John Simon

Mr. Graham Rutherford

Oxford Scholars

26 March 2023

On Government Regulation of AI

When OpenAI's ChatGPT burst onto the scene in November 2022, it created a massive stir in society, revitalizing long-standing fears about humanity's relationship with Artificial Intelligence, or AI. People's reaction to ChatGPT was so intense that one might believe it was the first major AI to enter the lives of the American people. On the contrary, artificial intelligence technology has been deeply integrated into American society for years now – it's become so prevalent that the CEO of job site Indeed has said that, "the most important role of a job application and resume is to appeal to the automated system that reads it, before it ever gets to human eyes" (Reidy). Between the prevalence of AI and the newfound commotion caused by ChatGPT, it's no surprise that governments around the world are taking notice, and are beginning to propose regulations on AI. The regulation of AI is a complex topic, and questions about what to regulate, how to regulate it, and whether regulation is even necessary, are still debated.

Along with ChatGPT, other AI systems that have made a big splash recently have been image generation models like Stable Diffusion and Midjourney, which can create realistic-looking images based on simple descriptions. What makes these generative models so special – and dangerous – is their ability to create human-like content to exact specifications in the blink of an eye. Between Stable Diffusion for images and ChatGPT for text, it's become dangerously easy to use AI to impersonate real people. This has plenty of real-life implications, with varying degrees of harm. Companies could use AI as a cheap customer service interface, or

thousands of human-realistic twitter bots could be used to spread misinformation across the internet. Some people may feel perfectly comfortable interacting with AI in the same way they would interact with a person, but a significant portion of Americans feel uncomfortable about AI in their daily lives. Some argue that everyone has a right to know when they are interacting with an AI system, or when an AI system is making a decision that influences them. Governments could allay these concerns by mandating that notice is always given to consumers who are interacting with AI systems. For example, mandating that companies report when they use AI to filter through resumes, mandating that news outlets report when their articles are written by AI, and mandating that AI customer service systems notify consumers that they are not human.

Governments may want to ensure that AI, especially AI used in sensitive areas, gives explainable results. Most AI systems are built using complex mathematical structures called neural networks, which are sometimes described as mimicking the structure of a real-life human brain. As AI has developed over time, neural networks have become so widely used that they are almost synonymous with AI itself. Neural networks were used in IBM's chess program DeepBlue, which famously defeated world champion Gary Kasparov in 1997, and they are used today in facial recognition systems, algorithms that predict stock prices, and ChatGPT. Neural networks are so ubiquitous because they are extremely powerful, but with that power comes a cost. Without major modifications, neural networks function as the quintessential black-box system. Computer scientists know what work they can do, how to make them work, and even why they work. However, nobody, not even the computer scientists that create them, can understand why a neural network might give the response it does – the networks are simply too complicated. This black-box nature has led to many stories about AI going wrong unexpectedly. For example, a team of researchers (Esteva) trying to train a neural network to identify skin

cancer were shocked when they found that their algorithm would confidently label any photo of a ruler or meter stick as showing skin cancer, even if the photo had no skin in it at all. It turns out that in the dataset their neural network was trained on, most of the photos of cancerous skin included a ruler for scale (Narlal). On the other hand, the photos of non-cancerous skin, which had nothing to measure anyway, never included a ruler. In essence, the neural network had learned to identify rulers, not cancerous skin. Had the researchers used an explainable AI system, the researchers would have immediately been able to identify the issue, allowing them to take appropriate action much earlier in the development process and ensuring without a doubt that the erroneous algorithm was never used by real doctors. Given their black-box nature, neural networks are not the best choice for every situation, especially those where errors such as this are unacceptable, like automated systems that read over job applications, bank systems used to determine eligibility for loans, or facial recognition systems for law enforcement. When it comes to regulation of AI, making sure that black-box systems are kept out of such sensitive areas is one consideration for governments.

Data privacy has been a well known issue for a while now, but what's not so well known is how closely related it is to AI. In general, AI systems need massive amounts of data to function properly. Sometimes that data can be gathered in a benign way: the world's leading chess engine created a dataset of millions of games simply by playing itself repeatedly and storing the results. However, more often the data is harvested in ways that break the privacy expectations that most Americans have. Things like real-time location tracking, online purchase monitoring, or internet history tracking are disconcerting to many. In addition to basic privacy concerns, that data is commonly fed straight into massive AI systems that are able to glean all sorts of information about a person. At a time when nearly 80% of Americans report concern

over how data on them is collected and used (Kennedy), there is pressure on the government to address these concerns.

One example of government regulation of data collection is the EU's General Data Protection Regulation. The GDPR is described by the EU as "the toughest privacy and security law in the world", and it has earned that title (Wolford). The GDPR seeks to ensure certain data processing practices, like maintaining proper security and encryption of sensitive data, ensuring that only the minimum amount of data is collected for a given task, and ensuring that data is not stored for longer than is necessary for a given task. Crucially, the GDPR puts the onus on the data collectors to prove that they are GDPR compliant: if a company believes they are GDPR compliant but can't show why that is so, they are not GDPR compliant. While some data privacy laws like the Health Insurance Portability and Accountability Act of 1996 (HIPPA), which protects sensitive patient health information, exist in the US, there are no national laws that provide blanket regulation of data in general. The EU's GDPR could act as a blueprint if America decided to implement such wide-ranging protections.

Perhaps the most difficult yet most important way governments might regulate AI is to enforce quality guidelines that ensure AI systems are working as they should. This is an incredibly difficult task for many reasons, but it mostly boils down to the fact that there are so many ways for AI to go awry. These issues can broadly be broken down into three categories: bad data, bad intentions, and bad logic. Given the hugely important role that data plays in AI, bad data can completely cripple an AI system. The earlier story of the skin-cancer detection algorithm is an example of bad data in action. There was an inherent flaw in the dataset – the presence of rulers in only the cancerous skin photos – that on its own was able to completely destabilize the entire algorithm. Recognizing bad data is an extremely difficult task, and it is not

always clear who is responsible for the issues caused by it. Oftentimes, algorithms are trained on data that is curated by a different organization than the one creating the algorithm. If a government were to attempt to make regulations to prevent bad data, it's not immediately clear who would be responsible for a violation: the data curator or the AI creator. Bad intentions doesn't mean evil intentions, but simply that the stated intentions of the AI system don't align with what it is actually trained to do. For example, in 2014 Amazon debuted a new AI system to review job applications that was trained on past company hiring practices (Dastin). Essentially, the AI was taught to mimic the company's recruiting patterns from the last ten years. It was marketed and talked about as being able to pick the most qualified candidates for the job, but in reality that's not what it was being trained to do. It turns out that Amazon's hiring practices from 2004 to 2014 were biased against women, and so that sexism was passed along to the AI recruiter. If the AI had actually been trained to pick the most qualified candidates, say by comparing resumes of past hires to their measured productivity as an employee, things likely would have turned out fine. However, the AI wasn't trained to pick qualified candidates, it was trained to mimic fallible human recruiters. Misaligned intentions are difficult to identify, but rigorous testing, potentially enforced by government regulations, can help identify bias in AI. The final way AI quality can falter is simply with bad logic, which can mean anything where the AI algorithm itself has logical flaws that cause it to fail. A perfect example of this is Tesla's self-described "full self driving" (FSD) system that, despite how it is marketed, Tesla has asserted should not be treated as full self driving. There are plenty of examples of FSD going awry, including abrupt braking in the highway's fast lane and even running into a million-dollar jet while in "summon mode" (Reidy). These logical errors are the result of a system that is seriously untested and flawed. Governments have already begun to regulate FSD, by forcing

Tesla to report crashes involving it and even prosecuting Tesla's overzealous marketing as false advertising.

Given the potential dangers of AI, it's not surprising that there is overwhelming support for regulation: 82% of Americans support government regulation of AI technology. Perhaps more surprising is that when the population is narrowed to American tech experts, a full 91% support regulation (Singer). However, not everyone believes regulation is the best way to control AI. One of the most common objections is that regulations will stall innovation, simply forcing companies to relocate. People argue that if the US regulates AI and countries like China do not, the US will fall behind the curve. Others point to Congress' aging population and well-documented lack of basic tech literacy as evidence that Congress is not up to the task of regulating AI in an efficient and effective manner. Perhaps most interestingly, there are calls that AI regulation is unfair because it would enforce unnecessarily higher standards for AI than human systems. For example, a government might regulate AI systems that filter through resumes to ensure that the system is always able to give a human-interpretable explanation of why a person passed or failed the filter. However, there are no such requirements that humans who read resumes have to have a concrete explanation for every decision. Job applicants have been rejected for many years now with no expectation of an explanation; opponents of AI regulation argue that there is no need for explanations now.

Debates about what, how, and whether to regulate AI are becoming increasingly common. Society must choose to grapple with unexplainable neural networks, or to mandate more complicated systems that give explainable results. It must have an understanding of data privacy concerns and the possible ways to protect the privacy of ordinary people, or risk a life where privacy is sacrificed in the name of big data. It must ensure that AI systems work as

intended, or accept automated systems that give erroneous or even prejudiced outputs. These debates are certainly worthwhile and interesting: the best way for society to come to a satisfying answer to these questions is to openly discuss and debate them. As fears of AI grow, it seems more and more as though AI regulation is only a matter of time.

Works Cited

Wolford, Ben. “What is the GDPR, the EU’s newest data protection law?.” *GDPR EU*, https://gdpr.eu

Castro, Daniel, and Micheal McLaughlin. “Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence.” *International Technology and Innovation Foundation*, 4 February 2019, https://itif.org/publications/2019/02/04/ten-ways-precautionary-principle-undermines-progress-artificial-intelligence/

Auxier, Brooke, et al. “Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information.” *Pew Research Institute*, 15 November 2019, https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/

Kennedy, Brian, et al. “Public Awareness of Artificial Intelligence in Everyday Activities.” *Pew Research Institute*, 15 February 2023, https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities/

Singer, Jeremy. “MITRE-Harris Poll Finds Lack of Trust Among Americans in AI Technology.” *MITRE*, 9 February 2023, https://www.mitre.org/news-insights/news-release/mitre-harris-poll-finds-lack-trust-among-americans-ai-technology

Esteva, Andre, et al. “Dermatologist-level classification of skin cancer with deep neural networks.” *Nature*, 29 June 2017, https://www.nature.com/articles/nature21056#main

Narla, Akhila, et al. “Automated Classification of Skin Lesions: From Pixels to Practice.” *Science Direct*, October 2018, https://www.sciencedirect.com/science/article/pii/S0022202X18322930?via%3Dihub

Dastin, Jefferey. “Insight - Amazon scraps secret AI recruiting tool that showed bias against women”, *Reuters*, 9 October 2018, https://www.reuters.com/article/amazon-com-jobs-automation/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idINKCN1MK0AH?edition-redirect=in

Gold, Aaron. “This Is Your Tesla FSD and Autopilot Crash Mega Thread”, *MotorTrend*, 19 January 2023, https://www.motortrend.com/news/tesla-fsd-autopilot-crashes-investigations/

Reidy, Gearoid, et al. “The CEO of the world’s biggest job portal says he has a solution to the labor shortage: getting rid of the resume.” *Fortune*, 16 November 2021, https://fortune.com/2021/11/16/recruit-holdings-ceo-indeed-glassdoor-labor-shortage-hiring-resumes/