Written Testimony of Jack Clark
Strategy and Communications Director
OpenAI

House of Representatives Oversight & Government Reform Committee
Subcommittee on Information Technology
Hearing on
Game Changers: Artificial Intelligence Part III – AI and Public Policy

April 18, 2018

Chairman Hurd, Ranking Member Kelly, and Members of the Subcommittee, thank you for the opportunity to discuss the crucial subject of artificial intelligence and policy. Today's hearing comes at a critical time for the development of AI: we have a dramatic future ahead of us, filled with opportunities and challenges. I'm going to concentrate on three main areas for this hearing: ethics, workforce issues, and the importance of AI measurement and forecasting. I think these are areas where success will translate into ensuring the US remains a globally competitive place to invent and apply AI. I see this as the start of an important conversation, and one which I hope continues.

Introduction:

First, a bit about me: I work as the Strategy and Communications Director for OpenAI, a non-profit artificial intelligence research company whose goal is to ensure that powerful artificial intelligence benefits all of humanity - both through direct technical work and through analysis of its impacts.

Much of the research OpenAI does today is about pushing the frontiers of AI capabilities in specific areas and our contributions have so far included new algorithms and associated code, new tools used by other researchers, public policy work about some of the challenges and opportunities of the technology, demonstrations of the capability of powerful AI systems via our 'Dota 2' project where we have developed AI systems capable of out-competing human champions at a complex and widely played strategy game, and AI safety which is essentially the study of how to ensure that increasingly powerful systems will continue to be predictable and interpretable in what they're doing and how they are doing it while acting with greater degrees of autonomy.

I also help produce the AI Index, an AI measurement and forecasting project that is part of the Stanford One Hundred Year Study on AI. Working on AI measurement has given me some insights as to the critical importance of the measurement and forecasting of disruptive technologies and has significantly influenced my thinking about ways the government could be more involved in this critical area.

Technical Background:

it's important to clarify why we're paying attention to AI: AI's capabilities also define AI's threats and opportunities, so the rate of progress of AI conditions the environment in which we make policy decisions. As a quick refresher: in 2012 several researchers, including the co-founder of OpenAI Ilya Sutskever, showed that they could use a technology called a neural network to obtain unprecedented results on a widely-studied image recognition task. At that time, their system was able to correctly identify the objects in an image about 85 percent of the time. In the five years that have elapsed since that paper was published, accuracies have climbed to around 98 percent - surpassing the performance of humans that evaluate themselves on this task. We've seen similar trends in speech recognition as well. These innovations aren't limited to rote classification tasks - similar advances have occurred in the area of AI-aided synthesis, with new techniques invented in 2014 leading to unprecedented advances in the ability for computers to create fake images, fake audio and fake videos with levels of fidelity that approach photorealism. All of these advances have relied on the same essential machinery, much of which is available in the public domain and which runs on widely available types of computer hardware. This means progress, if anything, is set to accelerate in the coming years.

Safety and Reliability:

These new capabilities have their own unique flaws which continue to befuddle and challenge researchers, so we must remember that while this technology is capable of amazing feats of 'intelligence' it is also capable of making mistakes that seem alien to its human developers and may limit its deployment in domains that require ironclad guarantees of reliability and predictability, or may lead to the technology causing accidents when deployed. Further research in the field of AI safety (of which OpenAI has made a substantial investment) may deal with (some of) these problems.

Given that, I'll use the rest of this testimony to discuss three key areas with specific policy recommendations where success will let the US lead development of AI, which will let the US coordinate the global response to the challenges and opportunities of the technology. I'll finish my testimony by providing OpenAI's Charter, a document we recently published that governs how we as an organization approach our mission and the broader community.

#1 We need to maximize AI's societal benefits and minimize its potential harms by building strong channels for dialogue between policy-makers and the community of researchers about ethical norms for responsible innovation in AI.

#2 We must support our academic institutions to allow us to meet the evident demands for AI talent and to ensure the US continues to define the frontier of AI research and applications, while also supporting the retraining of American workers to take advantage of this technological revolution.

#3 We must measure the progress and capabilities of AI to guide effective policy making.

#1: Ethical Norms.

Stakeholders in the development and deployment of AI technology, including the research community, actors in the private sector, and policy-makers, must jointly develop a set of ethical norms to govern their conduct and ensure that AI is both developed and deployed responsibly. If we deliberately pursue the creation of ethical norms around the research, development, and deployment of artificial intelligence, then we will be able to shape the culture in which the technology is developed, creating a shared sense of valid and invalid approaches to the technology, with specific best practices for specific areas. If we are successful in this then it will be significantly easier to apply regulations to artificial intelligence's consequences in the future as the community will have settled on a set of pre-agreed upon conventions that can become templates for law.

Without these norms it's likely that AI could be built or used in ways that cause harm to people or destabilize fundamental aspects of civil life. Last year, we saw the emergence of DeepFakes[1], technology based on artificial intelligence that made it relatively easy for people to take the face of a well-known individual, such as a celebrity, and superimpose it onto the faces of actors in pornographic films. This technology made it relatively easy for people to create unpleasant, unethical media, and while there is no indication anyone was harmed as a direct consequence of DeepFakes, I can assure you that the same technology could be used in other contexts - what if similar techniques were used to make it easy to take the face of a politician and superimpose it on another person in another context? We already have some examples of this occurring, such as a recent demonstration at MIT of President Trump's face being mapped onto the face of President Obama[2], and vice versa. The potential ramifications with regard to automated, synthetic propaganda are chilling and real.

At OpenAI, we have been thinking about these issues and, along with a wide set of stakeholders, recently published a report, the Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation[3]. In this report we have tried to think through some of the ways in which today's AI technologies could be re-purposed by malicious actors to cause harm, whether that be by using off-the-shelf AI technology to augment existing hacking techniques, or to use some of the aforementioned synthesis technologies to create more convincing 'advanced persistent threat'-style attacks, or to take soon-to-be-viable cleaning

---

[1] Oberoi, Gaurav. "Exploring DeepFakes." Hackernoon. March. 5, 2018. https://hackernoon.com/exploring-deepfakes-20c9947c22d9

[2] Clark, Jack. A video recording of a demonstration of AI-based facial mapping. Twitter. March. 1, 2018. https://twitter.com/jackclarkSF/status/969285043502878722

[3] Brundage, Avin, et al. "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation." Arxiv. February. 20, 2018. https://arxiv.org/abs/1802.07228

robots and convert them into bomb-delivery systems. Why did we do this? Because we wanted to highlight the ‘dual use’ nature of much of today’s technology; the same AI technology that can be used to diagnose tumors from x-rays can also be used to train systems to surveil or target individuals, and because these capabilities are embodied in software, proliferate widely, run on standard compute hardware, and are developed by a global set of actors, then traditional control regimes and other policy tools don’t seem to apply. Instead, we need to change the ethical norms with which developers approach their work on AI, so that they think twice before releasing something that can be trivially repurposed for negative uses, and so they try to use an ‘adversary mindset’ to view their own contributions through the perspective of a prospective attacker. I think this neatly illustrates how advances in AI capabilities also create new threats which are difficult to deal with through traditional regulatory tools, but may be dealt with by other means.

Another area for ethical norms, which OpenAI works on directly, is ensuring that the global development community is aware of and implements AI safety techniques in their own work to ensure that AI systems act in a reliable way in line with the intentions of their operators.

The government can and should, in our view, become more-active participants in the dialogue around developing shared ethical norms. As we’ve noted, the risks are quite substantial, so much so that we believe they’re worthy of government attention, forecasting, and mitigation. Fostering the discussion around ethical norms is one of the ways that government can help reduce risk, and its impact-to-cost ratio is attractively high. The nature of threat forecasting and norm creation is by its very nature extremely interdisciplinary, and the government is well-positioned to bring together all of the stakeholders for productive discussions that lead to specific recommendations and clearly-identified focus areas. Furthermore, as AI becomes more central to information technology in general, the government will become a direct stakeholder in using it for some of the most sensitive applications possible---eg health care, law enforcement, and national security. This increases the importance of involving the government in these conversations and of government not only helping to define norms, but adopting them itself in its own practices.

Why do we think developing ethical norms will make a big difference? Because we know that it works. In the areas of bias, we have seen similar research in recent years in which people have highlighted how today’s existing AI systems can exhibit biases, and the surfacing of these biases has usually led to substantive tweaks by the operators of the technology, as well as stimulating a valuable research discipline that provides a set of ‘checks and balances’ on AI development without the need for hard laws. We’ve also seen this happen in the area of algorithm deployment where organizations such as AI Now have embedded with communities likely to be affected by the use of automated algorithms and via joint research have devised recommendations that can be applied by the whole of the public sector. The more we can increase participation in these areas, the better prepared we will be for the significant challenges of this technology, and this will let us get ahead of some of its more obvious downsides so that we don’t end up needing to do reactive regulation.

Ethical Norms: Recommendations:

My specific recommendations for next steps here are, beyond hearings like this, to have more dialogue between the AI community and political appointees and staffers here on the hill, as with specific groups within the government's agencies that are currently implementing their own AI plans and initiatives.

I would also be delighted to facilitate a workshop with people here in Washington to better understand the areas where people are particularly concerned about malicious actors re-purposing AI technology. Through such a workshop I would hope to better incorporate the views of the stakeholders represented here into the technical community's research into this important area, and thereby allow us to create norms that are sensitive to concerns you may be hearing from constituents or other parties.

#2: Strengthening the AI workforce.

Compared to 2013, there are now 4.5 times as many US jobs listed on the job search website 'Indeed.com' that require AI skills, according to the AI Index[4]. This is a symptom of a challenge the AI sector in America faces: demand for people with AI skills is outstripping the supply of people with those skills, which is causing the private sector to hire people away from academia and to hire students earlier in their careers, and we are not increasing funding to let academia keep pace with demand from students[5]; these three factors, combined together, potentially weaken AI education in the US in the future. The private sector has independently sought to meet this demand via the emergence of a range of online education courses for AI, and companies such as Google and Microsoft have also released tools and training courses to help other people transition into AI careers. These indicate a huge amount of demand for people with AI skills. Without aggressive investment into enabling basic scientific research, the US risks squandering its opportunity to lead in the development of AI and the education of its most important practitioners, and thereby might miss out on the upsides of the technology as well as the opportunity to shape both national and global regulations and norms for the technology.

AI will have a significant influence on the economy. The one trillion dollar question is whether it is going to be a good influence or a bad one for the workforce. Here, experts are divided. By 2020 the World Economic Forum believes AI may destroy around 7 million jobs worldwide while creating 2 million jobs, while Gartner believes it will destroy around 1,800,000 jobs while creating 2,300,000; by 2030 McKinsey believes it could destroy anywhere between 400 million

---

[4] AI Index 2017 report, page 18. Accessed April. 10, 2018.
http://aiindex.org/#report

[5] In 2011, the University of California at Berkeley received 474 applications for PHDs in artificial intelligence and was able to admit 13 students. By 2018 this had risen to 2229 applications with 46 admitted students, showing how demand has scaled faster than educational capacity.

and 800 million jobs and create anywhere between 555 and 890 million jobs.[6] What these divergent estimates make clear is that AI will cause a significant amount of disruption in the labor markets around the world as some jobs are automated and other, new fields are created. But what we have learned from the current technology revolution is that many of the new jobs created require a significant set of skills than those which are automated. If we are to benefit as a nation and as a global community from AI then we must make sure this revolution is an inclusive one in which we equip as many people as possible with the skills needed to benefit from the changing economy, rather than be sidelined by it.

Strengthening the AI Workforce: Recommendations:

One of the best ways to support the basic research ecosystem - which all commercialization depends on - is to increase the number of funded PHD fellowships for AI across the country. This will make it easier for the United States to maintain its position as the global leader for AI education and will increase the chance of the world's next great artificial intelligence companies being founded here. We should pay particular attention to ensuring we fund students applying from abroad, as AI is of such opportunity it would be sensible to ensure that the US can educate and support the smartest people in the world. Government could choose to directly endow more funded PHD positions at universities that have displayed excellence in AI. This would have immediate beneficial effects and would avoid the additional overhead introduced by increasing grants which will generate significant logistical overhead on the part of professors that wish to apply for the new funding.

Government should take steps to support further commercialization and spin-outs of University research to further our already diverse ecosystem of AI startups, potentially via providing specialized funding to public universities to allow them to do this. As an example, the University of California at Berkeley recently formed The House[7], a startup commercialization and funding entity which launched in 2016 and has subsequently backed more than 50 startups which have collectively raised over $400 million. If government were able to supply relatively small amounts of capital to let public universities spin-up other, similar initiatives then it's fairly likely that the private sector would respond with additional funding and interest.

The government should invest more heavily in retraining programs both for its own workforce as well as for the workforce across America. We already know that in our current technology revolution the gains have been unequal, with certain regions, such as Silicon Valley, or the upper Pacific Northwest, or the North East of the United States, benefiting from job creation, and other regions languishing as jobs are automated and not replaced with new opportunities. This

---

[6] Winick, Erin. "Every study we could find on what automation will do to jobs, in one chart". MIT Technology Review. January. 25, 2018.
https://www.technologyreview.com/s/610005/every-study-we-could-find-on-what-automation-will-do-to-jobs-in-one-chart/

[7] The House. Accessed April. 16, 2018.
https://thehouse.build/

lack of equality drives discontent and reduces the faith of the worst-affected citizens in their political representatives which ultimately leads to a rise in extremism - we must avoid this.

#3 Measurement and AI Progress.

In my work on the AI Index it has become clear to me that we currently lack a bunch of the basic inputs we'd need to measure aspects of AI and how it is likely to evolve in the future. These measurement challenges include:

- A lack of knowledge about the rate of progress of self-driving cars as the companies developing them are keeping performance metrics secret as they view this as revealing proprietary data about their systems.
- A lack of standardized environments in which to test and evaluate robotics research and applications.
- Little information at both a state and country-level about the deployment of AI systems by both the government and the private sector.
- Little coordinated evaluation of the readiness of AI to be deployed in transformative areas like healthcare, and so on.

While I hope that the AI Index will motivate some further data gathering in these areas, I think the task is so large that it's an area where government could - and should - be more involved. If more people in government were focused on measuring and assessing the impact and progress of AI, then there would be more people in government aware of its progression and able to create smart policy to ensure it benefits as many people as possible.

Today, agencies like DARPA and IARPA are performing regular, subject-specific assessments of aspects of AI via hosting competitions and funding research, and NIST is developing some of the standards and assessment metrics as well. There are also bodies like the Congressional Research Service and the GAO which are looking at this area - last year I participated in a GAO-led report[8] to try to assess AI's impacts and opportunities in a few areas across America.

Measurement and AI Progress: Recommendations:

The government should create and host more competitions to galvanize activity in the academic and private sectors, as DARPA did with its 'Cyber Grand Challenge' initiative in 2016, or DIUx did this year with its 'xView' dataset release. In particular, I can imagine that competitions relating to robot manufacturing, drone navigation and control, and the safety and predictability of AI systems would also serve to catalyze progress which would lead to commercial innovations and an increase in the robustness and usability of the technology. We should fund and encourage organizations, including those named above, to conduct more competitions in this area.

[8] GAO, "Artificial Intelligence: Emerging Opportunities, Challenges, and Implications". March. 28, 2018. https://www.gao.gov/products/GAO-18-142SP

Additionally, we should be providing further funding for long-term AI measurement and analysis schemes, which could include empowering a specific agency to be the pre-eminent measurer of AI. For this, NIST may be a logical agency.

We may also want to explore even more unconventional ideas - for instance, might it be worth reviving the Office of Technology Assessment? OTA operated between 1972 and 1995 and during its lifetime produced more than 700 reports on technology areas of interest to the government. Having that kind of unbiased, bipartisan, science-driven organization would benefit the US today, as it would provide another means by which the government can educate itself about AI. Regardless, OTA provides a template for some of what can be achieved given a sufficiently powerful mandate and specific focus.

In particular, we should be seeking to measure and analyze the impact of AI on the economy and the workforce as, as discussed in section #2, this is an area with a diverse range of possible outcomes and where more knowledge will better prepare us as a nation to take advantage of this technology.

As well-meaning as I and my fellow panelists for this hearing are, I have a strong belief that there's no greater means of guaranteeing that you're getting the most correct view on a subject than by learning about it yourself. If the US government can take on a global leadership role in the measurement and forecasting of AI then it will be better positioned to identify the technologies challenges and opportunities, will be well positioned to create the competitions, datasets, and challenges that motivate progress in key areas, and can define global standards for what we, as a global community of actors, should be measuring and forecasting to ensure the continued stability of the world. There should be a concerted cross-agency effort to systematically analyze, measure, and forecast aspects of artificial intelligence deemed critical to the areas the agencies have purview over, and in areas that representatives hear frequent questions from constituents about.

Closing Statement:

If America can lead the world in the shaping of ethical norms around AI, strengthen its education system to ensure it remains the best place to study and develop AI technology, and ensure that the US government can be the most informed government in the world about AI's progress, then I think we'll be positioned to help not just American citizens, but the entire world benefit from this technology. The choice is, thankfully, ours to make.

Finally, I shall enclose the text of the OpenAI Charter, a document which we recently published that commits our own organization to a set of principles with which we'll approach the development of this technology. I think that another norm that we should all work on creating is making it easier and more acceptable for actors in the private sector to publicly commit themselves to standards that befit the immense impact of this technology.

OpenAI Charter:

OpenAI's mission is to ensure that artificial general intelligence (AGI) — by which we mean highly autonomous systems that outperform humans at most economically valuable work — benefits all of humanity. We will attempt to directly build safe and beneficial AGI, but will also consider our mission fulfilled if our work aids others to achieve this outcome. To that end, we commit to the following principles:

Broadly Distributed Benefits

- We commit to use any influence we obtain over AGI's deployment to ensure it is used for the benefit of all, and to avoid enabling uses of AI or AGI that harm humanity or unduly concentrate power.
- Our primary fiduciary duty is to humanity. We anticipate needing to marshal substantial resources to fulfill our mission, but will always diligently act to minimize conflicts of interest among our employees and stakeholders that could compromise broad benefit.

Long-Term Safety

- We are committed to doing the research required to make AGI safe, and to driving the broad adoption of such research across the AI community.
- We are concerned about late-stage AGI development becoming a competitive race without time for adequate safety precautions. Therefore, if a value-aligned, safety-conscious project comes close to building AGI before we do, we commit to stop competing with and start assisting this project. We will work out specifics in case-by-case agreements, but a typical triggering condition might be "a better-than-even chance of success in the next two years."

Technical Leadership

- To be effective at addressing AGI's impact on society, OpenAI must be on the cutting edge of AI capabilities — policy and safety advocacy alone would be insufficient.
- We believe that AI will have broad societal impact before AGI, and we'll strive to lead in those areas that are directly aligned with our mission and expertise.

Cooperative Orientation

- We will actively cooperate with other research and policy institutions; we seek to create a global community working together to address AGI's global challenges.
- We are committed to providing public goods that help society navigate the path to AGI. Today this includes publishing most of our AI research, but we expect that safety and security concerns will reduce our traditional publishing in the future, while increasing the importance of sharing safety, policy, and standards research.

Committee on Oversight and Government Reform
Witness Disclosure Requirement — "Truth in Testimony"

Pursuant to House Rule XI, clause 2(g)(5) and Committee Rule 16(a), non-governmental witnesses are required to provide the Committee with the information requested below in advance of testifying before the Committee. You may attach additional sheets if you need more space.

Name:

1. Please list any entity you are representing in your testimony before the Committee and briefly describe your relationship with each entity.

Name of Entity	Your relationship with the entity
OpenAI	Employee, Strategy and Communications Director

2. Please list any federal grants or contracts (including subgrants or subcontracts) you or the entity or entities listed above have received since January 1, 2015, that are related to the subject of the hearing.

Recipient of the grant or contact (you or entity above)	Grant or Contract Name	Agency	Program	Source	Amount

3. Please list any payments or contracts (including subcontracts) you or the entity or entities listed above have received since January 1, 2015 from a foreign government, that are related to the subject of the hearing.

Recipient of the grant or contact (you or entity above)	Grant or Contract Name	Agency	Program	Source	Amount

I certify that the information above and attached is true and correct to the best of my knowledge.

Signature Jack Clark Date: April 16, 2018

Jack Clark is the strategy and communications director for OpenAI, where he focuses on AI policy and strategy. He frequently participates in fact-finding studies and forums relating to AI, including events in recent years with the GAO and the Army Cyber Institute. He recently joined the Center for a New American Security task force on AI and national security. A frequent public speaker, Jack has given numerous talks about artificial intelligence and its impact on policy, ethics, and security, with recent talks covering AI and dual-use for a CNAS event in 2017, and issues of AI policy for the opening keynote of a Princeton conference on AI and Ethics in March 2018. He also helps run the AI Index, an initiative from the Stanford One Hundred Year Study on AI to track and analyze AI progress. In addition, he writes a weekly newsletter about cutting-edge AI research and applications called Import AI (www.importai.net), which is read by more than ten thousand experts around the world.