Computing Community Consortium's Response to RFI "Information to the Update of the National Artificial Intelligence Research and Development Strategic Plan"

Written by: *David Danks (University of California, San Diego), Maria Gini (University of Minnesota), Odest Chadwicke Jenkins (University of Michigan), Sven Koenig (University of Southern California), Daniel Lopresti (Lehigh University), Melanie Mitchell (Santa Fe Institute), Katie Siek (Indiana University, Bloomington), Ufuk Topcu (University of Texas at Austin), Holly Yanco (University of Massachusetts Lowell) and Maddy Hunter (Computing Community Consortium).*

The National Artificial Intelligence Research and Development Strategic Plan articulates how the United States can accelerate advances in Artificial Intelligence (AI) through strategic investments in research, infrastructure, assessments, education, and partnerships. We commend the National Science and Technology Council on updating The National Artificial Intelligence Research and Development Strategic Plan since our advances in AI technologies and our understanding of the unintended consequences of AI have significantly changed over the last six years. We agree that AI provides "tremendous opportunities" to improve many facets of life and that this is largely due to the federal government's investment in research. However, we emphasize that federal funding is crucial in *all* scientific research - from fundamental computing to social science - to ensure we are creating innovative, reliable, trustworthy, domain relevant, and socially responsible AI innovations.

Of interest specific to this RFI (and an update to our RFI response in 2018[1]), the CCC recently published a 20-Year Community Roadmap for Artificial Intelligence Research in the United States[2] that details the need to invest in fundamental AI research, educate a diverse, inclusive pipeline to make these needed innovations, and develop the infrastructure needed to support transformational AI. We point to the AI Roadmap as a prime example of the outcome from a community-based process, which we believe would work well for answering the open questions we raise below, as well as for its approach of identifying ambitious "stretch" goals requiring a long-term view and

[1] https://cra.org/ccc/wp-content/uploads/sites/2/2018/10/CCC-Response-to-AI-RFI.pdf
[2] https://cra.org/ccc/visioning/visioning-activities/2018-activities/artificial-intelligence-roadmap/

associated milestones that we think is vital for the strategic plan. More specifically, we offer that the plan should emphasize three important themes that we explore in more detail below:

- **Diversity** provides a foundation for the success of the U.S., and those many dimensions of diversity must be reflected in our AI efforts. We do not mean only diversity of people, but also diversity of ideas, theories, methods, projects, evaluation systems, and much more.
- **Education** is critical to addressing both pipeline challenges for the next generation of researchers and developers, and also the need for AI literacy across the population.
- **Foundational Research** enabled the AI revolution of recent years, but such research requires a longer-term view. Progress in AI requires both use-inspired research for the needs of now, and transformational basic AI research on the ideas of the future. Moreover, this foundational research must span across many disciplines so that we can understand, shape, and create AI for the betterment of all.

We now explore how these three themes can be incorporated into the eight strategies laid out in the National Artificial Intelligence Research and Development Strategic Plan.

Strategy 1: Make long-term investments in AI Research

The 2019 Plan talks about "America's scientific leadership" and "addressing societal issues" while mitigating negative effects caused by "bias, equity, or other concerns" without mentioning any investment in the **education** and training of the pipeline. The current **education** system is not equitable, and this puts future **diverse** scholars, who would contribute to this vision, at risk of not being able to participate. Successful research and development (R&D) starts with the pipeline. PhD applications are low across the country and a more robust pipeline is needed to create better tech/AI in the future, to train the next generation of AI researchers, and to **educate** the general population on how AI technologies will impact their lives and about the positive and negative aspects of AI. In addition to specifying what the goals are, it is important to provide some guidelines on how to achieve them. The processes we have now are not sufficient to bring our current goals to fruition, we must focus on creating new strategies and processes that will.

Strategy 1 is goal-centric and focused on specific capabilities, but research is always uncertain and surprising. The focus should be on creating processes that will sow many seeds and cultivate scientific exploration in addition to scientific rigor. An investment in

AI has to produce both ideas and people that move AI forward in positive directions. Even if a large investment is given, the current peer and merit review practices will not cultivate transformational ideas in AI or produce the necessary AI workforce. The review processes we have in place reward siloing into intellectual comfort zones among like-minded groups that gravitate to similar ways of thinking. Given how neural networks catalyzed over the last 20 years (mostly through international non-US support), we worry that our nation will miss the boat on the next transformations in AI if we continue to choose investments in this manner. Such practices tend to reward short-term results rather than **longer-term exploration**. This partially explains the rise and dominance of industry-led research groups in publication venues. Furthermore, the people being trained in AI tend to overfit to the current "hot" topics, and this does not lead to research labs that produce a workforce with **diverse** sets of ideas or skill sets. There is a risk that focusing solely on the AI workforce will miss the mark for creating systems that (1) address national needs and (2) utilize justice-oriented frameworks to solve societal issues.

Many topics are discussed in the report under Strategy 1. Additional thoughts are:

- *Data focused methodologies for knowledge discovery,* overlaps with Section 5,
- *Fostering research on human-like AI*, is connected to Strategy 2.
- *Pursuing research on general-purpose artificial intelligence* addresses the issue of needing more than machine learning methods, but is vague and talks about general intelligence in a way that does not provide specific ideas on what should be done. The concept of general intelligence is not sufficiently specific and there are no guidelines on how to move in that direction.
- *Advancing hardware for improved AI* and *Creating AI for improved hardware* and *understanding theoretical capabilities and limitations of AI.* There is emphasis on the hardware needed to assess performance. While the importance of hardware should not be underestimated, the sections minimize the importance of theoretical results and focus mostly on performance.
- *Developing scalable AI systems*. Scalability is critical, but communications, networks, and interactions with other systems are mentioned briefly, together with mentioning that distributed systems are more robust to failures. Issues related to multi-agent systems, interactions of agents, game theoretic approaches, norms for the behaviors of the agents, etc. are not mentioned.
- *Perception* and *robotics* emphasize collaboration, which is obviously important, but neglects issues such as adversarial systems, or systems that operate in disguise to collect data or to interfere with legitimate operations. Those issues should be addressed in Section 4.

Strategy 2: Develop effective methods for human-AI collaboration

No matter how many resources are applied to the development of fundamental and applied AI research, poor human-AI collaboration will prevent the system from achieving its full potential, as people will mistrust, misuse, and discontinue using systems they do not understand. Human-AI collaboration requires **foundational research** on the development of explainable AI systems that still achieve the required performance. Systems need the ability to answer, at varying levels of complexity, human questions about how decisions are made.

Human-AI collaboration goes far beyond the future of work, a focus of the 2019 Plan. People are increasingly interacting with AI in their homes (e.g., Amazon Echo, Google Home), in their cars (e.g., computer vision systems for safety), on their phones (e.g., mapping, shopping), for healthcare (e.g., automated diagnosis systems), in online shopping and entertainment (e.g., recommendation systems), and when shopping in the real world (e.g., Amazon Go). This revision of the plan needs to recognize the **diversity** of ways that people are already interacting with AI, and develop frameworks that can be applied to enable common interaction methods.

The revised plan must also recognize the **diversity** of interaction roles; much of the 2019 Plan frames the human as an operator of the AI. We aim for fundamental discoveries that enable the same richness of interaction between a person and an AI that currently exists between two people, with a shifting of roles depending on the current situation and the abilities of each. Ultimately, this would allow us to achieve a level of human-AI collaboration where the AI is a teammate of the person.

Not only do we need to create AI systems that are usable by all people, but we must also **educate** the public, starting at the K-12 level, about AI, probability and statistics, and computational thinking. Even with the best designed human-AI collaboration, without AI literacy, people will still effectively be bystanders to the technology.

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

AI has the power to provide enormous benefits, but also to cause significant harm. The inclusion of ethical, societal, and legal issues in the 2019 Plan is crucial in ensuring that our AI systems provide substantial benefits while conforming to relevant legal requirements. However, the current structure of the 2019 Plan separates these issues into a distinct strategy, rather than recognizing the ways in which ethical, societal, and legal issues arise throughout the lifecycle of AI creation, including the design, research,

development, and deployment stages. As a result, this particular strategy has a narrower focus that omits central ethical and societal issues. For example, the current strategy element on “building ethical AI” focuses on explication of principles to characterize ethical behaviors by AI systems. However, ethical AI performance depends on the values that the AI is intended to support, the contexts in which it is deployed, and many other aspects that go far beyond principles. Similarly, “designing architectures for ethical AI” focuses on systems in which ethics is explicitly represented, but the ethics in an AI usually arise from design and development decisions (e.g., about what errors are deemed “acceptable”) rather than explicit representations. We encourage the next revision to adopt a “whole lifecycle” approach to ethical, societal, and legal issues, with an emphasis on the ways that myriad decisions that might appear “merely technical” can actually have significant ethical, societal, and legal import.

This approach would be greatly enhanced by **foundational research** on best practices for designing, developing, and deploying AI systems that implement and support people’s values. Ethics and societal impact cannot be reduced to simpler measures of, say, reliability or security. At the same time, we do not currently have clear best practices or other ways to guide AI researchers and developers towards more ethical and socially beneficial systems. This much-needed research should also provide mechanisms, processes, and practices to ensure that **diverse communities** are supported by AI, rather than providing benefits for only a few.

We also note that the current strategy contains surprisingly little focus on legal and regulatory issues, despite the current title. There is increasing recognition and understanding that current regulatory and policy frameworks are insufficient for present-day and near-future AI systems. The revision should thus call for additional research, including the creation of regulatory “sandboxes” to test different ideas, on novel legal and regulatory approaches to ensuring AI safety and benefit.

Finally, we suggest that this strategy should call for the development and assessment of novel **educational programs** to ensure that people have appropriate knowledge and skills to assess and respond to ethical, societal, and legal concerns or opportunities with AI. If we want to have more ethical AI systems that better support people’s values, then relevant people need appropriate training. Most obviously, these issues should be incorporated into university curricula and upskilling/reskilling programs around AI. However, the needed **education** extends beyond the developers–for example, people engaged in software acquisition need to understand the technical, ethical, and legal possibilities and risks with AI systems. AI creators must learn how to do better, and the rest of us must learn what to demand so that they do better.

Strategy 4: Ensure the safety and security of AI systems

As an overarching theme and essential piece in this response, re-strategizing **education** and curriculum development is a vital piece of ensuring the safety and security of AI systems. Many issues pertaining to safety and security will not be resolved through piecemeal advances, but instead require a fundamental change in mentality in the way that we design AI systems and view consequential responsibility of these systems if things go wrong.

It is challenging if not impossible to ensure the safety and security of a system that was built without factoring in the safety- and security-related constraints into the development of the system. Therefore, ensuring safety and security cannot be an afterthought and can be achieved only through principled development processes that diffuse the need for safety and security in each step of development from specifications to certification.

The portion of the 2019 Plan that states, “the notion of safety (or security) by design might impart an incorrect notion that these are only concerns of system designers,” may come across as support to reduce the burden on designers rather than imposing responsibilities on all parties. The document should not have an indication that designers might be responsible for less. What they do is not sufficient for safety/security (or any other feature that contributes to the trustworthiness of AI) but good design is absolutely necessary for safety/security. The responsibility of ensuring the safety and security of AI systems needs to start during the design process, beginning with specifying conceptually and technically what we expect from them. These decisions made in the early stages restrict the decisions later in the lifetime and may have more severe and at times unpredictable and unintended consequences. It is definitely not easy to specify acceptable or desirable properties for systems with AI, but we need to start somewhere and adopt an iterative approach where the need for certification is a central driver of the design process.

In addition, the prominence of adversarial machine learning in Strategy 4 will downplay the importance and difficulty of securing systems with AI. It is rare that an AI functionality is used in isolation. We cannot overemphasize the notion of **diversification** of data sources and tools for building intelligence and methods for it. It is likely that integration of functionality is a security vulnerability, i.e., we may secure each in isolation but, when integrated, new vulnerabilities emerge. Most importantly, we need to leverage what we already know and incorporate that into techniques for AI. Current ML techniques reinvent the wheel whenever they face a new task. Therefore, it is

necessary to develop theory and mechanisms to address the vulnerabilities that arise when AI functionality is integrated into bigger systems.

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

We strongly agree that socio-technical infrastructure is needed “to support reproducible research in the digital area." This is an excellent start, but we encourage the administration to think more broadly to capture better sociological and contextual information to help us understand what is going on and how this goal will be implemented. The strategy acknowledges that more support is needed for semi-structured and unstructured data; we emphasize that researchers in social sciences should be consulted to get a better understanding of the rich, contextual data needed for future knowledge generation. In addition, multiple unstructured data streams need to be easily integrated for improved sense-making. These are open research challenges that need to be carefully addressed to answer more challenging, worthwhile societal issues (e.g., climate change, proactive healthcare interventions).

We highlight NAIRR considerations[3] for “building inclusive datasets and governance approaches that treat equity as a core design principle,” but also call for reviews of current attempts at equity in data collection (e.g., AllOfUs[4]) to identify best practices and issues, and to ensure historically exploited groups receive benefits from participating.

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

Benchmark datasets are very important and the strategy rightly emphasizes this. However, it’s important to note that there are a lot of problems with the kinds of evaluations done currently with widely used benchmark datasets in AI / ML.

Here are some of the issues:

- **Assumption that data is IID (Independent and Identically Distributed):** Today, most ML systems are evaluated by splitting data randomly into training and test sets, and evaluating accuracy on test sets - assuming the data is IID. However, real world data is often (almost always) not IID. More research needs

[3] https://www.whitehouse.gov/ostp/news-updates/2021/12/15/readout-of-the-fourth-national-artificial-intelligence-research-resource-nairr-task-force-meeting/

[4] https://allofus.nih.gov/

to go into how to construct benchmarks that test for more than just accuracy under IID assumptions.

- **Shortcut learning:** Many current standard benchmarks in ML have been shown to allow for "shortcut learning"—that is, a system can perform very well on the benchmarks due to subtle statistical correlations that are unrelated to the actual tasks we want the system to learn. These shortcuts are often hard to uncover.
- **Overfocus on standard benchmarks in evaluating research:** The reliance on standard benchmarks in today's ML research sometimes translates into an attitude that if a system is not tested on the standard benchmark, or if it scores below "state of the art" on that benchmark, the research is not worth publishing. This can have the effect of stifling important new ideas and encouraging research that provides only very incremental progress.

Additional thoughts:

- The strategic plan says that we need standards and benchmarks to ensure "accuracy, reliability, robustness, accessibility, and scalability." This list should also include generalization, robustness to shortcut learning, and robustness to adversarial attacks.
- We should not simply increase the availability of AI testbeds, as recommended in the report; we need to rethink the entire enterprise of how to create test beds that actually test for what we want systems to do. This is currently a big unsolved problem in AI/ML.
- The report says "Government leadership and coordination is needed to drive standardization and encourage its widespread use in government, academia, and industry." But how will the government actually do this in a way that creates standards and testbeds that really test what we want?

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

It is necessary to resolve the current curricular bottlenecks that limit growth of the AI workforce. Modernization of our approach to AI curriculum is desperately needed to grow to meet needs for national technological competitiveness. Currently, the expansive growth of AI has led to a two-sided demand crunch for developing an AI workforce. This AI demand crunch has pushed our current curricular structures beyond their limits. In particular, AI has grown from being a specialty research area into a basic literacy needed by people across society. The needs for education spans across the needs of people who will be leaders in the innovation of AI, practitioners that will design and develop AI products and services, and users and consumers who need to make

effective choices. There is a swelling demand for enrollments in courses[5] and to understand and to serve such a broad need in AI **education** for all people we must do the following:

- Cultivate a new populous of educators in AI
- Data collection more than anything else - particularly are there specific places we know where the pressure is bottlenecking curriculum (e.g., not enough people to provide stewardship to get people through the system)?
- There needs to be an interdisciplinary approach to address the workforce needs and create equitable systems. The pipeline has to be thought of more broadly and there needs to be training of those who help the pipeline).[6]
- Relative to the other strategies, Strategy 7's language is passive and non-committal - merely understanding, rather than trying to do something about it. It would be better to say "Realize a competitive AI R&D workforce" and involve the creation of programs, support for research about how to build skills etc.

Strategy 8: Expand public private partnerships to accelerate advances in AI

Public-private partnerships involve some significant tradeoffs. We recognize some notable positives (e.g., access to computing resources and data that only industry may possess, first-hand awareness of problems that have practical applications benefiting society, financial resources to make large investments in applied research), but at the same time there are risks, including the potential for conflicts of interest, whether real or perceived, that may impact the independence of university-based researchers. This kind of independence – to question and to criticize – has never been more important given the growing pervasiveness of commercial AI applications and certain well-publicized failures of recent years, discovered only after the fact. This relates to the **diversity** theme mentioned earlier.

In addition, while several major success stories are called out in the 2019 Update to National AI R&D Plan (e.g., CNNs), it should be noted that funding for basic AI research in the US has been uneven over the decades, and important contributions made as a result of non-US funding and by non-US researchers needs to be acknowledged. Sustained funding over long durations (much longer than typical grant award periods) is vital; the path from the germ of an idea to commercial application is a long one, and many US-based researchers have had to be very creative where they seek ongoing support. This relates to the **foundational research** and **diversity** themes mentioned earlier. (For an example of the time scales involved in transitioning from research to

[5] https://www.nytimes.com/2019/01/24/technology/computer-science-courses-college.html
[6] #noCode.org

practice, see the famous “tire tracks” diagram produced by the National Academies: Information Technology Innovation: Resurgence, Confluence, and Continuing Impact.)[7]

While existing efforts at establishing public-private partnerships have shown promise (as suggested in the sidebar on Page 42 for the 2019 Update), we are curious about the extent of the impact on the US research community as a whole, and in particular the involvement of researchers who are faculty members in “standard” academic settings. If the number of researchers who benefit is relatively small, and, more significantly, if these researchers are not also serving the urgent educational mission we face (as tenure-track faculty who conduct research and also teach undergraduates and graduate students), then they are contributing only part of the solution. We need our students – the future AI workforce – trained at the same institutions that are being funded to conduct leading research. These relations to the **education** theme mentioned earlier.

The financial “pull” of well-funded tech companies, along with the attractive environments they provide to researchers, creates a tension that has led to a “brain drain” in academia. This same effect has been visible at the level of domestic graduate students as well. Public-private partnerships could help address this issue, or they can exacerbate it. This concern should be a topic whenever such partnerships are discussed. This is another link to the earlier **education** theme.

For additional thoughts offered on behalf of the computing research community regarding the increasing closer connections between private industry and academic research, we cite the recent CCC whitepaper, “Evolving Academia/Industry Relations in Computing Research”.[8] Quoting from the summary of this whitepaper: “Particular attention needs to be focused on issues related to department culture, potential conflict of interest, intellectual property, and ensuring that students continue to have sufficient faculty mentoring and contact to prepare them for their career.”

[7] https://www.nap.edu/read/25961/chapter/1

[8] https://cra.org/ccc/wp-content/uploads/sites/2/2019/06/Evolving-AcademiaIndustry-Relations-in-Computing-Research.pdf