

The Bulletin

On Bias & Artificial Intelligence

Issue 13, November 2020



Spotlight of the Week: Our EqualAI Board

Robert LoCascio, Karyn Temple, Miriam Vogel, Malcolm Frank, Monica Greenberg (L/R)

Hot Topics in Business and Technology 🔥

- **NIST to release AI Ethics Guidelines in 2021:** The National Institute of Standards and Technology (NIST) will release guidance on implementing trustworthy AI technologies in the coming months (11/13).
 - Elham Tabassi, chief of staff at NIST's Information Technology Laboratory, told industry event attendees that the agency's goal is to establish standards that will help the international AI community better understand the issue of ethical bias in algorithms and data.
- **Messaging App Telegram fails to respond to deepfake porn condemnation:** Researchers hoped to eject a bot, which has not yet happened despite mounting pressure, after demonstrating creation of non-consensual nude images of women. More than 100,000 such images have been publicly shared through the bot's messaging app, Telegram. (11/18).

- **CSET on AI and Cybersecurity: report looks at how machine learning could reshape cyber attacks.** CSET released a report to encourage considerations of advancing technologies in machine learning when developing security measures against cyber attacks. The report also highlights how AI platforms could actually improve the discovery of the software vulnerabilities that enable cyber operations (11/19).
- **YouTube updates Terms of Service:** YouTube announced more explicit Terms against the collection of facial recognition information on its platform (11/18).
- **Google drops system to mitigate AI bias:** Google released MinDiff, a new framework for mitigating unfair biases when training AI and ML models. The model's training objective is augmented with an objective focused on removing biases. This new objective is then optimized over a small sample of data with known demographic information. Given two data slices, MinDiff penalizes the model for differences in the distributions of scores between the two sets to minimize the penalty by bringing the distributions closer together as the model trains(11/16).
- **The latest on Big Tech in Politics:**
 - Along with Google, Facebook confirmed last week that it would extend its political advertising ban over the post-election period (11/11).
 - Facebook is using machine-learning algorithms to prioritize which posts human content moderators should review first, rather than handling them in chronological order (11/13).
 - Snapchat: over 30 million Snapchat users used its voter engagement programs, registering to vote over 1.2 million users (11/16).
- **Mounting calls for greater AI transparency:** A growing body of scientists expressed frustration at large tech firms for refusing to share the underlying code and processes in AI developments. They say it questions the science behind the models, making it more difficult for experts to assess for bias and safety. In particular, the report calls out OpenAI and DeepMind for keeping code under wraps.(11/12).

AI & Public Health 🏥

- **Government to pay doctors using AI in treatment:** The Centers for Medicare & Medicaid Services announced its plans to pay for doctors using algorithms for diabetes diagnoses and stroke detection (11/10).
- **AI improves pregnancy health:** Research from the University of Seville shows that disorders such as congenital heart birth defects or macrosomia, gestational diabetes and preterm birth can be detected earlier with AI (11/6).
- **Health execs predict “widespread implementation of AI” in next few years:** A Black Book Research survey of 540 US health IT executives reports that 47% of US healthcare organizations are using AI, 90% of respondents expect widespread implementation of AI in the next five years (11/6).

US Federal & State Government News 🇺🇸

- **Democratic senators demand Facebook enforce its hate speech policies on anti-Muslim posts:** Sen. Chris Coons led an effort, including Senators Mark

Warner, Amy Klobuchar, and Elizabeth Warren and other senators, to send a letter to Facebook expressing "deep concern" and requesting that the social network better audit and monitor its practices.

- **TikTok:** The US Committee on Foreign Investment extended the deadline for TikTok to become a US company until Nov. 27, as Oracle and Walmart plan to own minority shares (11/13).

AI Around the World 🌐

- **UK:** Synthesized, a London-based startup, has launched a tool that identifies and mitigates bias in datasets for scientists working with AI programs. Their software scans datasets in minutes, provides an analysis of the way different groups of people are identified within that dataset, and generates synthetic data at the end of the process to re-balance the dataset, filling in the gaps where bias was identified (11/13).
- **Luxembourg:** The government of Luxembourg announced plans to “stay ahead of the [ethical] risks” of AI by forming its AI Legal and Ethics Committee: a team of legal experts on data and technology.
- **Thailand:** National AI governing ethics principles were approved by the prime minister and will head to the cabinet for review/endorsement (11/17).
- **Europe:**
 - **Cities:** Local European governments are undertaking an effort to allay public concerns over bias and privacy in the AI algorithms used to help run city services. For example, Amsterdam launched a website listing algorithms it uses, including a camera system that monitors social distancing during the Covid-19 pandemic (11/11).
 - EU: Officials announced plans to propose legislation in 2021 addressing privacy and discrimination risks associated with the use of AI in public services, such as facial recognition.
 - **Privacy:** Activist group NOYB filed complaints in Germany and Spain against Apple for using software that tracks how iPhone users utilize their devices (11/16).

Op-Eds: Top Picks of the Week 🤖

- **You can't eliminate bias from machine learning, but you can pick your bias:** The author summarizes in “a nutshell, data science is hard, machine learning is messy, and there is no such thing as completely eliminating bias or finding a perfect model”. The intent is not to discredit AI use but to argue that there “needs to be more awareness of how bias functions.. When we bring awareness to data science and model creation, we can make informed decisions about what to include or exclude, understanding that there will be certain consequences”. 11/14
- **Health Care AI Systems are Biased:** A *Scientific American* opinion discusses concerns over algorithms trained with imbalanced data in health care such as chest x-rays for an underrepresented gender and skin-cancer detection algorithms that perform with less accuracy with darker skin. Part of the problem is that data used to train most health AI came from just three states, California, New York and Massachusetts, as cited in JAMA. They recommend more diverse data to avoid perpetuating inequality in medicine (11/17).
- **Why AI can't move forward without diversity, equity, and inclusion:** Evolve, a new VentureBeat forum on bias in AI on 12/8, will highlight the challenges of bias in AI. One of the speakers, Huma Abidi, senior director of AI software

products and engineering at Intel shares her steps to combat bias and promote equity in AI (11/12):

- 1) companies have to think about who their target population is and whether the data is representative of the target population.
 - 2) Companies need to ask if they have the right interdisciplinary team, including personnel such as AI ethicists, including ethics and compliance, law, policy, and corporate responsibility.
 - 3) Companies need a “measurable, actionable de-biasing strategy that contains a portfolio of technical, operational, organizational actions to establish a workplace where these metrics and processes are transparent.”
- **Tech and Ethics: The World Economic Forum’s Kay Firth-Butterfield on Doing the Right Thing in AI:** WEF Chief AI ethics officer discusses the democratization of AI, the values of good governance and ethics in tech, and the importance of having people understand the technology across their organizations on an MIT Sloan Management Review podcast (11/17).
 - **The Women Influencing AI:** Forbes presents a list of female “thought leaders” in the AI/ML space who have “paved the way” for females entering the work space. Among them is Timnit Gebru, the Technical Lead of Google’s Ethical AI Team (11/17).

Upcoming Events 🎉

- “Designing Ethical AI and Healthcare Technologies,” a Fireside Chat with the Professor & Director of the Bioethics Program, Editor-In-Chief of the Canadian Journal of Bioethics, hosted by Impetus: [Nov. 24](#).
- Ethics of AI Webinar Series, hosted by the University of Helsinki and the Finnish Center for AI: [Nov. 27](#).
- Evolve: Ensuring Diversity, Equity, and Inclusion in AI: a digital series on how to ensure racial fairness and equity while building products, teams, and companies with AI, ML, and large data: [Dec. 8](#).