Computing Research Association - Industry (CRA-I) Response to Request for Information (RFI) on Developing a Roadmap for the Directorate for Technology, Innovation, and Partnerships at the National Science Foundation

July 2023

Lisa Amini (IBM), Elizabeth Bruce (Microsoft), Hank Korth (Lehigh University), Chris Ramming (VMware), Vivek Sarkar (Georgia Tech), Helen Wright (CRA), Ben Zorn (Microsoft)

The following is the Computing Research Association (CRA)'s CRA-Industry (CRA-I), a community-based organization, response to a Request for Information (RFI) on Developing a Roadmap for the Directorate for Technology, Innovation, and Partnerships (TIP) at the National Science Foundation (NSF). The mission of CRA-I is convene industry partners on computing research topics of mutual interest and connect them with CRA's academic and government constituents for mutual benefit and improved societal outcomes. Please note any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the authors' affiliations.

This response pertains to questions 1, 2, 3, 4, 6, and 7 from the RFI.

1. Prioritization

Extending TIP evidence sources beyond academia. The NSF community is well-known for its collaborative, evidence-driven approach to opportunity identification, prioritization, and project selection in an academic setting. But because the TIP mission extends to research translation and innovation (putting inventions into practice) it will need to extend its community and evidence-gathering processes to be inclusive of industry factors. And because the TIP mission relates to societal goals it will likely be advantageous to extend its evidence-gathering community to the non-profit sector for citizen engagement. Finally, because the TIP initiatives are likely to involve sociotechnical systems, it will be important to extend the evidence-gathering community to include legal, policy, standards, and regulatory organizations across government and industry. One systematic way to incorporate these diverse perspectives could be to selectively partner with visioning and planning groups that have been previously been established in stakeholder communities (e.g., the Computing Community Consortium (CCC), the Engineering Research Visioning Alliance (CRA), DARPA's study groups, Semiconductor Research Corporation, etc).

Regularly monitoring evidence for course correction. To be useful, “evidence” must be leveraged in the context of a decision analytic framework. Although no framework or method can be expected to produce “right answers,” these can be useful tools to surface evidence-based insights that result in higher-quality decisions. Importantly, decisions about large investments need to be revisited regularly in light of new evidence and in this respect TIP may need to operate differently than is customary in the lightly managed, largely open-loop research initiatives that are common in today’s NSF directorates.

Clarifying TIP’s mission relative to other agencies and organizations. To be effective and to win an enduring role in the panoply of government agencies, TIP needs to develop and advertise a strong sense of its mission, unique capabilities, and charter that can be used to triage opportunities and create effective interfaces with those other agencies. A low bar for success is the avoidance of duplicative work; the higher bar is to effectively develop synergies through partnership. As a role model, DARPA has effectively developed an organizational culture that allows program managers to triage new ideas by dismissing those best pursued by other actors (e.g., industry), focusing on projects with defined goals rather than open-ended research agendas, or not fitting the agency’s particular defense mission. A similar clarity of purpose for TIP needs to be developed, presumably with regard to TIP’s strengths and weaknesses. In taking these steps, TIP needs to be cognisant of synergistic government-related activity in this domain, particularly those of the DoD (DIU) and the US Intelligence Community (In-Q-Tel). There is definite overlap among these missions. TIP should look to partnering with these agencies where appropriate, but also to seek out critical areas those agencies may be less likely to consider.

As an illustrative example, TIP is uniquely advantaged by its close connection to researchers with a breadth of expertise to identify new directions; meanwhile other research agencies (DARPA, IARPA, NTIA, etc.) have a narrower community and technical scope. It could therefore leverage its relative breadth by pursuing priorities that fall through existing agency gaps. Meanwhile one TIP disadvantage compared to NIST, NTIA, DOE, etc is its relatively weaker industry understanding and connection, suggesting that it needs to explicitly coordinate with other agencies to increase its industry influence. Meanwhile there are gaps to be filled: for example, the Office of Science and Technology Policy (OSTP) arguably has the charter (but not a direct budget) to identify and pursue national priorities, and to coordinate agencies. In this context, TIP could potentially find a differentiated role in executing on those OSTP-identified national priorities in areas that are closely connected to TIPs prior research funding. TIP could differentiate itself by partnering explicitly with OSTP to coordinate with other agencies in developing and funding components of a coherent national strategy that go beyond what TIP could achieve in its present form.

Specific areas ripe for NSF attention. OSTP can be a source of illustrative and promising areas of investment, such as those found in the recently produced national strategy to advance privacy-preserving data and analytics. The strategy, developed with input from NSF and a broad set of stakeholders, proposes a range of worthwhile investment ideas. TIP could make it a priority to follow through on a subset of those investment connections, and to coordinate with other agencies to ensure that the resulting investments are coherently influencing the adoption of compelling ideas stemming from prior NSF investment. OSTP is following a similar RFI process to identify national priorities in artificial intelligence, and again NSF will be uniquely positioned to draw on its previously funded research and coordinate with other agencies to drive adoption of compelling research results. Within these general priority areas, NSF could also prioritize initiatives for maximal impact, for example, topics at the intersection of multiple industries that may be underinvested without explicit championship (e.g., RINGS at the confluence of cloud, networking, and telecommunications); topics ripe for transference from one field to another in the spirit of many other cross-sector "tire track" jumps (e.g., generative AI with application to accessible technology, biology, physics, etc.); implications of disruptive opportunities that are nearing comprehensive industry adoption (e.g., confidential computing as a not-yet-standardized paradigm shift in computing security).

2. Suitability

Renewed emphasis on public-private partnerships. NSF has traditionally invested primarily in the academic research community, leveraging peer review as its primary selection technique independent of the specific investment topic (e.g., basic research, translational research, research infrastructure). Pursuant to some of the observations above, both the "peer" community and the funding vehicles need to evolve to include a broader set of participants.

A renewed emphasis on public-private partnerships led or sponsored by "champion" companies could increase the sense of ownership by industry in shaping and leveraging research results. Examples (with lessons learned in some cases) include the Semiconductor Research Corporation's MARCO/FCRP/JUMP programs, the NSF RINGS initiative, various bilateral programs initiated by CISE, and longstanding NSF funding models such as I/UCRC and InTrans. An important feature of these programs is that they do extend the NSF's community to include new communities with independent perspectives on the available "evidence." The requirement that partner companies commit funding increases the odds of follow through and the gravitas of formational activities. However, several variations on the theme could further bring benefit.

First, in recognition that breakthroughs often involve developments at the intersection of two or more communities (with many examples shown in the National Academies "Tire Tracks" analysis) TIP may wish to refine its approach to multi-sector program management. For example, instead of aligning multiple stakeholders to agree on a single request for proposals or programs, multiple initiatives could be developed and tailored to address sector-specific opportunities. In addition, instead of assuming that all solicitations need to be issued by the NSF, or even that all strategic priorities are best pursued via calls for proposal, TIP could work with other agencies each with its own unique responsibilities. For example, if TIP pursues opportunities in a specific area like security and privacy, it may well be that a partner agency such as NIST should engage in a coordinated action (for example, developing reference architectures on a path to implementing new ideas, which it has done to good effect with the concept of "Zero Trust"). Meanwhile TIP could partner with narrow project agencies such as DARPA to advance the related state of the art with explicit consideration of "Dual Use" across industry and defense. TIP could likewise engage with regulatory agencies to develop compliance policies and government purchasing guidelines to encourage rapid development of specific ideas. Overall, TIP might be the agent behind coordinated "solicitation sets" across a range of agencies.

3. **Workforce**

Broaden support for education and workforce development across TIP programs. NSF TIP should consider incorporating an education and workforce development component into all programs. NSF TIP should intentionally design programs to reach a broader and more diverse set of communities, including women, minorities, and people with disabilities. Workforce development among these communities can have a virtuous cycle effect of having those individuals then contribute innovations that bring their unique perspectives that may otherwise be lost. Investing early in a diverse and skilled talent pipeline is critical for industry especially in areas like AI where the workforce needs are changing rapidly.

Address the need for diverse pathways. NSF TIP should support a diverse range of institutions including: universities and colleges; community colleges; and technical colleges when it comes to addressing education and workforce development needs. Increase support for students to access and pursue different learning pathways, including 2-year and 4-year degree programs, graduate programs, certificate and credential programs, and online programs in each of the emerging technology areas. Encourage alignment of secondary and post-secondary education and curriculum to ensure alignment of pathways into in-demand fields. Disruptive technologies, such as generative AI, by their nature require workforce support for life-long learning as the skills acquired in traditional pathways (university education) may become obsolete and require retraining and new understanding and skills are needed.

Increase internships and apprenticeship programs with industry in emerging tech areas. Support programs that enable students to gain exposure to industry, such as apprenticeships, paid internships, work-study, and industry mentorships. Expansion of programs to support student internships in industry focusing on research and social impact projects. Leverage federal government funding to help galvanize private sector coordination to establish more industry-based internship programs. Invest in support staff and infrastructure to run and manage.

> Model an "Industry - Research Experience for Graduates" program based on NSF's current "Research Experiences for Undergraduates" program that supports students working on research-based projects in emerging tech areas in industrial research labs. Extend to 6-9 months to allow students to gain deeper knowledge and experience.
>
> Support Learn-Practice-Apply (LPA) model programs that enable students to gain fundamental knowledge (classroom), hands-on research experience (research/project based courses), and exposure to state of the art use-inspired challenges (through work with industry partners).

Support for curriculum development, training resources, and co-teaching opportunities with industry. Support programs that advance STEM curriculum in emerging technology areas and resources to incentivize development of open access curriculum. Open access /open-source curriculum provides resources for many educational institutions to leverage, build-on, and customize curriculum in these areas (vs. build new curriculum from scratch). Opportunities and incentives for industry and academia to co-develop curriculum/training, co-teach courses, and co-mentor student projects driven by industry use cases. Externship programs for faculty and teachers to work alongside industry professionals providing opportunities to gain exposure to industry use cases and challenges; relevant programs; and new tools and methods. Finally, support existing professional organization mentoring programs such as SIGPLAN-M.

Invest in AI for education and training to scale and improve personalized learning. Invest in developing and scaling AI technologies for education. Including AI designed to support faculty, teachers, to scale personalized learning and teaching support (autograders etc). Support the translation to pilots and proof of concepts that can be rapidly adopted by all academic/education communities. Support responsible development of Generative AI and LLMs for use in education and training and ensure the community develops best practices and standards for student and teacher-use.

Support use of labor market data to better understand future workforce needs 1-5 years out - focusing on key emerging tech areas. Combine and leverage data from government agencies and private sector data sources to ensure data about industry needs is shared with state, local and tribal governments, as well as educational institutions and policy makers to help inform priorities and investments are aligned around future workforce needs.

4. **Addressing Societal Challenges**

NSF TIP should consider prioritizing technology areas in which hold the highest potential for 1) emergent abilities, 2) eminent and sustained impact 3) combinatorial effects of technologies. The following paragraphs elaborate on these criteria and highlight technology focus areas that meet, or are anticipated to meet, these criteria.

Emergent abilities. Emergence occurs when a complex entity has properties or behaviors that its parts do not have on their own, and emerge only when they interact in a wider whole. The individual parts interact and begin to work together to form something new, causing complex behaviors to arise that were not exhibited in individual or smaller numbers of parts. In AI, emergent abilities have been observed in large language models (LLMs) that were not present in smaller models, such as few-shot prompted arithmetic. Emergent functions in quantum materials, such as high-temperature superconductivity and colossal magnetoresistance, are believed to be crucial for developing the next-generation quantum technologies and potentially energy-efficient computing. In rapidly developing materials and technologies especially, such spontaneous and unpredictable phenomena can have impact at the level of societal, national and geostrategic challenges.

Eminent and sustained impact. The current rate and pace of technological change is unprecedented. Individual inventions, such as the telegraph, may be short lived, but the evolution of a class of technology can define a period in history and hold the potential to disrupt or alter the balance of power at societal and geostrategic levels. In this example, the evolution of scaling communications from printing press, to telegraph, telephone, the Internet, etc had broad and sustained impacts over decades. While all of the potential technology focus areas are part of our ongoing, broad technological advancement, some areas stand out as most pressing at this time for significant, eminent and sustained impact include: AI; high performance computing and semiconductors; quantum information science and technologies; robotics and advanced manufacturing; advanced materials and energy efficiency technologies.

Combinatorial effects of technologies. Advances in one technological area often, in turn, drive and accelerate technological development in other areas. As an example, advances in materials are crucial to advances in quantum technologies, and in turn, quantum simulation, which can drive advances in molecular design, drug discovery, and materials manufacturing. Likewise, advances in high performance computing can drive advances in AI, and in turn robotics and automation, and in turn, energy and industrial efficiency technologies. Such combinatorial effects could hold greater potential for addressing the societal, national, and geostrategic challenges prioritized.

6. Crosscutting investments

The issue of privacy is a crosscutting area where technologies are being developed that are ripe for transfer into a variety of application domains. The foundational approach of privacy focuses on the encryption of information, the proper authentication of those with authorized access to information, and issues around differential privacy. A more contextual approach of privacy includes focus on not only the information itself but the flow of that information and on proving certain properties of a collection of data without disclosing anything more about the data than the truth of those specific properties. Applications range from accounting/finance (proof that a set of transactions comply with a stated set of sanctions while keeping the transactions themselves secret) to logistical (proof that all components came from validated suppliers, but with the details of who supplied what kept secret). New cryptographic systems, many based on the concept of *zero-knowledge proofs,* have the potential of automating processes that now require trusted intermediaries. An emerging application domain merges this technology with AI and ML in systems where model outputs can be shown to come from inputs that meet certain properties while retaining the privacy of the inputs themselves. This can help ensure privacy in LLM applications in health and also enable privacy-preserving explainable AI.

Digital assets form a crosscutting area in which physical assets in the "real world" are represented by a digital token. Some "authority" provides an enforceable tie between the real-word asset and the token, and then, with that guarantee in hand, the asset can be sold, rented, used, etc. entirely in the digital domain without any need to interact with the non-digital enforcement authority except for resolution of real-world (that is non-digital) disputes.

Because advances in generative AI have improved to the point that LLMs can generate program code and invoke external software capabilities via plugin frameworks, it is increasingly important to be able to apply standards of privacy, security, fairness, and robustness to systems that include both AI (in the form of LLMs) and other traditional software components. Existing approaches to software robustness via specification, testing, and verification need to be revisited given the less predictable behaviors of LLMs and the lack of visibility about why the LLM has generated a particular output. Collaborative investments that allow the creation of more robust hybrid LLM/software infrastructure will accelerate the application of LLM technology in all areas where traditional software systems are currently being used.

7. Other Topics

Partner with small firms, especially startups and efforts more in “pre-startup” mode. There is a tendency for public-private partnerships to gravitate toward larger players in the technology space in question not only due to their size but also as a result of their ability to support a team with the broad-based skills needed to move the concept to deployment. This implicit bias in favor of size does have its merits, but it tends to reinforce the current consensus on "hot topics" and the best approaches to them. The TIP Directorate should look also at ways to partner with small firms, especially startups and efforts more in a "pre-startup" mode. This is particularly important in areas that are just emerging and in areas that have potential impact in disciplines that traditionally have had less interaction with computing and information research. In taking these steps, TIP needs to be cognisant of synergistic government-related activity in this domain, particularly those of the DoD (DIU) and the US Intelligence Community (In-Q-Tel). There is definite overlap among these missions. TIP should look to partnering with these agencies where appropriate, but also to seek out critical areas those agencies may be less likely to consider. The current In-Q-Tel portfolio of companies has the expected focus on areas relevant to military and intelligence infrastructure and domains of potential attack vectors. Looking at their coverage in computing, there is a strong focus on AI, microelectronics, and trusted computation. DIU has a stated mission "to identify and understand critical national security challenges that can be solved with leading edge commercial technology within 12 to 24 months." The DIU portfolio is largely in physical systems, but also includes AI, cyber security, and software testing and validation. In-Q-Tel looks mainly at existing startups and venture opportunities. DIU works with "top companies" but likewise has a strong interest in the startup and venture domain. TIP should give particular consideration to technologies of great national importance that are at least somewhat outside the purview of the defense and intelligence communities. Among those domains are financial/economic systems (such as digital assets, digital currencies, online commerce), information infrastructure (data science, social media, remote collaboration, internet search), and non-military applications in climate and the environment. In areas that cut across many applications, such as AI, TIP might usefully focus on broadly applicable infrastructure in contrast to possibly more focused applications in those other agencies.

Emerging technologies today do not always conform to traditional research frameworks of industrial labs and academic publications. Focused R&D groups may "publish" via papers on *Medium,* blogs, tweets, etc. This makes partnership opportunities harder to find, but no less important. One example of such a space is the software side of confidential computing, where much of the leading-edge research in zero-knowledge approaches to privacy with transparency and regulatability is happening in small startups. Despite NSF's strong academic heritage, TIP should seek to meet developers of novel technologies in *their own* R&D culture and not push for conformity to more traditional academic norms.

Small firms can be overwhelmed by an excessive embrace by the government, so it is essential to make the partnership right-sized and well-focused. One model is to create a team that can link the innovation of the startup to a first customer that becomes a partner in innovation. Bringing in a customer – not many customers, just one – at this early stage can help focus the innovators. To keep the innovators from being distracted by the non-innovative surrounding framework (e.g., in the case of a back-end system, this would be the web app, mobile app, UX), teams of experts in early application development and pre-sales could be supported and made available to help innovators over the initial hump of getting to market. Later, when the innovation has matured, the team providing that first-customer linkage would move on to another innovation.

Partnerships in computing R&D should be considered not only in the development of technology but also for the education of decision-makers. Leaders in government and industry are facing policy decisions on novel and unfamiliar technology. Uninformed or ill-informed decision makers are unlikely to make optimal choices. Educational partnerships targeting non-CS decision makers is an important pathway for technological impact. Partnership structure here would necessarily differ from the standard R&D partnerships. A possible worthwhile goal would be to seed the creation of interdisciplinary courses spanning an emerging area of computing and an application discipline (finance, supply chain, policy, etc.)

Be aware of potential success disasters. TIP needs to be cognizant of the potential for “success disasters”, including not just problems that arise if an idea becomes successful too quickly, but also if its success leads to unintended consequences. For example, an obvious concern would be to shift the balance of funding away from basic toward applied research leading to a shortage of future discoveries worthy of innovation (“eating one’s seed corn”). Another form of success disaster might be an effective TIP that leads to the failures of central planning that have been attributed to other forms of government.

Look for lessons learned internationally and at the local level. Europe has long been driving industry-academic partnerships that meld translation and research that potentially are similar in purpose to some of TIP's stated goals. The EU has also experience with mechanisms up to and including funding for corporate entities, formulaic requirements for multi-stakeholder/multi-region engagement, etc. To explore: if TIP succeeds in its current goals, will the outcomes be similar to Europe's outcomes, and are these desirable or not? Similarly, many US state governments have initiated programs that blend research with industry engagement and are meant to drive regional progress; have these worked out and what are the intended and unintended consequences? Comparative analysis seems worthwhile if only to highlight the factors that are worth monitoring going forward (e.g., scientific output, research translation successes, economic development).

Conclusion

We firmly believe that considering and incorporating broader perspectives, including those from industry, in all of TIP's future programs and solicitations will be crucial for effectively representing the computing research community and driving advancements in the field that benefit all stakeholders involved. CRA-I is thrilled to have had the opportunity to provide comment to this RFI and we look forward to watching TIP continue to impact the community as it grows.