CRA and CCC's Response to the Request for Information (RFI): Inviting Comments on the NIH Artificial Intelligence (AI) Strategy

Written by: Kevin Butler (University of Florida), Ovidiu Daescu (The University of Texas at Dallas), David Danks (University of California, San Diego), Catherine Gill (Computing Community Consortium), Ming Lin (University of Maryland), Brian Mosley (Computing Research Association), Manish Parashar (University of Utah), Mona Singh (Princeton University), Michela Taufer (University of Tennessee Knoxville), and Pamela Wisniewski (Florida International University).

This response is prepared by the Computing Research Association (CRA), assisted by the Computing Community Consortium (CCC). CRA is an association of over 270 North American computing research organizations, both academic and industrial, and partners from six professional computing societies.

CCC's mission, a CRA subcommittee, is to enable the pursuit of innovative, high-impact computing research that aligns with pressing national and global challenges.

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, or of the National Science Foundation, which funds the CCC.

The Computing Research Association (CRA) and Computing Community Consortium (CCC) appreciate the opportunity to provide input to the National Institutes of Health (NIH) regarding your Request for Information on developing an AI strategy. We commend NIH's proactive approach in exploring how artificial intelligence can enhance biomedical discovery and public health protection, and we hope our recommendations prove useful in developing an AI strategy.

Foundational Themes for an NIH AI Strategy

To effectively integrate AI, NIH's strategy should center on safe, secure and effective AI, prioritizing tangible outcomes and practical implementation. This approach inherently encompasses principles often associated with ethical AI, such as fairness and transparency, by focusing on real-world needs, stakeholder engagement, and rigorous testing.

Recommendations:

- **Emphasize Data Readiness:** NIH should invest in a comprehensive and compliant data infrastructure, a cataloging system, and discovery mechanisms that integrate diverse data formats, sources, and security levels. This includes developing clear standards and schemas for data reporting, collection, and archiving to ensure reproducibility and broad accessibility for researchers, even those without direct NIH funding. However, this level of broad accessibility should be balanced with considerations for privacy of patient data and compliance with HIPAA standards. The NIH should also support initiatives that create federated national infrastructures for NIH-related data to break down data silos and accelerate scientific inquiry through ease of access for researchers. Data should also be evaluated for potential biases, completeness, and whether the data was responsibly obtained to avoid harming individuals who may be underrepresented or misrepresented within an AI system’s data. Crucially, the NIH must prioritize assuring the security and privacy of training data and guarantee that trained models cannot leak private information, which could compromise user safety.

- **AI for Augmentation, Not Replacement:** NIH should unequivocally position AI as a tool to enhance human performance, not to substitute for human expertise or judgment. AI should not be used to replace clinical decision-making. These tasks require nuanced intellectual evaluation and human expertise. Rather, AI should be used to automate mechanical administrative tasks (e.g., checking grant submissions for formatting errors) to free up staff for more critical work, or to supplement human efforts in areas such as drug discovery (e.g., enumerating combinations), correlating vast datasets, and improving healthcare delivery scalability.

- **Focus on Reproducibility and Trust:** The strategy should strongly emphasize reproducibility and trust in AI-assisted research, meaning AI methods that are developed for healthcare contexts should be subjected to rigorous evaluation standards, akin to randomized controlled trials (RCTs) for medical interventions, to ensure their efficacy and safety before widespread deployment. Even

thoroughly tested models may function very differently or fail all together when being deployed in a new location due to differences in population demographics and data formats, so it is important to conduct rigorous testing at every new deployment to ensure algorithmic robustness (see CCC's Algorithmic Robustness whitepaper). During testing, experts in both AI and medicine should be involved to judge the efficacy of the models from an array of viewpoints.

High-Impact Use Cases for AI

AI holds immense promise for accelerating discovery and improving public health. However, a nuanced and interdisciplinary approach is essential to leverage its strengths while acknowledging its current limitations. Below, we list several use-cases for which AI solutions may accelerate research and operations at NIH:

- **Accelerating Biomedical Discovery:** AI can significantly advance scientific discovery by enumerating and exploring vast combinatorial possibilities in areas like drug discovery. This uses past data to reduce exponential possibilities, leading to more effective and efficient research. However, any discovery guided solely by past data risks limiting innovation and leading to a "saturation of knowledge" or biased research directions. AI should complement, not govern, the pursuit of new knowledge.
- **Enhancing Public Health Protection:** AI offers substantial potential for public health education, advocacy, and protection. These may include improving public health messaging and advocacy and enhancing the scalability of healthcare delivery, especially in rural and underserved areas where access to medical professionals is limited. AI can provide vital personalized support, helping to guide patients on healthy practices, and offering complementary tools for diagnosis and care.
- **Supporting Clinical Decision-Making:** AI can become an amazing tool to aid patients in seeking medical expertise and help healthcare professionals make faster, more informed decisions. For example, AI could enable doctors to access insights from diverse medical decisions and outcomes across the country, providing crucial context for unique patient symptoms or lab results. However, human judgment remains paramount; AI should assist, not replace, clinical decision-making. Careful consideration and evaluation are needed when deploying models, as their effectiveness can vary significantly across different populations and demographics.

Implementing Test Beds

Test beds are crucial for validating AI's utility and ensuring responsible deployment.

Recommendations:

- **Apply Medical Standards:** When implementing AI test beds and pilots, NIH should apply rigorous medical standards, such as those used for randomized controlled trials. This will ensure robust evaluation and efficacy before wider rollouts, preventing haphazard implementation driven solely by cost-cutting goals. AI systems should also be required to uphold the same ethical standards as US medical practitioners, to ensure that AI will not deliver recommendations in scenarios where a hallucination or misleading recommendation may risk lives.

- **Focus on Use-Inspired Cases:** Test beds should focus on use-inspired applications that address genuine needs within biomedical research and public health, demonstrating clear and measurable benefits. AI researchers should proactively work with medical practitioners, patients, caregivers, and other stakeholders to determine problems that may benefit from AI solutions rather than creating AI solutions and trying to retroactively fit them to the needs of stakeholders.

- **Identify Negative Use Cases:** In addition to identifying appropriate applications for AI, it is extremely important to identify use cases where implementing AI would be inappropriate, or even dangerous. For instance, AI should not be deployed in situations where the consequences of an error are catastrophic and human oversight or intervention is not reliably possible. Such a safeguard requires establishing clear boundaries for AI autonomy, especially in sensitive areas like medical diagnosis or treatment planning, where the potential for hallucinations, biases, or misinterpretations by an AI system could have life-threatening implications. The NIH should lead efforts to develop guidelines that define these "no-go" zones for AI, perhaps by creating a framework for assessing inherent risks in specific applications. This framework could consider factors like the level of interpretability required, the potential for algorithmic bias to exacerbate existing health inequities, and the irreversibility of AI-driven decisions. By clearly defining these negative use cases, we can ensure AI is deployed responsibly and ethically, safeguarding patient well-being and maintaining trust in technological advancements.

Workforce Development and Organizational Structure

Effective AI integration requires significant investment in workforce development and a thoughtful organizational approach.

Recommendations:

- **Study Anticipated Effects of AI System Deployments on the Medical Workforce:** When the U.S. mandated Electronic Health Records (EHR) systems, it changed the landscape of medical care. Physicians were suddenly dedicating up to 2 hours of work using the EHR systems for every hour spent delivering patient care, according to a study published in the Mayo Clinic Proceedings. The poor usability of this system, combined with the number of hours physicians were required to use it, led to a substantial increase in professional dissatisfaction and burnout among physicians, a pattern that contributed to the US's physician shortage and access-to-care problems. Beyond testing AI systems for usability and accuracy, it is imperative that the NIH test these systems' impact on healthcare professionals to ensure that these systems do not also contribute to physician burnout, further exacerbating the nation's doctor shortage. The NIH should also study anticipated effects related to job creation and job loss. What new jobs will be created in the healthcare space when AI systems are implemented? Which jobs will become redundant or will need to be restructured? These questions should be answered before AI systems are widely implemented.
- **Dedicated Computational Focus:** Consider establishing a dedicated institute or broader study sections for computational biomedical science. This would provide a home for grants focusing on advanced computing, data security, high-performance computing, and AI methodologies applied to biomedical problems. This structure ensures specialized review and support for highly technical proposals.
- **Embrace a Socio-Technical Approach:** Recognize that advancing AI requires integrating technical solutions with an understanding of human behavior and societal implications. This means bringing sociologists and humanities scholars into the development and deployment process, not just from a technical perspective. In addition to experts, patients, patient family members, caregivers, and other key stakeholders should be involved in providing feedback about these systems during the development and deployment process, and after, as the systems continue to be evaluated on a regular basis.

Interagency Collaboration

Collaborating with other federal agencies can streamline expertise sharing and program development.

Recommendations:

- **Expand Cross-Agency Partnerships:** Actively explore and expand ongoing collaborations with agencies like the National Science Foundation (NSF). Programs such as Smart Health and Wellbeing, which have already demonstrated success, serve as excellent models. This fosters knowledge migration and efficient resource utilization, leveraging existing pools of expertise rather than duplicating efforts.
- **Maintain Distinct Mission Focus:** While encouraging collaboration, it's crucial to maintain the distinct mission of NIH. NSF, for example, often prioritizes novel techniques to push research boundaries, while NIH emphasizes reliability and direct applicability to medical domains. Expanding existing, successful cross-agency programs is preferable to broadly embedding AI experts across all panels, which could blur critical distinctions and potentially hinder progress if less reliable, cutting-edge methods are prioritized over proven approaches for medical applications.

Balanced AI Portfolio

NIH's AI strategy must contribute to a balanced research ecosystem, not dominate it.

Recommendations:

- **Diversified Funding:** Ensure the AI strategy explicitly accounts for sufficient funding and resources for non-computational and non-AI research. Overemphasis on AI could inadvertently hinder innovation and new findings in other critical areas of biomedical science. A balanced portfolio ensures continued progress across the entire spectrum of NIH's mission.

We believe that by adopting these recommendations, NIH can develop a robust and effective AI strategy that advances biomedical research, improves public health, and enhances the agency's operations.