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.

National Science Foundation Networking and Information Technology Research and Development (NITRD) National Coordination Office (NCO) Request for Information on Update to the 2016 National Artificial Intelligence Research and Development Strategic Plan

Comments of Microsoft Corporation
October 26, 2018

Microsoft appreciates the opportunity to provide input to the National Science Foundation Networking and Information Technology Research and Development (NITRD) National Coordination Office (NCO) Request for Information on Update to the 2016 National Artificial Intelligence Research and Development (R&D) Strategic Plan. We applaud the timeliness of NITRD NCO's request: continued innovation in artificial intelligence (AI) technologies is an essential element in any national strategy to maintain U.S. competitiveness and global leadership in this area.

Microsoft's vision for AI is to amplify human ingenuity with intelligent technologies. We are working toward a "human-centered" vision of AI—where human intelligence is augmented through advances in areas such as computer vision, speech recognition, natural language understanding and machine learning. We view AI as a range of general-purpose technologies that can enhance any decision-making process with additional insights and intelligence, leading to breakthroughs in innovative applications, services and business models, as well as societal challenges.

Microsoft has made deep and sustained investments in research and development in AI in recent decades, and is a major contributor to its advancement as well as its application more broadly in diverse areas including education, healthcare, agriculture, environment and sustainability, and accessibility. For example, in 2017 Microsoft's researchers developed a speech recognition system that understood spoken words as accurately as a team of professional transcribers, with an error rate of just 5.1 percent using the standard Switchboard dataset; researchers are also partnering with epidemiologists to develop innovative techniques such as using mosquitoes as sensors to detect pathogens in the environment and prevent disease outbreaks. We strongly believe in the potential of AI to empower and create new opportunities for every person and every organization, leading to advances in nearly every field of human endeavor and progress in many existing societal challenges. When made accessible to everyone, AI will transform industries, increase productivity, and help solve society's biggest challenges. To fuel AI innovation for these goals, Microsoft launched its AI for Good initiative in 2017 to encourage development of AI innovations that can help solve humanitarian issues and create a more sustainable and accessible world. We are investing $115 million in a five-year commitment to three programs: AI for Earth, AI for Accessibility, and AI for Humanitarian Action.

Like other technological advances, AI will raise a number of new societal issues and exacerbate some existing ones. It also has the potential to be misused in ways that can cause harm. These are valid concerns, but if handled reactively and without thoughtful consideration, they may lead to policy and regulatory approaches that could adversely impact continued innovation in AI. Often overlooked in such policy dialogues is that AI innovation should be an essential part of any solution.

Realizing the promise of AI is a shared responsibility that requires government, business, academia, civil society and other interested stakeholders to come together to shape its development, with a firm foundation in long term R&D of the technology, guided by principles and policy frameworks that can foster trust and broad adoption.

Since 2016, recognition of AI's transformative impact on national economy and competitiveness is evident from multiple national strategies that have been published, and the AI declarations and joint statements that have been made in bilateral, plurilateral and multilateral agreements. Many of these are concerned with enabling long-term AI innovation as an anchor for sustained economic growth, creating AI research centers, and developing the next generation of AI breakthroughs. For example, in 2017 Canada announced a CAD$125 million Pan-Canadian AI strategy focused on AI research and talents, complemented by a CAD$950 million funding for innovation superclusters around the country. In March 2018 France announced a €1.5 billion AI strategy to establish its global leadership in AI research, training, and industry, including creating a network of research institutes across the country. In April 2018 the EU Commission committed to increase its investment in AI to €1.5 billion by the end of 2020, with incentives to mobilize an additional €3 billion in combined public and private-sector funding; the Commission and Council are working to develop a coordinated action plan on AI by early 2019. In May 2018 Korea released a US$2 billion national AI R&D plan.

Microsoft firmly believes that **a sustained focus on AI R&D, in both the public and private sector, is fundamental to preserving U.S. competitiveness and global leadership, and to shaping development of AI policy and regulatory frameworks in ways that can maximize its potential benefits**. Governments around the world are expressing growing interest in AI, and especially in establishing regulatory and ethical frameworks related to it. With technologies often perceived as exacerbating global inequalities and societal instability, governments are increasingly seeking to regulate U.S. businesses and innovations. In this challenging environment, it is essential that the U.S. government, in a coordinated manner, continue its strong and proactive engagements in international organizations and other bilateral and plurilateral negotiations to help support U.S. business and enable continued growth and competitiveness. The U.S., with a policy environment that promotes market-based innovation, inclusive economic growth, and a preference towards ex-post regulation and enforcement, has led the world in deployment of innovative digital technologies and experienced first-hand their impact on national GDP.

We urge the U.S. government to take steps that would strengthen its active participation in these global dialogues, including a **stronger inter-agency coordination process that is focused on opportunities for AI to enhance U.S. competitiveness and economic prosperity, and budgetary allocations within respective federal agencies that would correspond with stated AI objectives and R&D priorities**.

Within this context, we would like to reinforce some existing strategies from the current AI R&D plan, as well as raise new areas for consideration. We group these broadly into the following dimensions:

1) **Technology development**: short- and long-term investment in strategic AI technologies, wherever applicable, should emphasize fostering trust and broad adoption, including development of practices that can identify and mitigate issues that would cause lack of trust, including detection of bias and adversarial AI.
2) **Enabling broad AI innovation and adoption across all sectors and by organizations of all sizes**:
    a. Shared public datasets and testing/verification environments that will enable innovation of AI technologies and applications, especially by entrepreneurs and small and medium enterprises (SMEs);
    b. Global standards that are voluntary and market-driven, which can help to inform common practices and development of holistic policy frameworks, and form the bases for market efficiencies;
    c. Best practices in adoption of AI across sectors, including use of proprietary solutions, open-source software and open standards-based solutions.

3) **Sociological research**: although the overall societal impact of AI will be positive, gaining better insights into this transformation and potential downsides will require non-technical research, including sociological and longitudinal studies, that can help more effectively and innovatively identify and address issues such as AI workforce and reskilling needs.
4) **Research challenges, centers and public-private partnerships**: these are effective and efficient mechanisms to encourage technology advances and commercialization in select strategic areas.
5) **AI R&D and policy making**: use of AI for more effective and evidence-based policy making.

We believe that enhancing the National R&D Plan with the above considerations will increase U.S. competitiveness and economic prosperity. They are discussed in more detail below.

1. Technology development (Strategies 1-4)

For AI to be deployed on a global scale, its development needs to be shaped to be trustworthy, i.e., AI systems must function in such a way as to inspire trust in their users. This can only be realized if relevant stakeholders from business, government, civil society and the research community work together on shared principles and ethical frameworks. In January 2018 Microsoft released the book "The Future Computed: Artificial Intelligence and its role in society" to contribute our perspectives to this global dialogue and encourage a sense of shared responsibility that we all have in shaping this development. We believe that commitment to the principles of fairness, reliability and safety, privacy and security, and inclusiveness – supported by an underlying foundation of transparency and accountability – provide a solid framework for building trust and should guide the development of AI.

Based on this, we propose that "Trustworthiness" be considered part of the "cross-cutting R&D foundations" in Figure 4, and is a superset of "Safety and Security." We appreciate the focus on both short- and long-term research priorities described in Strategies 1-4, and encourage all federally-funded research to be guided under the context of fostering trust and broad adoption of AI technologies, and technologies that are human-centered. For example, researchers "*pursuing research on general-purpose artificial intelligence"* or "*fostering research on human-like AI"* should be guided by these principles. In addition, the research priorities should also include development of tools and practices that can detect and address potential risks that may arise, e.g., detect and mitigate bias that may be present in training data, and address and mitigate adversarial AI.

Bias and fairness
AI systems should strive to treat everyone equally and not result in disparate impact towards any groups of people. However, because AI models are driven by data, the cultural biases and behavior of society at large, as well as the biases of the individuals or organizations that collect, curate and deploy training data sets, can be learned, reinforced, and in some cases amplified by the resulting models. When the resulting recommendations are applied without full understanding of these potential system flaws and limitations, the outcome can lead to disparate impact.

Moreover, once such biases are detected, determining appropriate actions and responses, if any, that need to be taken is a complex technical and policy issue that will most likely need to be context-dependent, and this introduces other challenges. What is essential is to raise the awareness of fairness issues that can arise during the development and deployment of AI systems, so that they can be appropriately identified and addressed. *As these are sociotechnical issues, research in this area needs to be holistic in developing both*

1) *analytical approaches to detect and mitigate potential unfairness, such as methods that (a) systematically assess the data used to train AI systems for appropriate representativeness and document information about its origins and characteristics; and (b) establish comprehensive fairness quality criteria and confirm that a system meets such criteria before release; and*
2) *sociological approaches to increase diversity in design, development, deployment and operation of AI systems, and to collect corresponding evidence on their effect on bias in system recommendations and decision-making.*

Transparency and intelligibility

When AI is used to help make decisions that impact people's lives, it is important that they can understand the logic behind system recommendations. Behind this request is typically the need to verify qualities such as fairness, safety and reliability – and intelligibility of system behavior is a means to that end. Transparency and intelligibility are also essential for enabling human accountability and redress, which are fundamental to fostering trust of AI and enabling its broader adoption.

Calls for satisfactory explanation of the operation of machine-learning models, training algorithms and automated systems are growing louder, fueled in large part by increasing anxiety over the ubiquity of "black-box" technologies. These demands have manifested themselves in recent regulatory initiatives. For example, Europe's General Data Protection Regulation (GDPR) mandates the disclosure of "meaningful information about the logic of the decision-making" in certain automated systems. And, more recently, the European Commission has proposed that online platforms such as social media and search services make information available about the functioning of their models and algorithms in order to promote competition and consumer protection.

Enabling human understanding of AI is, and must be, an area of intense, cutting-edge research.
Promising tools and practices for understanding AI systems are emerging, but more research is needed to advance this work and to determine which approaches can help people actually obtain meaningful and contextual explanations of system behavior. Indeed, early research involving human-subject experiments suggests that the landscape is not as straightforward as originally expected, as some system-design choices commonly thought to influence intelligibility have been shown not to have an appreciable effect on human outcomes. Researchers have even shown that, at least in some contexts, literal exposure of system internals can prevent people from noticing when a model makes a mistake due to an overload of information. Accordingly, as more methods for enabling human understanding of AI systems are developed and refined, it will be even more important to consider the full context in which a system is used and the reasons for needing intelligibility before selecting a particular method, as well as the utility of that method to end-users.

Safety and cybersecurity

AI-powered security innovations can enable greater levels of defense and empower users to better manage cybersecurity risks in their AI deployments. However, as AI becomes more widely available, concerns about malicious use of the technology are growing. *Investments in AI security research need to advance the use of AI for cyber defense in public- and private-sector networks, as well as the state of secure software-development practices to address new threats to AI.*

There are increasingly urgent questions that require research on how to forecast, prevent and mitigate the harmful effects of malicious uses of AI. Additional considerations on AI systems research related to cybersecurity include:

- All types of machine learning improve with the quality of the training data. Such data needs to have rigorous quality standards applied and be sanitized against changes, whether benign or malicious. New detection methods must be brought to bear against contaminated and manipulated training data so that it is not ingested blindly, and systems must be able to discern the difference between these two types of anomalies.
- The interpretability and understanding of AI-based decision-making grows in importance as the sophistication of deep-learning systems increases. A system that can hide its self-modification or is unable to provide visibility into its decision-making processes in a verifiable fashion will be more susceptible to manipulation. New research that combines system transparency and intelligibility with cybersecurity detection approaches are needed.
- In deployment, even a simple AI-driven system can require complex interactions between multiple algorithms created by different organizations. Research on verification of the interaction of these adaptive systems, some of which may be executing automatically based on pre-determined conditions, along with opportunity for human oversight or intervention, will greatly help to improve the safety and security of such systems.
- Human-machine interaction is an important area of research in ensuring safety. However, AI should exercise discretion in its handling of human interactions, whether stemming from over-sharing information, a human-initiated intervention to correct bias in a system, or when acting as an agent on behalf of a user. New research is needed to further the concepts of AI resilience and discretion in this area.
- The threat landscape related to the malicious use of AI will evolve both rapidly and constantly. Innovative and continuing approaches to cybersecurity development and training will be essential.

2. Enabling broad AI innovation and adoption across all sectors and by organizations of all sizes (Strategies 5-6)

Enhancing competitiveness and economic prosperity requires that an appropriate infrastructure be in place to facilitate innovation and broad adoption, for both research and commercialization purposes, especially for the entrepreneurs and SMEs that form the core of healthy and sustainable digital ecosystems. This infrastructure must incorporate the following elements:

a. Shared public datasets and testing/verification environments that will enable continued innovation of AI technologies and applications;
b. Global standards that are voluntary and market-driven, which can help inform common practices and development of holistic policy frameworks, and form the bases for market efficiencies;
c. Best practices in adoption of AI across sectors, including use of proprietary solutions, open-source software and open standards-based solutions.

As AI is data-driven, a national data strategy must be a part of any AI strategic plan. The strategy should accelerate existing open data initiatives, and simplify access to and encourage use of publicly owned and funded datasets, including by U.S. government agencies. This will require additional R&D investments in the development of tools and resources to make it easier to identify, use and manipulate relevant datasets, including work to facilitate interoperability of data sets, verify data provenance, and comply with appropriate policy on use of the data sets. To help reduce the risk of privacy intrusions, governments should support and promote the development of de-identification techniques that enable systems to access personal data for developing AI algorithms without knowing the identities of

individuals. Governments can also invest in and promote methods and processes for linking and combining related datasets from public and private organizations while preserving confidentiality, privacy and security as circumstances require.

Given the current state of AI technology and policy development, U.S. government leadership in developing global standards through voluntary consensus-based standards organizations that have the appropriate expertise, both technical and business, to address foundational AI concepts will be helpful in informing ongoing concerns about AI. For example, the AI project under ISO/IEC JTC 1/SC 42 that is developing AI terminology, reference models, and risk awareness and management frameworks can provide a common structure to guide ongoing discussion, including informing ways in which risks can be identified and mitigated. The lack of such work may actually hinder broader adoption of AI and lead to regulatory approaches that may impede further innovation.

With the intense interest in AI and ethics, it should be noted that although specifying a standard process for the design, development and operation is a good approach for increasing awareness of these issues and encouraging better outcomes, it is essential that any such standards definition does not dictate that systems must implement a given set of values. Furthermore, such standards should not be aimed only at engineering/computing professionals – the design, development, operation and deployment of trustworthy AI is a shared responsibility that must be driven by the leadership of the organization.

Interoperability standards can come at a later stage, when AI technology is more mature, to avoid adversely impacting innovation, and open-source software (OSS) should be considered as a viable and more flexible alternative to providing interoperability.

In general, AI standards should be performance-based and establish desired outcomes, but not be prescriptive on how the outcome be achieved. It should also be noted that the U.S. government should not develop its own standards according to OMB Circular A.119, and that internationally harmonized and voluntary standards are the preference.

An additional consideration for the R&D plan to promote earlier adoption of AI is to encourage sharing of best practices in adoption of AI across sectors, especially across U.S. federal government agencies, and further development of software libraries and toolkits, including use of proprietary solutions, open-source software and open standards-based solutions. Domain-specific benchmarks may be helpful here. Federal agencies should prioritize the use of open data for training and open data standards in AI, and explore ways to improve the capacity of key agencies with the use of AI.

3. Sociological research

Although the overall societal impact of AI will be positive, much has been written about its potential downsides, e.g., exacerbating existing bias, promoting income inequality and social behavior in gaming AI systems. These considerations must be an integral part of any AI R&D strategic plan, as technological advancements need to be informed by sociological and behavioral insights. *As technology becomes part of the foundation of our everyday lives and interactions, its effects must be understood, and a sustained multidisciplinary research program is required.* An example of such research that was raised above is studies on the effect of integrating more-diverse perspectives in the design, development and deployment of AI systems. Other research, especially longitudinal studies, can further inform issues in workforce and skills development including barriers to online learning, effective learning habits of different age groups, socio-cultural barriers to career transitions, and examples of career pathways.

4. Research challenges, centers and public-private partnerships

Research challenges, such as the DARPA Grand Challenge, have a history of promoting technology development, innovation and commercialization in strategic areas. The existence of dedicated research centers in different strategic areas can serve the same purpose. *With U.S. competitiveness in AI at stake, such challenges and centers should be considered as parts of federally funded research programs.*

Public-private partnerships in AI research, whether as part of a research challenge or as part of a specific research initiative, can help focus the research on concrete (but not necessarily short-term) problems, and lead to more efficient use of federal research funding to address competitiveness and economic issues. Furthermore, such partnerships can often guide the development of outcomes that are more practical, implementable and scalable – all important considerations in technology commercialization. *We recommend that the NCO conduct periodic dialogues between public and private sector experts to share current view of the state of AI R&D and emerging challenges, especially those that can enhance U.S. competitiveness.*

5. AI R&D and policy making

The R&D plan needs to consider the role that AI technology can play in enabling more effective and evidence-based policy making.

AI technology development as an integral part of AI policy development
As AI is still at a nascent stage of development, open dialogues between government, business, civil society and academic researchers are essential to shape continued development of the technology to be trustworthy and realize its potential benefits. Working together we can identify and prioritize issues of societal importance as AI evolves, enable sharing of best practices and motivate further research and development of solutions as new issues emerge. This would require that AI technology development, which includes all of the above elements, be considered an integral part of any AI policy development process.

As an example, consider that in addressing the transparency of AI systems, researchers at Microsoft and elsewhere continue to explore state-of-the-art methods for explaining AI systems. A number of promising technical methods that involve examining both individual system components and full AI system behavior have begun to emerge. This includes concepts like "datasheets for datasets," where every training dataset is accompanied by a description that includes its motivation, composition, how it was collected and processed, and any limitations in the data collected that can result in unintended outcomes such as know biases or privacy considerations. *Keeping policy stakeholders informed of such research and other technology developments discussed above (e.g., best practices, standards development, open source software, etc.) will lead to more informed policy making that would not adversely impact, but rather encourage continued AI innovation and broader adoption.*

Recent advancements in the development of data de-identification techniques such as differential privacy and homomorphic encryption provide another example of the complementary relationship between technology development and AI policy. Privacy is a key pillar in establishing trust; however, AI systems that are used in contexts that involve people would need access to data about people to make informed decision. As more techniques that enable data access to be anonymized are developed, policy makers can leverage and encourage use of such privacy-preserving technologies by allowing de-identified data to be used for training AI systems, as long as they are not re-identified.

Leveraging AI to enable evidence-based policy making
With data becoming increasingly available through a variety of means, federal R&D investments should consider how AI can be used to enable evidence-based policy making in every part of the process, including identifying needs, budget development, implementation oversight and evaluation, in ways that enable government to improve funding of policies, initiatives and programs. Evidence-based policy making can reduce wasteful spending, expand innovative programs that can help to enhance U.S. competitiveness and economic growth, and strengthen accountability for program operations. Examples of roadmaps and evidence for improved policy making can be found in the Pew Charitable Trust report on this topic.

Adoption of AI in government services
AI has the potential to improve how the U.S. government interact with citizens, deliver services more inclusively, and address difficult societal issues. As these are challenges that may not be adequately addressed by business, they should be considered as part of the R&D plan.

By being early adopters of AI, the government can promote responsible and effective uses of AI, and help to develop some of the best practices to foster trust and broad adoption of these technologies. Deployment across multiple Federal agencies can lead to increased inter-agency coordination as well as further collaboration on AI research initiatives and strategies.

6. Conclusion

Microsoft appreciates the opportunity to provide this response to assist the NITRD NCO in updating the 2016 National Artificial Intelligence Research and Development Strategic Plan. We firmly believe that a sustained focus on AI R&D, in both the public and private sector, is fundamental to preserving U.S. competitiveness and global leadership, and is an integral part of shaping the development of a national AI policy and regulatory framework. This response provides our input on considerations for the R&D plan that would help achieve these objectives and establish a firm foundation for U.S. leadership in AI. Their implementation, along with a stronger inter-agency coordination process that is focused on opportunities for AI to enhance U.S. competitiveness and economic prosperity, and budgetary allocations that would correspond with stated AI objectives and R&D priorities, can also strengthen U.S. leadership in global dialogues on AI.

Respectfully submitted,

MICROSOFT CORPORATION
Rich Sauer
Corporate Vice President, Deputy General Counsel
3720 159th Avenue NE, Redmond, Washington 98052

For correspondence:
M-H. Carolyn Nguyen
Director, Technology Policy
901 K Street, NW, Washington, DC 20001
(202)831-6475
cnguyen@microsoft.com