Statement by
Meredith Broussard
Associate Professor, New York University
Research Director, NYU Alliance for Public Interest Technology
before the
Task Force on Artificial Intelligence
of the
Committee on Financial Services
U.S. House of Representatives

Representative Foster, members of the Task Force, thank you for hosting this important hearing on ethics in artificial intelligence, and for giving me the opportunity to submit this testimony. My name is Meredith Broussard and I am an associate professor at the Arthur L. Carter Journalism Institute of New York University, the research director at the NYU Alliance for Public Interest Technology, and an affiliate of the NYU Center for Data Science. I'm also the author of the book, *Artificial Unintelligence: How Computers Misunderstand the World,* which has been widely adopted as a text in AI ethics courses. I began my career as a computer scientist at AT&T Bell Labs and the MIT Media Lab before turning to journalism, where I now teach investigative journalism using data and code. As part of my research, I create artificial intelligence for investigative reporting, and I do a lot of science communication work around computational literacy in order to empower people to understand the algorithms that are increasingly used to make decisions on our behalf. I also consult on algorithmic audits of commercial systems; I am working on developing a regulatory sandbox in order to audit AI systems for legal compliance; and I founded a summer program for early and mid-career scholars called the NYU Institute for Public Interest Technology.

In this testimony, I'm going to explore a practical vision for regulating artificial intelligence that builds on the wide-ranging testimony that has already been presented before this Committee. I'll do a few things:

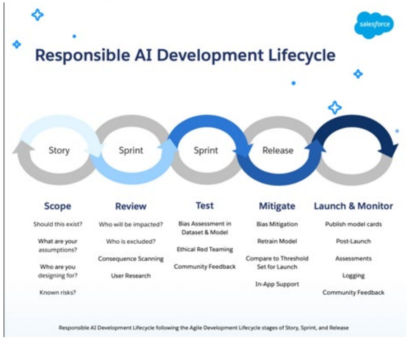
- Explain what AI is and isn't
- Talk about discrimination by default
- Talk about algorithmic auditing
- Explain that some frameworks for AI ethics exist, and note how they can be integrated into business processes
- Talk about regulatory sandboxes, which are a promising development for auditing and compliance

The first thing I want to say is that AI is not what we see in Hollywood depictions. There is no robot apocalypse coming, there is no Singularity, we do not need to prepare for artificial general intelligence (AGI) because these things are imaginary. What is real is that AI is math. General AI is the Hollywood sci-fi version, and it is entirely imaginary. Narrow AI is what we have, and it is very complicated and beautiful math. It is math that is computed on machines. I say this in order to underscore the fact that computing is a terrestrial process; it does not take place in a literal cloud. The cloud is someone else's computer. The process of "doing AI" is something that humans do with machines. As a sub-field of computer science, AI itself has many sub-fields. Machine learning is currently the most popular. Machine learning is a sub-field of AI, the same way that algebra is a sub-field of mathematics. However, the terms "AI" and "machine learning" tend to be used interchangeably today. Both are poorly-chosen terms, because they suggest there is a brain, or sentience, inside the computer. There is not. When we do machine learning, we take a large set of historical data, and instruct the computer to create a model based on patterns and values in that dataset. The model can then be used to predict or make decisions based on past data. The more data you put in, the more precise your predictions will become. Computer scientists refer to this as the “unreasonable effectiveness of data.” However, all historical datasets have bias. For example: if you feed in data on who has gotten a mortgage in the past in the United States, and ask the computer to make similar decisions in the future, you will get an AI that offers mortgages to more white people than BIPOC people.

AI is a terrestrial process, and it needs to be regulated ASAP because it has all of the flaws of any human process, plus some. The previous recommendations offered before this committee have offered detail on additional factors such as identity verification and cybersecurity, which are of course important parts of the landscape in regulating AI.

My current regulatory vision begins with frameworks, high-level governance models that guide a company’s use of AI and data. A company can make sure its frameworks are implemented by performing regular algorithmic audits, ideally using a regulatory sandbox. The process could be monitored by regulators using tools we already have, namely compliance processes inside existing regulatory agencies.

Many frameworks for AI ethics and bias detection exist. Salesforce’s AI ethics lead, Kathy Baxter, has helpfully gathered many of them online. I like the ethical AI checklist developed by Equal AI, an organization that is led by Miriam Vogel, who is also testifying today. NIST is developing an Artificial Intelligence Risk Management Framework intended for voluntary use and to improve the ability to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems. Framework concepts can be integrated into normal business processes. The Salesforce site that I mentioned also includes a diagram showing how bias detection and auditing can be integrated into agile software development methods, as follows:


Responsible AI Development Lifecycle

Responsible AI Development Lifecycle following the Agile Development Lifecycle stages of Story, Sprint, and Release

Every company needs a framework for AI governance, and needs to implement the concepts. The next question becomes: inside a company, which AI needs to be regulated and monitored? This depends on the user and the context. Automated license plate readers used at toll booths by the local department of transportation, with data stored for only a short time, is a reasonable use of AI. Automated license plate readers used by police as dragnet surveillance, with the data stored indefinitely, is an unreasonable use of AI.

The EU’s proposed AI regulation calls for categorizing AI into high and low risk. A low-risk use of facial recognition might be using facial recognition to unlock your phone. This is fairly inoffensive, and there is a fallback (a PIN code) for when the facial recognition technology fails. A high-risk use of facial recognition might be the police using facial recognition on real-time surveillance video feeds. Facial recognition technology has been shown to consistently mis-identify people with darker skin; people of color are at a high risk of being harmed by facial recognition when it is used in policing. High-risk AI would need to be registered and regularly audited to ensure it is not harming citizens. This EU regulation is a good start, and I would recommend the US adopts a similar strategy of characterizing AI as the high-risk and low-risk, and regulating the high-risk uses in each industry.

After deciding which AI gets regulated, it is necessary to look for specific kinds of bias. The process for uncovering algorithmic bias is called algorithmic auditing. O'Neil Risk Consulting and Algorithmic Auditing (ORCAA), a company I consult with, performs bespoke algorithmic audits that are tailored precisely to a company's needs. ORCAA audits algorithms in context, asking how an algorithm might fail and for whom. This is a way of identifying how an algorithm might be racist or sexist or ableist or might discriminate illegally—and once we identify the problem, it can be addressed, or the algorithm can be discarded. Software like Parity or Aequitas or Fairness 360 can evaluate algorithms for one of 21 known kinds of mathematical fairness.

The proposed EU legislation calls for the use of a regulatory sandbox, which I am particularly enthusiastic about. A regulatory sandbox is a protected environment where companies can test their algorithms for bias. If and when the bias is discovered, they can then address the issue in their code and re-run the test until they are in compliance with acceptable thresholds. Currently, heavily regulated industries like insurance make the claim that they are not collecting race data, and thus their AI can't be biased. Other factors like zip codes operate as proxies for race, however. If an AI uses zip code in order to determine the price of an insurance policy, it is using race as a factor, which is a problem. Using a regulatory sandbox would allow an insurance company to see if they are inadvertently using a protected characteristic to make a coverage decision, and would allow them time to address the issue instead of pretending it does not exist. I'm currently working with ORCAA to develop a regulatory sandbox prototype. In our version, regulators would also have a limited view inside the sandbox, to see that companies are auditing their algorithms for bias and fixing the problems that they find. Our concept also allows regulators to see reports showing algorithmic audit results without the companies revealing any trade secrets.

An open secret in the AI world is, everyone knows these systems discriminate. Any conversation about robot apocalypse is a deliberate distraction from the harms that AI systems are causing today, right now. AI is preventing people from getting mortgages: a recent investigation by The Markup found that nationally, loan applicants of color were 40%–80% more likely to be turned down by mortgage-approval algorithms, as compared to their White counterparts. In certain metro areas, the disparity was greater than 250%. When the International Baccalaureate used AI to assign student grades during the pandemic, high achieving low-income students received terrible grades, which prevented them from getting college credits that would allow them to graduate early and incur less student loan debt. AI is used to generate secret predictive consumer scores, like health risk scores, consumer prominence scores, identity and fraud scores, or summarized credit statistics. It is likely that BIPOC people are systematically disadvantaged by most of these scoring systems. AI is used in so-called predictive policing. One particularly egregious example is found in Pasco County, Florida, where the Sheriff's office used AI to generate a list of people who were predicted to be at risk of becoming criminals in the future, though they had done nothing wrong. The police then pre-emptively harassed the people on this list, which included students who were identified based on their educational records. AI is not a magic bullet; it may seem to solve certain business problems, but it inevitably causes new problems and has unintended consequences.

A useful frame is found in Ruha Benjamin's book *Race After Technology*, in which she argues that automated systems discriminate by default. If we adopt this vision, it becomes dramatically easier to spot discrimination and bias inside AI systems. It is not a question of whether, but a matter of looking for the obvious.

Companies should be evaluating the potential benefits, harms, and greater implications of their AI technology, rather than enthusiastically adopting every new technology in a mad scramble to emulate Big Tech firms. I'd like to encourage a space for technology refusal, normalizing and rewarding the firms who refuse to use AI systems that are biased or discriminatory.

I've laid out a vision here. In my vision, companies adopt meaningful AI ethics frameworks; algorithmic audits are seamlessly integrated into business processes; we have a comprehensive regulatory policy that mandates algorithmic auditing for AI; we have regulators who are trained to spot bias in AI systems; and a new kind of technology will have been developed to facilitate and monitor the process. The final piece is education. Companies need to educate their workers in AI. This might mean calling on groups like the NYU Alliance for Public Interest Technology for professional development. It might mean executives reading books like *Artificial Unintelligence*, or *Algorithms of Oppression* by Safiya Noble, to get better informed about the limits of AI and how bias operates inside sociotechnical systems. Education will also be needed around new AI compliance measures, so that people understand better how bias manifests and how to detect it and address it inside AI systems.

The technology to achieve dramatically better algorithmic insight already exists. Various mathematical definitions of fairness exist. Parity and Aequitas and Fairness 360 are all platforms for algorithmic auditing and bias detection. ORCAA's regulatory sandbox builds on their excellent work. All of the necessary pieces have been made incarnate through the hard work and creativity of scholars, activists, and concerned parties. The missing link is the policy mandate. I would welcome the opportunity to talk more about AI regulatory policy. Thank you for the opportunity to testify today on this important topic.