AMIA Response to NIH RFI: Inviting Comments on the Future of NLM's Biomedical and Data Science Extramural Research Programs

Informatics domains such as **bioinformatics, translational bioinformatics, clinical informatics, public health informatics, and consumer health informatics** are at the forefront of innovation that enable precision medicine, early disease detection, and scalable public health solutions. However, unlocking their full potential requires a clear, strategic vision that aligns emerging research with long-term impact. In response to the world's rapid advancements in these areas, NLM is embarking on a transformative journey to define the next phase of priorities for its Extramural Programs that will catalyze groundbreaking research and foster interdisciplinary collaboration to address the most pressing challenges in public health. Central to this endeavor is a robust **science vision** that reflects emerging opportunities in informatics and data science and advances the impact of new tools such as artificial intelligence (AI).

This RFI invites input from interested communities to inform the NLM science vision, ensuring that its funding priorities are forward-looking, impactful, and positioned to accelerate/support groundbreaking discoveries in biomedicine and health.

1. Vision for the Future

A forward-looking science vision requires anticipating emerging trends, critical challenges, and transformative opportunities. Responses in this section will help define baseline knowledge, pinpoint unmet needs and identify foundational advancements needed to accelerate progress. These insights will shape the foundation of the science vision by aligning future investments with unmet needs and emerging opportunities.

Input Requested:

- Significant unmet needs in biomedical informatics and data science, particularly those that require innovative, cross-domain solutions to drive transformative advancements across multiple areas of research.
- Key challenges in biomedical informatics and data science that necessitate groundbreaking solutions, especially those that bridge multiple domains and have broad transformative implications for research and practice.
- Foundational tools and methods that can be developed to support broad, scalable applications and enable next generation research.

Please provide a response on areas of future investment that NLM should consider.

AMIA RESPONSE to #1:

Significant Unmet Needs: Multi-modal data integration at population scale requires frameworks spanning molecular biology, epidemiology, and social sciences to handle genomics, proteomics, environmental exposures, social determinants, and clinical data while preserving privacy and ensuring equity. Current approaches fail to address temporal dynamics, missing data patterns, and heterogeneous data types simultaneously. Temporal health modeling across lifecycles demands collaboration between pediatrics, geriatrics, genetics, behavioral sciences, and health economics to create predictive models accounting for critical periods, cumulative exposures, and life transitions from preconception through end-of-life. Precision prevention at community scale necessitates integrating genomics, environmental science, sociology, urban planning, and public health for individually tailored yet community-contextualized interventions.

Key Challenges: Causal inference in complex health systems must move beyond correlation using novel methods leveraging natural experiments, instrumental variables, and counterfactual reasoning across epidemiology, economics, and clinical research, particularly for long-term outcomes and rare diseases where randomized trials are infeasible. Real-time adaptive learning in clinical environments requires bridging machine learning, clinical workflows, human factors engineering, and regulatory science to develop safely adapting AI systems maintaining performance while ensuring patient safety. Computational equity and bias mitigation demands sociotechnical approaches integrating computer science, biostatistics, medical anthropology, and community engagement to detect, measure, and mitigate algorithmic bias across the AI lifecycle for equitable outcomes.

Foundational Tools: Universal health data commons architecture should enable federated, privacy-preserving queries across institutions using homomorphic encryption, differential privacy, and secure multi-party computation for rare disease research and population health studies. Explainable AI for multi-stakeholder healthcare must provide contextually appropriate explanations for patients, clinicians, researchers, and regulators through cognitive science, human-computer interaction, and regulatory science integration. Synthetic health data generation platforms should create realistic, privacy-preserving datasets preserving complex relationships, temporal dependencies, rare events, and demographic diversity while enabling collaborative research without privacy risks.

2. Strategic Focus Areas
NLM aims to identify high-impact research areas with the greatest potential for impact while addressing gaps in current funding priorities. This section includes transformative opportunities in AI, data science infrastructure, interdisciplinary collaboration and emerging scientific challenges that require strategic investment.

Input Requested:

- Underexplored areas or specific gaps in NLM's current funding investments and how addressing these research gaps can lead to transformative impact or advancements.
- How existing biomedical informatics categories (i.e., bioinformatics, clinical informatics, translational bioinformatics, public health informatics, and personal health informatics) can evolve to align with contemporary challenges or evolve to address emerging scientific and technological challenges.
- Emerging or underexplored areas in data science and biomedical informatics that have the greatest potential to drive transformative advancements and should be prioritized for future research investments.
- Emerging AI technologies that hold the most promise for advancing biomedical discovery, clinical decision-making, or public health interventions.
- New methodologies that are needed and key challenges to ensure AI systems are reliable, generalizable, secure and scalable across a variety of datasets, populations, and environments.
- How AI can be leveraged to enhance the integration of complex, multi-modal data and address scalability challenges across research domains.
- The most significant cross-domain challenges where AI could have a transformative impact, and what foundational advancements are needed to address them.
- Critical challenges in data access, integration, and representation that hinder research advancements in data science and biomedical informatics.
- Strategies and or frameworks that NLM can implement to enable scalable and innovative data sharing while addressing privacy, security and accessibility concerns.
- Infrastructure, tools or frameworks that are needed to support interdisciplinary and multi-institutional collaboration in biomedical informatics research.

Please provide a response on strategic focus areas that NLM should consider.

AMIA RESPONSE TO # 2:

Computational Health Equity Research: While NLM funds health disparities research, there's insufficient investment in computational methods specifically designed to identify, measure, and mitigate algorithmic bias in healthcare AI systems. Current approaches treat bias as post-hoc consideration rather than designing equity into algorithms from inception. Addressing this gap requires developing bias-aware machine learning frameworks, fairness

metrics for healthcare contexts, and validation methods ensuring equitable performance across diverse populations. This could transform AI development in medicine, making equity a core design principle.

Environmental Health Informatics: Integration of environmental exposures, climate data, and social determinants with clinical and genomic data remains underfunded despite mounting evidence of environmental health impacts. NLM should invest in frameworks combining satellite imagery, air quality sensors, built environment data, and real-time climate information with longitudinal health records. This could revolutionize understanding of gene-environment interactions and enable precision prevention strategies accounting for location-specific health risks.

Narrative Medicine and Computational Linguistics: Patient narratives, clinical notes, and qualitative health data contain rich information that current NLP approaches inadequately capture. Investment in advanced discourse analysis, sentiment analysis, and narrative structure understanding could unlock insights from unstructured clinical text, patient-reported experiences, and social media health discussions, transforming patient-centered care by providing computational tools to understand patient perspectives and quality of life.

Evolution of Traditional Categories: Clinical informatics must evolve beyond EHR optimization to encompass ambient intelligence, IoT integration, and predictive healthcare delivery. Bioinformatics should expand from sequence analysis to multi-omics integration and dynamic biological system modeling. Translational bioinformatics must include reverse translation using real-world clinical data to inform basic research. Public health informatics should embrace predictive modeling, social media monitoring, and community-engaged digital interventions moving beyond traditional surveillance.

3. Additional Comments:

To ensure no critical areas are overlooked, we invite respondents to provide additional insights that may not fit within the categories above but are critical to advance the fields of biomedical informatics and data science.

Please provide any additional insights that NLM should consider.

AMIA RESPONSE TO #3:

Regulatory Science and AI Governance: The rapid advancement of AI in healthcare outpaces current regulatory frameworks. NLM should invest in research developing evidence-based guidelines for AI validation, approval processes, and post-market surveillance. This includes studying optimal regulatory pathways for different AI systems, developing standardized evaluation metrics, and creating frameworks for continuous monitoring of AI performance in real-world settings. Such research would inform FDA and other regulatory bodies while ensuring safe AI deployment and public trust in healthcare AI systems.

Foundation Models for Healthcare: NLM should invest in developing and evaluating healthcare-specific foundation models that can be fine-tuned for diverse clinical tasks, integrating text, imaging, genomics, and physiological data. Unlike commercial models, these should be openly available and designed with healthcare-specific requirements for safety, privacy, and regulatory compliance. This represents a unique opportunity for NLM to establish public infrastructure that democratizes access to advanced AI capabilities while maintaining scientific rigor and ethical standards.

Workforce Development Crisis: The field faces a critical shortage of professionals bridging clinical expertise with advanced computational skills. NLM should expand support for dual-degree programs (MD/PhD, RN/PhD, PharmD/PhD in informatics) and create pathways for practicing clinicians to gain informatics expertise. Following AMIA's commitment to diversity, funding should explicitly address underrepresentation through targeted training programs, mentorship initiatives, and partnerships with minority-serving institutions.

Open Science Infrastructure: Consistent with AMIA's advocacy for FAIR data principles, NLM should require all funded research to publish datasets with metadata, release software under open-source licenses, and provide reproducible analysis pipelines. Additionally, NLM should support the development of cloud-native research platforms that democratize computational access for smaller institutions and international collaborators, establishing sustainable models for long-term data stewardship beyond typical grant periods.