

AI RFI Responses, October 26, 2018

Update to the 2016 National Artificial Intelligence Research and Development Strategic Plan RFI Responses

DISCLAIMER: The [RFI public responses](#) received and posted do not represent the views and/or opinions of the U.S. Government, National Science and Technology Council (NSTC) Select Committee on Artificial Intelligence (AI), NSTC Subcommittee on Machine Learning and AI, NSTC Subcommittee on Networking and Information Technology Research and Development (NITRD), NITRD National Coordination Office, and/or any other Federal agencies and/or government entities. We bear no responsibility for the accuracy, legality or content of all external links included in this document.

Subject: RFI Response: National Artificial Intelligence Research and Development Strategic Plan
From: Jack Clark

Hello,

Please find within the body of this email OpenAI's RFI Response regarding the National Artificial Intelligence Research and Development Strategic Plan. <https://www.federalregister.gov/documents/2018/09/26/2018-20914/request-for-information-on-update-to-the-2016-national-artificial-intelligence-research-and>

RFI Response: National Artificial Intelligence Research and Development Strategic Plan

Submission on behalf of OpenAI[1].

This submission proposes one significant augmentation of the existing AI strategic aims, and proposes the addition of a new one. This submission is from OpenAI, an artificial intelligence research organization in San Francisco which conducts research at the frontier of artificial intelligence. OpenAI's goal, as stated in the organization's Charter, is to ensure that increasingly powerful AI systems benefit all of humanity.

OpenAI's mission is to ensure that "artificial general intelligence - by which we mean highly autonomous systems that outperform humans at most economically valuable work - benefits all of humanity"[2].

Therefore, our policy recommendations are designed to strengthen the US's AI capacity in such a way to benefit both the nation as well as creating the infrastructure for further productive international collaboration on AI, giving America the chance to play a stabilizing role in an increasingly fast-paced AI technology development environment. As former Secretary of the Navy Richard Danzig recently cautioned: "superiority is not synonymous with security"[3].

Recommendation #1:

Prioritize investment in AI Safety research.

This is an area that fits the natural government function of funding under-supplied public goods. AI safety, the ability to orient AI systems towards the safe achievement of goals, is in the interests of all parties. However,

companies and academics have traditionally avoided major investment in this area, perhaps

expecting it to be solved by others in the future. Only a few organizations have seriously committed to investing in long-term AI safety research, despite growing recognition of its importance by prominent researchers. Such investments could also serve to more closely align the interests of the US government with those of AI researchers.

AI Safety is a research area dedicated to improving the predictability and robustness of increasingly autonomous systems. Though commercial actors have incentives to ensure that specific products are safe when deployed into the market, coordination between multiple private actors is a challenge without natural standardization of AI Safety techniques, research agendas, and applications. AI Safety is an area where targeted government investment can help define a cross-cutting suite of technologies that can be accessed by a variety of other actors - private sector companies, research organizations such as OpenAI, academics, and governments themselves. Government is well positioned to - via the various measurement and standards initiatives proposed in the existing strategic plan - synthesize the available data about

AI Safety research and make investments that increase the standardization of this technology.

AI Safety is also an area of concern for AI researchers.

Organizations such as Alphabet's DeepMind and OpenAI have invested significant sums in forming their own "AI Safety teams" to conduct research into this area, and in recent years a number of organizations - including the Future of Life Institute, the Machine

Intelligence Research Institute (MIRI), and the Future of Humanity Institute - have received significant amounts of philanthropic donations to further their own work in AI safety; this highlights the pent-up demand in the AI ecosystem for more investment into AI Safety research. Our belief is that targeted funding of AI Safety research by the US government would increase its ability to have productive conversations with AI researchers who are keen to work on matters relevant to national security and national competitiveness,

but are currently wary of engaging with institutions that don't appear to prioritize safety.

Recommendation #2:

Strengthen the US AI ecosystem by investing in AI talent.

There are an immediate high value actions that policymakers can take to help ensure we have a healthy AI R&D workforce: support sensible, skill-based immigration initiatives to ensure that the US AI ecosystem has a sufficiently diverse and skilled talent pool that it can reliably generate new research ideas, which can serve to further support US industry and academic leadership in AI.

Across the board the AI industry is full of open positions, and would love to be hiring faster. High levels of access to the global AI talent pool would enable market mechanisms to meaningfully reduce the gap between how quickly the industry can hire and would like to hire. If the uncertainty and wait times for visas to work in this strategic industry could be reduced, it'd be much easier for AI organizations to rapidly grow and create value. In economic terms, talent and funding are the two primary drivers of the AI industry.

Attracting top talent is extremely difficult and competitive globally.

Talent

tends to follow other talent and concentrate, so policy actions taken in the next couple years here have the potential to shape the long term global landscape of this industry.

In addition to addressing the hiring needs of US companies, progress on this front would have valuable side effects for facilitating international cooperation on the many challenges surrounding the development of beneficial AI technology. Allowing talent to freely flow into US institutions will make it easier to create an international social fabric around AI research, and put America in a position to take the lead on coordinating norms and best practices for AI safety and ethical use of AI.

Recommendation #3:

Augment existing measurement and analysis initiatives with a specific “dual use” focus.

For America to lead in international coordination on AI development - and thereby be able to shape the norms and laws which define the technology - it must tackle the challenge of AI technologies being exploited by bad actors.

Applications of AI technology, like any technology, are

prone to misuse. However, many of the applications of AI technology are of a scale and power that is of a different kind to typical technologies - AI systems are typically embodied in software, can be deployed onto generic infrastructure (CPUs and GPUs and commodity storage), and are frequently available via “open source” (free) code. This lowers the barriers to deployment of the technology and broadens the range of actors that can feasibly deploy it.

We also know that AI technologies are becoming increasingly more powerful, as advancement in AI is correlated with a growth in the amount of compute being used to train certain transformative applications.

AI’s capabilities and potential applications might develop at an increasingly rapid rate. A recent analysis by OpenAI shows that the amount of compute used to develop significant AI capabilities has grown by approximately 300,000X in the past six years (Moore’s Law would yield a 12X increase over this same period)[4].

We believe that it is in the government’s interest to analyze, measure, and forecast such technical progress as this will make it easier for government to anticipate areas to invest in AI to benefit society, and can serve as an early warning system for the potential downsides of AI that can be addressed through discussion, investment, or potentially regulatory actions.

[1] For more information, please refer to:

www.openai.com

[2]

<https://blog.openai.com/openai-charter/>

[3]

<https://www.cnas.org/publications/reports/technology-roulette>

[4]

<https://blog.openai.com/ai-and-compute/>
