site took action on, 95% was proactively identified and stopped by artificial intelligence (AI) without a user reporting it (11/19).
- **The rise of the generative adversarial network:** The *NYT* created an interactive AI system to demonstrate the ease of generating fake faces using a new AI technology called a generative adversarial network (GAN) software, which is increasingly being sold to indicate fake virtual demographics and other more nefarious purposes (11/21).

## AI & Public Health

- **AI models predicting the pandemic’s end are a “health hazard”:** Authors warn that algorithms being used to model the pandemic’s lifetime and near “end” should not be used to inform policy decisions given that they are false and misleading. Additionally, they cite a recent paper in the *Journal of American Medical Informatics Association* finding that biased models may exacerbate the disproportionate impact of the pandemic on people of color. (11/20).

- **Health from your smartphone:** Two new Israeli apps that use AI and your phone's cameras to screen for UTIs and kidney disease have been launched in the UK to help diagnose patients who are homebound during lockdown (11/20).

## US Federal & State Government News

- **The Pentagon and Howard University partner to create AI Center of Excellence:** The DoD, Army and Howard University announced their collective effort to push machine learning-rooted research and technologies with a \$7.5 million center of excellence (11/25).
- **NYC:** NYC's city council is trying to restrict the use of algorithms to screen job applicants, considering a bill that would require potential employers to notify job candidates about the use of "automated decision systems." Companies must also complete an annual audit to demonstrate the technology doesn't result in bias (11/20).
- **Top intel official expresses concerns about bias in military AI:** The deputy chief of staff for intelligence, surveillance, reconnaissance and cyber discusses algorithmic bias concerns in machine learning used by military systems. "[W]hat if the military builds an intelligence algorithm that is unintentionally biased toward Russian intrusion methods versus Chinese or Iranian... What if the military builds an algorithm to locate ballistic missiles, but developers only used North Korean imagery data to train it? Will it be able to accurately locate ballistic missiles originating from other adversaries?" (11/18).

## AI Around the World

- **Australia:** The Australian Human Rights Commission released a new paper *Addressing the problem of algorithmic bias* for companies. Their approach includes acquiring more "appropriate" data, increasing model complexity, and modifying AI to consider societal inequalities (11/23).
- **UK:**
  - **CDEI:** The UK government's advisory body on the responsible use of AI released a major proposal for government measures to mitigate AI bias. One key recommendation includes a "mandatory transparency obligation on all public sector organisations using algorithms that have an impact on significant decisions affecting individuals" (11/27).
  - **New Digital Market Unit (DMU):** As part of the nation's plans to regulate big tech, the UK will set up a new Digital Market Unit (DMU) to oversee a "pro-competition" regime for internet platforms — including those funded by online advertising, such as Facebook and Google (11/27).
- **Finland:** The University of Helsinki launched a new Ethics of AI online course to help the public and businesses understand what the ethical use of AI means (11/27).

- **Singapore**-based company BasisAI has released an ML software, Bedrock, that enables organizations to “mitigate risk, ensure fairness through easy detection and correction of unintended bias that can creep into AI systems, and ultimately develop trustworthy AI.” (11/25).

### Op-Eds: Top Picks of the Week 🤖

- **Ethical AI isn't the same as trustworthy AI:** In this VentureBeat opinion, Kimberly Nevala addresses consumer “trust” in AI, as distinguished from ethical AI. She argues that “ethics are the guiding rules for the decisions we make and actions we take. These rules of conduct reflect our core beliefs about what is right and fair. Trust, on the other hand, reflects our belief that another person — or company — is reliable, has integrity and will behave in the manner we expect”(11/28).
- **We're Still Smarter Than Computers:** Computer science professor Dr. Melanie Mitchell argued in 2019 against letting the fears of futuristic “superintelligence” affect policy decisions. 12 months later, Shira Ovide examines the latest AI developments. According to her, the promise that technology will “out-human humans” is still only a maybe (11/25).
- **AI Fairness Review:** This review from Analytics India puts forth the “top” 8 initiatives by large tech firms to “ensure fairness in AI.” It includes Facebook’s AI-ethics based research center, Google’s defining AI principles, and an internal ethics board by IBM, among others (11/20).
- **9 Black Women That You Should Know in Data Science:** In honor of #BlackinDataWeek, the Sadie Collective offers nine data scientists you should know and how they’re shaping the field of AI & AI ethics (11/20).

### Upcoming Events 🎉

- **EqualAI Forum** - Join executives from leading global companies and industry experts to discuss bias in AI systems — why and how to address it and where to start: December 4.
- Assessing the AI Agenda: Policy Opportunities and Challenges in the 117th Congress, hosted by The Wilson Center’s Science and Technology Innovation Program: December 3.
- Human work in the age of smart machines: A conversation with James Merisotis, hosted by the Brookings Institution: December 3.
- The Future of Speech Online 2020, hosted by the Center for Democracy and Technology: December 7.
- Women in Tech Regatta, hosted by Amsterdam/EMEA: December 7.