October 26, 2018

Faisal D'Souza
National Coordination Office
Networking and Information Technology Research and Development
215 Eisenhower Avenue
Alexandria, VA 22314

Dear Mr. D'Souza,

On behalf of the Center for Data Innovation (datainnovation.org), we are pleased to submit comments in response to the Networking and Information Technology Research and Development's (NITRD) request for comments on updating the 2016 National Artificial Intelligence Research and Development Strategic Plan.[1]

The Center for Data Innovation is the leading think tank studying the intersection of data, technology, and public policy. With staff in Washington, D.C., and Brussels, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as important data-related technology trends. The Center is a non-profit, non-partisan research institute affiliated with the Information Technology and Innovation Foundation.

The United States is the global leader in developing and using artificial intelligence (AI), but it may not be for long. Other countries, such as China, France, and the United Kingdom have developed significant initiatives to challenge U.S. leadership in AI, while the U.S. government has not yet developed a comparable national AI strategy to address the challenges holding back greater adoption and use of AI in the United States. NITRD's National AI R&D Strategic Plan is a valuable opportunity to help the United States maintain its competitive edge in the absence of a broader national AI strategy.

[1] "Request for Information on Update to the 2016 National Artificial Intelligence Research and Development Strategic Plan," Federal Register, September 26, 2018, https://www.federalregister.gov/documents/2018/09/26/2018-20914/request-for-information-on-update-to-the-2016-national-artificial-intelligence-research-and.

STRATEGY 1: MAKE LONG-TERM INVESTMENTS IN AI RESEARCH

The 2016 National AI R&D Strategic Plan rightly recognized the need for R&D investments in areas of AI with potential long-term payoffs, as "some important areas of [AI] research are unlikely to receive sufficient investment by industry, as they are subject to the typical underinvestment problem surrounding public goods."[2] Though the AI R&D Strategic Plan is not a budget proposal, it should nonetheless stress the need for Congress to address the federal government's substantial underinvestment in basic and applied research that could generate large payoffs for AI.

There are some commendable AI R&D funding commitments. The Defense Advanced Research Projects Agency (DARPA) has announced that it plans to spend $2 billion on a multi-year "AI Next" campaign to advance the state-of-the-art in AI, such as by researching how to develop AI systems capable of common-sense reasoning.[3]

However, outside of defense applications, there is not nearly enough AI R&D funding. The National Science Foundation (NSF) funds $122 million in core AI research annually, however, this amount of funding is insufficient to pay for all of the high-priority R&D. For example, in 2017, NSF received $174 million in proposals for AI research that it deemed either competitive or highly competitive, but that it did not have the budget to support, indicating significant capacity for increased AI R&D.[4]

While the private sector supports some AI R&D, the federal government should play a larger role. The federal government's commitment to R&D, as measured by R&D intensity—the ratio of R&D to GDP—has fallen from a high of 1.86 percent in 1964 to just 0.66 percent in 2015.[5] While private sector R&D in the United States has grown steadily, surpassing federal R&D intensity in 1980 and eventually reaching 1.84 percent by 2016, public and private R&D are not interchangeable.[6] The federal government is the predominant funder of basic research, which is foundational to all other

---

[2] Networking and Information Technology Research and Development Subcommittee (NITRD), *The National Artificial Intelligence Research and Development Strategic Plan* (Washington, DC: NITRD, October 2016), https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf.
[3] "AI Next Campaign," U.S. Defense Advanced Research Projects Agency, Accessed October 24, 2018, https://www.darpa.mil/work-with-us/ai-next-campaign.
[4] *Game Changers: Artificial intelligence Part II, Artificial Intelligence and the Federal Government, Before the House Oversight Subcommittee on Information Technology*, 115th Cong. (2018) (statement of James Kurose, Assistant Director of Computer Science and Information Science and Engineering, National Science Foundation).
[5] "U.S. gross domestic product, R&D, and ratio of R&D to gross domestic product (and components): 1953–2016," National Science Foundation, Accessed October 25, 2018, https://www.nsf.gov/statistics/2018/nsf18309/pdf/np16-dst-tab001.pdf.
[6] "Federal R&D Budget Dashboard," American Association for the Advancement of Science, Accessed October 24, 2018, https://www.aaas.org/programs/r-d-budget-and-policy/federal-rd-budget-dashboard.

R&D. However federal R&D intensity in basic research fell to 0.21 percent in 2015, its lowest point since 1980.[7] Meanwhile, though private R&D intensity in applied research, which involves using known basic scientific concepts to solve a particular problem, surpassed federal applied R&D intensity in 1979, both public and private R&D investment in applied research has seen a downward trend over the last several decades.[8] Federal funding should support both basic and applied R&D. Notably, private R&D has overwhelmingly focused on developmental research, which refers to developing available scientific knowledge into a commercial product or process.

The United States should also look to other countries that have developed national AI strategies to direct its AI investments. In particular, China and France have developed proposals that emphasize interdisciplinary AI research, and the United States should make that more of a priority. China's 2017 New Generation Artificial Intelligence Development Plan recommends promoting exploratory research into the convergence of AI with psychology, economics, sociology, and other core disciplines, with a particular focus on supporting research into areas where there is no consensus and opportunities to make wholly original discoveries.[9] Similarly, France's AI strategy proposes the creation of four to six interdisciplinary AI research institutions, called 3IA, with the objectives of "(re)shaping attractive and prestigious research environments that are capable of significant breakthroughs at international level and are grouped under a single, high profile and renowned label; dispensing high-level scientific training in AI, for the researchers, engineers and entrepreneurs of tomorrow; [and] enabling smoother interfaces between disciplines and between academic research and industry, expediting the transformation of ideas into proofs of concepts (POC), scientific applications and groundbreaking technology and intellectual property, capable of forging the fabric of startups and SMEs on which the industry of tomorrow will depend." France's strategy also recommends providing three key types of support for 3IA institutions: "1) access to virtually unlimited computing means; 2) administrative procedures that have been streamlined as far as possible; 3) assistance with living conditions, not least for foreign researchers."[10] The United Kingdom also espouses an interdisciplinary approach, detailing in its AI Sector Deal to invest up to £20 million

[7] Ibid.
[8] Ibid.
[9] Graham Webster, et al., "Full Translation: China's 'New Generation Artificial Intelligence Development Plan' (2017)," New America, August 1, 2017, https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/.
[10] Cédric Villani, *For a Meaningful Artificial Intelligence* (France: Mission assigned by the Prime Minister Édouard Philippe, March 28, 2018), https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf.

($25.8 million) to spur research into how AI can benefit services sectors such as law and insurance.[11]

STRATEGY 2: DEVELOP EFFECTIVE METHODS FOR HUMAN-AI COLLABORATION

Developing effective methods of human-AI collaboration is an important and worthwhile goal for the National AI R&D Strategy. However, the strategy should recognize that it is similarly important to develop effective methods for AI-AI (i.e. machine-to-machine) collaboration as well. As more technologies and services integrate AI, ensuring different AI systems can easily and seamlessly interact and work with one another will enable a wide variety of useful consumer, business, and government applications.

STRATEGY 3: UNDERSTAND AND ADDRESS THE ETHICAL, LEGAL, AND SOCIETAL IMPLICATIONS OF AI

The 2016 National AI R&D Strategy included "improving fairness, transparency, and accountability-by-design" as a goal for AI R&D investments, citing the need to ensure AI systems are not susceptible to bias or used to discriminate unfairly.[12] While developing methods to detect and correct for bias in AI systems is worthwhile, transparency is only a useful method for doing so in limited circumstances. Though many have called for requirements that algorithms be made transparent so that regulators, journalists, and others could scrutinize their source code and spot evidence of harmful behavior, "pulling back the curtain" on algorithms would likely be ineffective for many uses of AI. While examining code can provide meaningful information about how some algorithmic systems make decisions, for many advanced AI systems that rely on thousands of layers of simulated neurons to interpret data, even their developers cannot explain their decision-making. For example, researchers at Mount Sinai Hospital in New York developed an AI system called Deep Patient that can predict whether a patient is contracting any of a wide variety of diseases.[13] The researchers trained Deep Patient on the health data from 700,000 patients, including hundreds of variables, which allow it to predict disease without explicitly having to be taught how. The system is substantially better than other disease-prediction methods, yet its own developers do not know how its decision-making process works.[14] The United Kingdom's Government Office for Science cautions against pursuing

---

[11] *AI Sector Deal* (United Kingdom: U.K. Department for Business, Energy, and Industrial Strategy, April 26, 2018), https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal.
[12] Networking and Information Technology Research and Development Subcommittee (NITRD), *The National Artificial Intelligence Research and Development Strategic Plan* (Washington, DC: NITRD, October 2016), https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf.
[13] Will Knight, "The Dark Secret at the Heart of AI," MIT Technology Review, April 11, 2017, https://www.technologyreview.com/s/604087/the-darksecret-at-the-heart-of-ai/.
[14] Ibid.

transparency as a means for protecting against the potential risks of algorithms, stating "Most fundamentally, transparency may not provide the proof sought: Simply sharing static code provides no assurance it was actually used in a particular decision, or that it behaves in the wild in the way its programmers expect on a given dataset."[15] In a nutshell, transparency guarantees neither accurate nor unbiased results.

More importantly however, by emphasizing these traits be considered "by design," this recommendation places the burden of ensuring AI does not cause harm on an AI system's developers, rather than the party responsible for deploying an AI system, or its "operator." This is a shortsighted approach because developers have little control over how their algorithms are used, while operators make the most important decisions about how their algorithms impact society.[16] For example, an AI system could be fair and accountable in certain conditions, but when exposed to new training data or used in other contexts—factors outside of a developer's control—it could cause significant harm.

Instead, the AI R&D Strategic Plan should encourage research into methods for achieving algorithmic accountability. Algorithmic accountability is the principle that an algorithmic system should employ a variety of controls to ensure operators can: verify it works in accordance with the operator's intentions; and identify and rectify harmful outcomes. Algorithmic accountability promotes desirable outcomes, protects against harmful ones, and ensures algorithmic decisions are subject to the same requirements as human decisions.[17] This approach is technology neutral, granting operators flexibility to employ a variety of different technical and procedural mechanisms to achieve algorithmic accountability. Federal R&D efforts should focus on developing and improving these various mechanisms, which include, but are not limited to, explainability, confidence measures, impact assessments, and error analysis.[18]

The 2016 plan recognized the value of improving the explainability of AI systems, which currently poses a challenge for AI developers and operators as there can be as-of-yet inescapable tradeoffs between the explainability and accuracy of advanced AI systems.[19] DARPA has taken encouraging

---

[15] *Artificial Intelligence: Opportunities and Implications for the Future of Decision Making* (London: Government Office for Science, February 12, 2016), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf.
[16] Joshua New and Daniel Castro, "How Policymakers Can Foster Algorithmic Accountability" (Center for Data Innovation, May 21, 2018), http://www2.datainnovation.org/2018-algorithmic-accountability.pdf.
[17] Ibid.
[18] Ibid.
[19] Ibid.

steps to overcoming this challenge, committing $75 million in 2017 for its "XAI" program to research how algorithmic explainability could be achieved. Federal R&D investments should expand on this work to make it easier for operators to adhere to the principle of algorithmic accountability.

STRATEGY 4: ENSURE THE SAFETY AND SECURITY OF AI SYSTEMS

The 2016 plan emphasizes the importance of ensuring AI in critical systems is secure from cyber-attacks, however the AI R&D Strategy should encourage research into both the cyber risks and opportunities of AI systems more broadly. AI provides attackers new cybersecurity vulnerabilities to exploit and new methods to automate cyberattacks. Conversely, AI is a powerful tool for automating cyber defenses, discovering unknown vulnerabilities, and augmenting the shortage of human workers available to address cybersecurity challenges. However, NITRD's 2016 Federal Cybersecurity R&D Strategic Plan only briefly mentions the role of automated systems and machine learning in cybersecurity.[20] NITRD should ensure that its AI R&D Strategic Plan and its cybersecurity efforts both complement each other and emphasize the importance of advancing research into the risks and opportunities posed by AI for cybersecurity.[21]

STRATEGY 5: DEVELOP SHARED PUBLIC DATASETS AND ENVIRONMENTS FOR AI TRAINING AND TESTING

NITRD should continue to explore how it can best make shared data resources available for developing AI systems. In particular, NITRD should fund research into developing innovative public-private data sharing models. For example, the United Kingdom is developing data trusts—a model where public and private parties have defined rights and responsibilities regarding shared data—as a means of facilitating access to data that businesses, government agencies, or researchers would otherwise opt to not share due to concerns about its sensitive or proprietary nature. In addition, NITRD should explore how to ensure these types of data sharing frameworks can foster not only domestic data sharing, but international data collaborations as well. Additionally, data produced by federally funded research should be made available as open data by default. However, in cases where the data is sensitive, it should be made available through a data trust or similar mechanism.

---

[20] Networking and Information Technology Research and Development Subcommittee (NITRD), *Federal Cybersecurity Research and Development Strategic Plan* (Washington, DC: NITRD, February 2016), https://www.nitrd.gov/cybersecurity/publications/2016_Federal_Cybersecurity_Research_and_Development_Strategic_Plan.pdf.

[21] Nick Wallace, "Integrating Europe's AI and Cybersecurity Strategies," Center for Data Innovation, September 26, 2018, https://www.datainnovation.org/2018/09/integrating-europes-ai-and-cybersecurity-strategies/.

Developing public datasets is one way that the federal government can help address concerns about bias in AI. For example, it should develop shared datasets that can serve as an unbiased resource for organizations developing facial recognition technology. Facial recognition technology can be incredibly useful, however historically, the training data available to developers overwhelmingly consists of white, male faces, causing many facial recognition systems to underperform for minorities and women.[22] Though the private sector has an incentive to make facial recognition algorithms as accurate and reliable as possible, developing a representative dataset of hundreds of thousands of faces requires considerable resources, and even if a company were to do that, it has little incentive to share this data with potential competitors. Recognizing this challenge, IBM announced in June 2018 that it would publish the world's largest annotated dataset of faces specifically for the purposes of studying bias in facial analysis.[23] This is encouraging, however not only is this insufficient to solve the problem of biased facial recognition systems but overcoming this challenge should not just be the responsibility of the private sector. NITRD should encourage agencies to develop public datasets of faces that reflect the diversity of the United States. In addition, it should launch a call for proposals to develop other training and testing datasets that could reduce bias through better public data.

STRATEGY 6: MEASURE AND EVALUATE AI TECHNOLOGIES THROUGH STANDARDS AND BENCHMARKS

While the private sector has made important strides in developing AI benchmarking tools—such as MLPerf, which provides an open suite of benchmarking tools for machine learning software frameworks, hardware accelerators, and cloud platforms—many of these benchmarks are still evolving.[24] The federal government can accelerate the development and adoption of effective and reliable benchmarks for AI systems. In addition, the federal government has an important role to play to ensure that the private sector continues to develop voluntary, consensus-based standards for new AI technology both domestically and internationally.

STRATEGY 7: BETTER UNDERSTAND THE NATIONAL AI R&D WORKFORCE NEEDS

The 2016 plan rightly acknowledges the need to better understand the current and future national workforce needs for AI, and agencies should continue to research these questions. However, certain workforce obstacles have already revealed themselves and agencies should address these

[22] Steve Lohr, "Facial Recognition Is Accurate, if You're a White Guy," *The New York Times*, February 9, 2018, https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html.
[23] IBM, "IBM to Release World's Largest Annotation Dataset for Studying Bias in Facial Analysis," new release, June 27, 2018, https://www.ibm.com/blogs/research/2018/06/ai-facial-analytics/.
[24] "MLPerf," MLPerf, Accessed October 24, 2018, https://mlperf.org/.

challenges proactively. For example, the extreme demand for skilled AI workers is creating a negative feedback loop for the AI education pipeline. Technology firms are willing to pay a premium for workers with highly sought-after AI skills, attracting leading academics away from universities, which limits the pool of AI experts available to teach these skills. For example, as of late 2017, six of twenty AI professors at the University of Washington are on full or partial leave to work for the private sector, and four of the most renowned AI researchers in academia have taken leave or fully left their professorships at Stanford University.[25] Data from the National Science Foundation suggests this trend is endemic, with 58 percent of new computer science PhDs taking jobs in the private sector in 2015, rather than staying in academia, up from 38 percent from 2005.[26] This is due in no small part to the fact that in 2014, the median annual salary for postdocs in computer science was $55,000 at universities and $110,000 in the private sector.[27] To help alleviate the brain drain from universities, the federal government should provide competitive early-career monetary awards for AI researchers that are conditional on remaining in academia for a fixed period of time. These awards would incentivize more AI researchers to stay in academia and help U.S. universities meet the demand for AI skills.

CONCLUSION

The ability of the United States to remain globally competitive in AI will depend in no small part on public R&D activities focused on accelerating the development and deployment of the technology. NITRD's 2016 AI R&D Strategic Plan was a welcome initiative to coordinate federal R&D activities in AI and focus them where they would be most effective, and it is encouraging to see NITRD build on this effort.

Sincerely,

Daniel Castro
Director
Center for Data Innovation
dcastro@datainnovation.org

[25] Cade Metz, "Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent," *The New York Times*, October 22, 2017, https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html.
[26] Daniela Hernandez and Rachael King, "University's Ai Talent poached by Tech Giants," *The Wall Street Journal*, November 24, 2016, https://www.wsj.com/articles/universities-ai-talent-poached-by-tech-giants-1479999601.
[27] Ibid.

Joshua New
Senior Policy Analyst
Center for Data Innovation
jnew@datainnovation.org