U.S. HOUSE OF REPRESENTATIVES
COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

HEARING CHARTER

Artificial Intelligence: Societal and Ethical Implications

Wednesday, June 26, 2019
10 a.m. – 12:00 p.m.
2318 Rayburn House Office Building

Purpose:

On Wednesday, June 26, 2019, the Science, Space, and Technology Committee will hold a hearing to discuss the impact of artificial intelligence (AI) on society and the ethical implications in the design and use of this technology. The hearing will examine the extent to which AI is already being deployed across different sectors of our society and economy, how biases, vulnerabilities, and other unintended consequences may manifest in these AI systems, and how Federal agencies, as part of their research programs, standards development efforts, and internal adoption of AI, can help ensure more ethical and responsible design and application of AI.

Witnesses:

- **Ms. Meredith Whittaker**, Co-Founder, AI Now Institute, New York University
- **Mr. Jack Clark,** Policy Director, OpenAI
- **Mx. Joy Buolamwini**, Founder, Algorithmic Justice League
- **Dr. Georgia Tourassi,** Director, Oak Ridge National Lab—Health Data Sciences Institute

Overarching Questions:

- In what applications and to what extent are AI systems already in use today? What are examples of AI use that touch people's lives every day that we don't often hear about?
- What are the different ways that bias can manifest in AI systems? What are the consequences of these biases? What are some of the other risks and concerns related to fairness, transparency, trust and other ethical considerations in the application of AI systems?
- How should we assess and address these risks and concerns in AI systems? How can we integrate ethical considerations at the earliest stages of research and education? What is the role of the Federal government in these efforts?

Background

Ubiquity of AI

All applications of artificial intelligence in use today can be considered "narrow AI," or AI that's designed to do a very specific set of tasks. (In contrast, general AI is a system that possesses generalized human cognitive abilities and, when presented with an unfamiliar and complex problem, can develop solutions drawing from contextual knowledge. We are still very far from achieving artificial general intelligence.) Machine learning is a technique most often used to achieve end-user AI applications, and involves developing an algorithmic model based on input data, then using that model to make certain optimizations or predictions. An example of this is image recognition, in which a set of human-labeled images (e.g. "bike", "cat", "lamp") can be fed into an algorithm, which then looks for patterns common to all images with a specific label. The algorithm builds a model ("learns") from this "training data", so when it is presented with an unlabeled image containing one of the objects that was in the training data, it is able to make a guess as to what the object is. This method of training algorithms with human-labeled data is called "supervised learning". There is also "unsupervised learning", in which no labels are provided, and the algorithm simply looks for similarities and groups images into clusters based on certain characteristics.

AI systems have been in use for a while in the commercial sector, the most prominent examples being targeted advertising and financial market predictions. More recently, thanks to rapid advances in computing speed and methodology (e.g. deep neural networks), as well as increasingly larger datasets generated and collected across a variety of platforms, AI-powered systems have grown increasingly capable and widespread. In healthcare, AI systems can aid in medical diagnoses[1,2], perform many duties of clinical assistants[3], and help first responders make critical decisions[4]. In transportation, AI algorithms can help predict and mitigate traffic[5], and autonomous vehicles that use a variety of AI technologies are rapidly becoming more advanced[6]. AI technology used in agriculture can improve crop quality and reduce workloads[7], and AI algorithms are increasingly used in scientific research to help sort and analyze massive amounts of data in fields such as weather prediction[8] and genetics research[9]. Businesses large and small

[1] https://www.nytimes.com/2019/02/11/health/artificial-intelligence-medical-diagnosis.html
[2] http://med.stanford.edu/news/all-news/2019/06/researchers-develop-ai-tool-to-help-detect-aneurysms.html
[3] https://www.meridian.edu/how-artificial-intelligence-ai-is-improving-medical-assisting/
[4] https://www.meritalk.com/articles/u-s-canada-ai-partnership-aims-to-help-first-responders/
[5] https://news.usc.edu/148660/usc-engineers-use-artificial-intelligence-to-reduce-traffic-jams/
[6] https://www.ucsusa.org/clean-vehicles/how-self-driving-cars-work
[7] https://searchenterpriseai.techtarget.com/feature/Agricultural-AI-yields-better-crops-through-data-analytics
[8] https://spacenews.com/ai-for-earth-observation-and-numerical-weather-prediction/
[9] https://www.deepgenomics.com/

are increasingly adopting AI technology to improve performance quality and workflow efficiency—AI analysis has even been used to improve beer brewing[10] and clean cat litter[11].

AI Associated Risks

AI-powered systems have the potential to drastically improve our lives, but also the potential to do significant harm if they are not vetted for bias and fairness. (This hearing is primarily focused on civilian and commercial uses of AI with a presumption of no intent to harm. There are many scenarios in which AI can be intentionally misused or abused.) There are many different causes and manifestations of bias, the most straightforward of which is bias in training data. An AI algorithm's performance depends heavily on the quality of its training data. In the image recognition example above, if the tagged training dataset included mostly cats but only a few dogs, the algorithm will be able to identify cats much better than dogs. In more practical examples, a self-driving car trained by driving on the roads of Boston may not recognize different patterns in other cities, and an AI diagnostic tool trained on x-ray images of younger patients may fail to perform well on older patients. Training data bias can have significant social implications as well. Facial recognition systems trained on mostly light-skinned faces have performed much worse in identifying faces with darker skin[12]. When such systems are used in law enforcement to, for example, identify criminal suspects from video footage, it can lead to a higher number of false arrests for people with darker skin.

One solution to this problem can be to "de-bias" the data by making sure the data is representative of real life. However, such an approach can quickly exacerbate societal biases, because real life data reflect existing social norms and structures of power. Targeted job advertising services have shown to men advertisements related to higher paying jobs than what is shown to women[13]. When a user searches black-identifying names in Google, they are more likely to see arrest-related ads than when searching white-identifying names[14]. Even when AI systems are specifically designed to mitigate human bias, hidden biases can arise. A well-known example is Amazon's attempt to build a resume screening AI algorithm to identify promising job candidates[15]. Part of the goal was to eliminate personal bias from human hiring managers, who might rate applicants higher if the manager relates to them more or if the candidate fits the manager's subjective standards of qualification. The algorithm was trained using resumes submitted to the company over a 10-year period. However, because the tech sector has been severely male dominated over the past decade, the algorithm quickly learned that male candidates were more preferable and demoted any resume that mentions the word "women" or

[10] https://www.forbes.com/sites/bernardmarr/2019/02/01/how-artificial-intelligence-is-used-to-make-beer/#904f8e370cf4
[11] https://www.techradar.com/news/this-ai-litter-tray-analyzes-your-cats-health-and-uses-nasa-tech-to-clean-itself
[12] https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html
[13] https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study
[14] https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/
[15] https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

indicates that the applicant was female (e.g. "captain of softball team" vs "quarterback"), even though the AI was not explicitly programmed to consider gender. Amazon eventually cancelled this project, but if the algorithm had been implemented in real life without being vetted for bias, it would have exacerbated the already significant gender inequality in the tech sector. If a dataset set is carefully curated and vetted for bias and fairness, it could solve some of the issues associated with biases manifesting in AI systems.

There are additional sources of bias that can be introduced in the design phase before an algorithm is ever trained on data. For instance, AI algorithms can be designed to optimize for a small set of parameters without considering the bigger context of the problem. An extreme example is, if an AI is tasked with developing a method to suppress a widespread disease, it might propose to eradicate an entire country's population. In this case, the AI optimized only for disease control without regard to the broader context of the goal, which is to save human lives. Bias can also arise when a measured characteristic is used as a poor proxy for another characteristic. For example, risk assessment algorithms are increasingly used in courts to determine bail or even jail time by evaluating factors such as gender, age, and prior convictions. However, the extent to which each of these characteristics contribute to a person's likelihood to commit a crime is still an active area of research, and therefore the risk assessments already in use are not clearly based on sound science.

These above instances are examples of poor alignment between the task assigned to the AI and the actual human goal. To better align the AI tasks and human goals involves not only technology expertise, but an understanding of the relevant social science and ethical considerations. Bias is not a technical bug, but a social problem. Because humans program AI, the programmers' biases can naturally carry through into the AI system, and it requires an interdisciplinary approach to mitigate these biases.

When AI systems are shown to produce biased results, the systems may be re-trained or re-designed to produce more equitable outcomes. However, biases may remain hidden in the AI "black box." In addition, many users of AI-driven products may lack the awareness and expertise to test for bias or fairness before implementing AI systems– i.e. they may have undue trust in the system. Finally, when biased AI systems are put into applications such as in criminal justice, schools, or financial sectors, the technology can discriminate against many more people and much faster than any one biased individual can, exacerbating existing inequities and perhaps creating new ones. All of these risks are greater when humans are out of the decision-making loop and there is no opportunity for the affected individuals to appeal the AI's "decision" – i.e. there is a lack of transparency[16]. Beyond any one application or algorithm, experts have also raised broader questions of who benefits from AI more so than others, if the widespread deployment of AI could further exacerbate existing inequities due to job loss and disparate access to the benefits of AI, and whether AI tools should be used at all in certain contexts.

[16] https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy

Ethical Design and Deployment of AI

Well before AI systems are deployed in our society, there are many ways in which ethical considerations can be integrated in the research and design processes, as well as in the education and training of the scientists and engineers who will ultimately design these systems.

Researchers deciding on what research questions to pursue and what kinds of systems to design can engage in the exercise of imagining every application – good and bad- to which the research or algorithm may be relevant and every way in which biases may manifest. There is no way to predict all such possible outcomes, but the very exercise of considering possibilities encourages researchers to put their research in a societal context and refine their path for the best possible outcomes.

Computer and data scientists can partner with ethicists, social scientists, legal scholars, and others in the humanities and social sciences to bring to bear their scholarly expertise and perspective in shaping research and designing systems. Diversity in personal experience and perspective is critical to minimizing bias. The representation of women and minorities in AI research and the tech sector more broadly is already very poor, as reported widely in recent years. This lack of diversity in those designing systems manifests in the technology in unintended ways, such as in the Amazon example provided previously.

Real-world environments for AI applications are almost always different from lab settings. Users of AI systems, especially public sector users such as schools, police, courts, and others can rigorously test their systems at the point of use and actively engage with the public in that process to uncover hidden or overlooked biases.

Achieving the responsible design and deployment of AI also requires integrating ethics into technology education at every stage of the AI education pipeline, from K-12 all the way up to current AI developers. It requires viewing AI as an interdisciplinary field rather than a purely technical field. The National Science Foundation (NSF), which funds university research across all non-biomedical disciplines (including social sciences) and also funds numerous STEM education programs, has a critical role in both of these efforts.

Standards around training datasets, performance measures, and best practices for assessing the impact of AI systems could help current AI developers and users design and use AI more responsibly. The National Institute for Standards and Technology (NIST) has begun broad stakeholder engagement in thinking about what standards and frameworks around AI could look like, as part of complying with the Executive Order on Artificial Intelligence[17]. This includes

[17] https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/

holding a workshop with relevant stakeholder[18] and issuing a Request For Information (RFI) regarding AI standards[19].

Many universities and think tanks are already considering the ethical issues surrounding AI R&D. For example, Stanford recently established its Institute for Human Centered AI (Stanford HAI)[20], which aims to take a multidisciplinary approach to AI research by bringing together faculty and researchers from across the university campus. The Harvard Berkman Klein Center and the MIT Media Lab have partnered to create the Assembly program[21], which brings together technologists, business managers, and policymakers to tackle emerging problems related to the ethics and governance of AI. The private sector is also attempting to tackle issues related to AI bias and ethics. Companies such as Microsoft, Google, and Intel have all published their own versions of AI ethics principles[22, 23, 24]. However, these principles are generally abstract and lack concrete governance structures and accountability measures.

Finally, there are also international conversations taking place surrounding the ethics of AI. The Organisation for Economic Cooperation and Development (OECD) recently adopted a set of AI principles for guiding governments in responsible stewardship of trustworthy AI[25]. Many individual countries have also established their own AI strategies that incorporate ethics to various extents. However, similar to private companies' attempts to address AI ethics, many of these plans and principles are high level and abstract, and more concrete steps are still in development.

[18] https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/
[19] https://www.federalregister.gov/documents/2019/05/01/2019-08818/artificial-intelligence-standards
[20] https://hai.stanford.edu/
[21] https://bkmla.org/
[22] https://www.microsoft.com/en-us/ai/our-approach-to-ai
[23] https://ai.google/responsibilities/responsible-ai-practices/
[24] https://newsroom.intel.com/articles/intels-recommendations-u-s-national-strategy-artificial-intelligence/#gs.jzrn1u
[25] https://www.oecd.org/going-digital/ai/principles/