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.

AAAI Response to NITRD RFI:
National Artificial Intelligence Research and Development Strategic Plan

October 26, 2018

This document is in response to the Request for Information (RFI) issued by the United States National Science Foundation in support of the Networking and Information Technology Research and Development (NITRD) National Coordination Office (NCO) and on behalf of the Select Committee on Artificial Intelligence through Federal Register Notice: 82 FR 48655 (https://www.nitrd.gov/news/RFI-National-AI-Strategic-Plan.aspx) to collect input from the public to update the 2016 National Artificial Intelligence Research and Development Strategic Plan (https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf).

This submission is an organizational response from the Association for the Advancement of Artificial Intelligence (AAAI). AAAI is the largest AI Society in the world, with over 3000 members. This response was developed by the President and Past-President of AAAI in coordination with the Government Relations Committee of the AAAI Executive Council.

The 2016 National Artificial Intelligence Research and Development Strategic Plan was a significant document to lay out priorities for government effort and investments in AI. As requested by the RFI, we comment here on each of the strategic goals in that document.

Strategy 1: Make long-term investments in AI research

Past investments in AI research have led to substantial advances that are now at the forefront of many important applications. Advances in speech recognition have led to ubiquitous conversational chatbots, in natural language processing to impressive machine translation systems, in knowledge graphs to knowledge-guided search engines and applications, in constraint reasoning to scheduling and manufacturing design engines, in robotics to both self-driving cars and interactive robots. Deep learning has revolutionized our ability to exploit very large labeled datasets. Many decades of research were needed for the technologies to mature and these applications to emerge.

There are still many important aspects of AI that present significant challenges:

- Common Knowledge: How can AI systems incorporate commonsense knowledge about how the world works that all humans possess? How can AI systems access up-to-date knowledge about notable entities in the world (people, institutions, places, etc) that are important for a given task? How can AI systems be aware of important knowledge of appropriate and inappropriate behavior in social interactions? Existing methodologies (supervised learning from text, hand-coded knowledge bases) have so far failed to provide this knowledge. Such knowledge is important for allowing AI systems to operate in open environments and especially to interact effectively with people.

- Meaningful Interactions: What kinds of interactions (context setting, explanations, visualizations, transparent structures) will make it easy for people to collaborate effectively, safely, and reliably with AI systems? How can AI systems best augment human decision-making and vice versa, and become “human aware”? How can AI systems master human language to communicate effectively? How can AI systems best support complex collaborations? How can humans teach AI systems to expand their knowledge? Interdisciplinary research that engages human factors, cognitive psychology, and AI research communities toward these challenges is needed.
- Robust and Sound AI: How can AI systems be made robust to un-modeled aspects of the world? No system can model (or be aware of) the full complexity of its surroundings. Living systems appear to behave robustly even in the presence of these “unknown unknowns”. One important direction is to develop ways that AI systems can introspect about their capabilities and limitations. Methods for continuous self-monitoring to detect failures and limitations are needed.
- AI Safety: Modern AI systems continue to learn from their experiences after they are deployed. Methods are needed for ensuring that this adaptation respects safety and functionality constraints. Formal verification techniques may be useful but are limiting for software systems that adapt, plan and learn and will require new methods; self-monitoring capabilities may be essential.
- AI Theory: What are the theoretical limits of AI systems? There is computability theory for all of Computer Science, the theory of inductive inference and Probably Approximately Correct (PAC) learning for machine learning, and intractability results for various logical representation systems. Can tighter formal limits or better theoretical understanding be achieved for specific classes of AI systems/methodologies (e.g., deep learning)?
- Intelligence: How can AI systems help us understand the brain and intelligent human behaviors, and advance fundamental understanding of intelligence? How can we model intelligence in all forms (human, animal, synthetic)? How can we use these models to develop AI systems that address differences in mental abilities (e.g., to help treat autism, or to support independent elderly living)?

The challenges of understanding intelligent behaviors have proven towering, profound and routinely underestimated. They are unlikely to be resolved through existing funding programs or industry research projects. They will not be solved without significant investments in shared resources and infrastructure.

The federal government, research institutes, universities, and philanthropies should create significant opportunities to fund multidisciplinary AI research. For example, the achievement of systems with robust common sense reasoning and human-level decision-making expertise will require sustained collaboration between disparate and (currently) largely disconnected research communities in psychology, social sciences, and AI. There is also increasingly a need for policy and technology research to mutually inform and align.

Past experience suggests that effective multi-disciplinary multi-institutional research require special programs significant and sustained funding. NSF has had several interdisciplinary research programs over the years that follow this model and are very effective (e.g., ITR, CDI, SEES) , but the impetus only lasts for a few years. In order for a faculty member to take the risk and effort of building an interdisciplinary research program, there needs to be the prospect of continuing funding opportunities over the long term. This prospect can also encourage universities to create interdisciplinary faculty positions to attract candidates that may not fit neatly into one discipline.

Many important research areas in AI cross government agency boundaries (e.g., the Departments of Justice, Commerce, Energy, and Defense as well as NSF and NIH). The government should create cross- agency working groups to develop research roadmaps and funding programs to promote this research. Important research aimed at social good crosses levels from city governments to regional utilities to law enforcement at all levels (including municipal, state, FBI, Coast Guard, and Border Control). Mechanisms need to be created that support the development of research programs spanning these levels.

Strategy 2: Develop effective methods for human-AI collaboration

Most of the significant AI research challenges outlined above are applicable to this point. We cannot underestimate the general intelligent capabilities expected of a collaborative assistant or partner, such as commonsense, language, robustness, etc.

Fundamental issues of trust lay ahead in human-AI collaboration research. Strong programs on AI safety, ethics, and privacy would surely open up AI systems to a much broader public benefit. Among the important research questions that should be addressed are the following:

- Data and methodological bias – Much of the potential of AI systems follows from the ability to extract patterns from large data sets and turn these results into forms of actionable information and advice. However, there are several sources of bias that can impact the accuracy of the conclusions that are drawn. If the data were collected in a biased way or if data quality (noise, missing values, precision) exhibits biases, then the extracted patterns can be biased. Likewise, biases can come from the assumptions made by the algorithms applied to extract patterns and draw conclusions (e.g., active learning methods, cost-sensitive methods, etc.). How can we define "bias"? How can we detect it? How can we eliminate or control it?
- Collaborative decision-making – In the short and medium term, mechanisms for AI systems interacting with and supporting human decision-makers (in contrast to fully autonomous AI systems) will constitute the primary path to application and benefit, and this requirement exposes several gaps in current capabilities. Very few AI systems are able to explain their reasoning, either through summarization of logical inference, visualization of key consequences, or simulation of expected decision behaviors. This

capability is fundamental to broader application of AI systems: (a) to allow people and computers to work well together (effectiveness, safety, reliability), (b) to enable people to attain appropriate levels of trust in AI systems and promote further automation, (c) to support post mortem examination of decision making for credit assignment and possibly for legal purposes, and (d) to help AI system developers detect and repair errors in the system.

- Ethical decision-making – As we move toward applying AI systems in more mission critical types of decision-making settings, AI systems must consistently work according to values aligned with prospective human users and society. Yet it is still not clear how to embed ethical principles and moral values, or even professional codes of conduct, into machines.

Strategy 3: Understand and address the ethical, legal, and societal implications of AI

There is tremendous potential for AI to serve the public good by creating decision making tools that incorporate a comprehensive set of sensor signals into highly-accurate models that enable both rapid response to crises, as well as medium and long-term planning. AI systems are already being applied to optimize many aspects of city services including utilities, transportation, law enforcement, and poverty mitigation. In many areas of science and engineering, AI systems are being used to transform research practices and accelerate discoveries. For decades, AI systems have been an important component of space exploration missions. AI systems have also been shown to be useful for detecting manipulation of social media and many forms of financial fraud. Robots are being used for information gathering and for search and rescue in catastrophic events.

Looking ahead, we anticipate many other high-impact applications with novel and profound social benefits in the short to medium term, including early detection of serious medical conditions from routine test data; more efficient preventive medicine and healthcare delivery including home-based care; improved ecosystem and resource management; personalized education; detection of public health hazards (e.g., presence of lead paint) from analysis of diverse data; and automated testing and safety checking of complex software/hardware systems that will be ultimately operated by people. In general, AI has had strong success (often surpassing human expertise) in problem domains that are narrowly scoped and well structured; and applications that possess these characteristics are prime candidates for short-term benefit.

It is crucial to expose to the current and potential beneficial impacts of AI to address diverse societal problems. AAAI holds an annual Conference on Innovative Applications of Artificial Intelligence that promotes the dissemination of AI work with practical impact, and provides a forum for researchers and practitioners to discuss challenges and lessons learned in embedding

AI systems in all aspects of society and public institutions. New forums that bring this research and applications closer to the public need to be fostered.

The deployment of AI systems in increasingly more complex decision-making settings raises important issues around agency, ownership, fairness and responsibility. These have been topics of long-standing interest in the AI community, but much more research is needed on characterizing AI systems in those terms, and understanding how to ascribe agency and responsibility. AI systems need to incorporate technical mechanisms when possible to capture quantitative metrics of their behavior as well as the environment where they operate. Another concern that requires study is that of providing protection against potential power asymmetries that might arise (e.g., through manipulation and/or exploitation of AI systems) between those with insight and understanding of AI technologies and those without it. Finally, since AI systems often rely on personal or sensitive data, research needs to continue to address general data privacy issues in software and database systems.

Given the potentially unique character of AI systems, any broader forum convened to discuss policies concerning AI should include representation from the AI research community. Laws and regulations in each of these areas will need to evolve over time, but individual cases make bad law, so it is important that legislatures put some reasonable statutes in place. The legal aspects of AI systems are complex, and as such they should be approached incrementally as a function of both (1) degree of the system autonomy permitted and (2) problem domain (e.g., autonomous vehicles, medical diagnosis) rather than pursuing discipline-wide blanket laws.

There is strong interest in the AI community to continue to be a leading force on these topics. The AAAI bylaws state that our organization promotes research and responsible use of AI. AAAI holds an annual Conference on Artificial Intelligence, Ethics, and Society that brings together multi-disciplinary researchers working in these areas. AAAI is also a founding member of the Partnership on AI, a consortium crystallized by industry that is focused on promoting best practices on AI technologies.

The socioeconomic implications of AI are difficult to predict. It is likely that AI-based technology will improve productivity in many industries, but it is unclear how the benefits of these productivity improvements will be distributed through the economy. AI systems continue to be developed to improve education, particularly in STEM fields, through personalization and one-on-one tutoring. AI systems can also improve access through natural language interaction and virtual presence. The government should fund research to monitor social/economic impacts of AI systems by collecting statistics and studying how AI systems affect the nature of work, the growth of productivity, and the distribution of wealth. Regular reports to government should be required, so that appropriate policies can be introduced if they become necessary.
Care should be taken to distinguish economic impacts due to AI systems from those that are due primarily to other factors (e.g., other information technology, outsourcing practices). The government should seek to build greater in-house technical expertise in AI as a practical means of gaining understanding and getting on top of these issues.

Strategy 4: Ensure the safety and security of AI systems

When AI technology is incorporated into systems that contribute to high-stakes decision-making, errors can have severe consequences. In the past, AI research and development has not always attended to these risks. Research is urgently needed to develop and modify AI methods to make them safer and more robust. A discipline of AI Safety Engineering should be created and research in this area should be funded. This field can learn much by studying existing practices in safety engineering in other engineering fields, since loss of control of AI systems is no different from loss of control of other autonomous or semi-autonomous systems. AI technology itself can also contribute to better control of AI systems, by providing a way of monitoring the behavior of such systems to detect anomalous or dangerous behavior and safely shut them down. Note that a major risk of any computer-based autonomous systems is cyber attack, which can give attackers control of high-stakes decisions.

There are two key issues with control of autonomous systems: speed and scale. AI-based autonomy makes it possible for systems to make decisions far faster and on a much broader scale than humans can monitor those decisions. In some areas, such as high-speed trading in financial markets, we have already witnessed an “arms race” to make decisions as quickly as possible. This is dangerous, and government should consider whether there are settings where decision-making speed and scale should be limited so that people can exercise oversight and control of these systems.

Most AI researchers are skeptical about the prospects of “superintelligent AI”, as put forth in Nick Bostrom’s recent book and reinforced over the past year in the popular media in commentaries by other prominent individuals from non-AI disciplines. Recent AI successes in narrowly structured problems (e.g., IBM’s Watson, Google DeepMind’s Alpha GO program) have led to the false perception that AI systems possess general, transferrable, human-level intelligence. There is a strong need for improving communication to the public and to policy makers about the real science of AI and its immediate benefits to society. AI research should not be curtailed because of false perceptions of threat and potential dystopian futures.

Strategy 5: Develop shared public datasets and environments for AI training and testing

The promotion of open and FAIR (findable, accessible, interoperable, and reusable) data initiatives (be it data about cities, government, biomedical experimentation, the environment, materials engineering, education, etc.) would likely accelerate AI application development in many problems of societal interest/benefit, since AI researchers often end up pursuing problems where data is openly available.

Strategy 6: Measure and evaluate AI technologies through standards and benchmarks

Benchmarks and datasets annotated with answer keys (solutions or class labels) are crucial for steady improvement of AI algorithms. To apply supervised learning to acquire broad, commonsense knowledge, labeled data sets are needed about common sense situations. Similarly, to give AI systems better understanding of appropriate (ethical) behavior, data sets are needed describing decision making situations and the ethical and unethical actions that could be taken in those situations.

Incentive prizes, if large enough, can have a major impact. However, they generally reward people who already have enough resources that they can bring to bear and who can take the risk of spending their own funds even if the probability of winning a prize is low. Providing some form of participant support for non-traditional teams that wish to compete for incentive prizes is critical for broadening participation.

Current government acquisition standards and rules, particularly in DoD, are acting as disincentives to the development of advanced technology. It can take upwards of a decade or more for AI systems to be proven and transitioned into operations. The government should define new processes for the certification of adaptive/AI technology so that DoD and other government agencies can easily acquire it.

Strategy 7: Better understand the national AI R&D workforce needs

Many academic institutions are struggling to keep up with the explosive growth in student enrollment in AI. Universities are creating faculty positions, but the demand for AI experts in both academia and industry remains a challenge.

There is currently a significant pull of academic research and teaching expertise toward AI companies due to financial packages that universities cannot match, computing facilities and other infrastructure that is not otherwise available, and sustained financial support. Computer science departments are struggling to accommodate special arrangements for their best AI faculty being courted by industry, and to retain strong undergraduates to pursue doctorates in AI rather than join the excitement offered by industry. This trend benefits short-term application of AI research and represents new symbiotic possibilities for industry and academic AI advances. However, it negatively impacts fundamental academic AI research as well as the training of future AI researchers and practitioners.

Universities are allocating faculty positions to AI-related areas. However, for prospective faculty members to succeed, they need to be able to compete for research funds from federal sources (NSF, ONR/ARL/AFOSR, DARPA, NIH, NIST, DOE, etc.) which are often prioritized for other areas of computing or applications. In some federal agencies, AI researchers participate in very small numbers because they prioritize specific science areas in detriment of broader AI research that could have significant impact. Congress needs to allocate additional funds to these agencies to enable them to invest in AI research programs. Further, it is important that the government continue to advocate and invest in longer-term, fundamental AI research. It often takes many years to achieve breakthroughs that are key to solving particular societal problems, and no one has the crystal ball to fully predict what these will be.

Universities find increasingly hard to attract graduate students and faculty that find a career in other countries more appealing. Difficulties in obtaining visas upon graduation deter many of the best students from coming to US academic institutions. International PhD graduates are leaving the US in larger numbers than before, in part due to immigration constraints but also due to the availability of attractive opportunities for AI overseas.

A commitment of the US government to significant and sustained AI funding would make such faculty positions more attractive both to potential faculty members and to their institutions, as well as to graduate and post-doctoral students. Government could also improve the training of future AI researchers by greatly increasing the funding available for graduate fellowships such as those provided by NSF and DoD.

There is increasing demand for computer science courses and extra-curricular robotics activities in K-12, and this enthusiasm should be seized to feed the nation's pipeline of AI researchers looking to the next many decades. Steps should be taken to make introductions to AI topics such as machine learning, planning, knowledge representation, and robotics part of the core undergraduate curricula for non-computing majors as well as in high schools.

Particularly important is to promote initiatives that increase participation of underrepresented groups, marginalized populations, and underserved regions in terms of opportunities for education and training, programs to lower barriers to join the AI workforce, and initiatives to broaden the applications and social impact of AI.

There is also need for more basic education and outreach activities to the general public on the capabilities and potential of AI technologies.

Thoughtful investments in AI education will have a profound impact in the nation's future. AAAI holds an annual Symposium on Educational Applications of Artificial Intelligence that provides a forum for teachers and researchers to discuss their work on improving materials and access to AI education. AAAI has recently partnered with the Computer Science Teachers Association (CSTA) to develop national guidelines for teaching K-12 students about AI. Much more needs to be done in these areas.

Strategic planning for the future of AI for education and academia is key to the health of the field and for the country at large.