PUBLIC SUBMISSION

Received: May 29, 2025
Tracking No. mb9-jaq3-waql
Comments Due: May 28, 2025 **Submission Type:** API

Docket: NSF-2025-OGC-0001
NITRD_FRDOC_0001

Comment On: NSF-2025-OGC-0001-0001
Request for Information: Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

Document: NSF-2025-OGC-0001-DRAFT-0190
Comment on FR Doc # 2025-07332

Submitter Information

Organization: Penn State University

General Comment

Please See attached file(s)

Attachments

Penn State Response to RFI National AI RD Strategic Plan

Securing the AI Frontier: Driving Innovation with Responsibility

AI Hub, Pennsylvania State University
Mehrdad Mahdavi and Vasant Honavar

Response to RFI Docket ID No. NSF-2025-OGC-0001 – National AI R&D

As AI capabilities evolve at an unprecedented pace, the field of AI research is undergoing a profound transformation—reshaping its core questions, methods, communities, and research environments. This comment outlines key initiatives in foundational and translational research, education, and infrastructure needed to ensure continued U.S. leadership in AI. It emphasizes both short- and long-term efforts to advance AI technologies, support their transformative applications, and cultivate a cadre of AI-savvy scientists, engineers, and professionals across all domains.

A. Shaping the Future of AI: Toward Responsible and Sustainable Intelligence

A. 1 Unprecedented Progress

Recent developments in artificial intelligence (AI) underscore both remarkable progress and the urgent need to address emerging challenges. The field is advancing rapidly in performance, integration into applications, and accessibility, reinforcing AI's central role in shaping the future of technology, science, and society.

Since the 2023 release of the *National AI R&D Strategic Plan*, AI systems have made significant strides—surpassing human performance in many tasks and achieving unprecedented capabilities in video generation and multimodal understanding. These technical breakthroughs have led to widespread real-world adoption. AI has moved beyond the lab, with growing regulatory approvals and commercial deployments: the U.S. FDA has cleared over 200 AI-enabled medical devices, and autonomous vehicles operate at scale in both the U.S. and China.

Private investment in AI reached record levels, with the U.S. attracting over $100 billion in 2024, including nearly $34 billion in generative AI alone. Business adoption of AI has also accelerated, with most organizations reporting active use. Evidence suggests that AI contributes to workforce productivity and may even help bridge skill gaps in some domains. While the United States remains the global leader in developing cutting-edge models, countries such as China have rapidly closed performance gaps, contributing to an increasingly globalized AI ecosystem that includes increasing activity in the Middle East, Latin America, and Southeast Asia.

Despite these gains, progress in the regulation of, and guidelines and best practices for, responsible development and deployment of AI remains slow and uneven. Incidents involving unsafe or unreliable AI outputs are on the rise, yet few developers adopt standardized evaluation frameworks. Nonetheless, new benchmarks are advancing the field's capacity to assess their validity and safety. Governments are responding with stronger regulations and significant public investment, recognizing that innovation must be matched by robust governance.

Public sentiment toward AI is improving worldwide, though regional differences persist. Optimism is highest in Asia and the Global South, while skepticism remains strong in North America and parts of Europe. At the same time, AI is becoming more efficient and accessible thanks to advances in smaller, more capable models, open-weight alternatives, and better hardware—reducing both costs and energy demands.

Educational initiatives are expanding, with more countries integrating computer science and AI into K–12 curricula. However, progress remains uneven due to disparities in infrastructure and educator preparedness—particularly in under-resourced regions.

At the cutting edge of AI, industry now dominates, producing nearly 90% of state-of-the-art models. Nonetheless, academia remains a key driver of foundational research. Competition among top-tier models is intensifying, with challengers rapidly closing the gap with established leaders.

AI's contributions to science have received global recognition, including Nobel Prizes and Turing Awards for foundational breakthroughs. Yet critical limitations persist. While current models excel at standardized benchmarks, they continue to struggle with complex reasoning and logic—capabilities essential for safe deployment in high-stakes domains. This enduring gap highlights the urgent need for sustained investment in foundational research and the development of robust reasoning frameworks.

A. 2 Emerging Themes and a Vision for the Future of AI

As artificial intelligence rapidly evolves from laboratory prototypes to real-world systems, the need for a cohesive and forward-thinking research agenda has never been more urgent. The *National AI R&D Strategic Plan in 2023* laid critical groundwork for advancing AI in ways that are ethical, equitable, and effective. As artificial intelligence systems grow in capability, complexity, and societal impact, a bold and integrated research vision is needed to guide the next era of discovery, development, and deployment. The following strategic themes outline a framework for advancing AI in ways that are scientifically rigorous, societally beneficial, and deeply human-centered. We identify nine interlinked research themes that together define a forward-looking vision for accelerating progress in foundational and translational research while safeguarding societal interests.

1. **Intelligent Agents and Mult-Agent Systems.** We envision future AI systems that seamlessly integrate aspects of intelligence that have traditionally been studied in isolation—perception, learning, reasoning, decision-making, action, interaction, coordination and communication. Achieving such unification is essential for building general-purpose intelligent agents and multi-agent systems capable of operating reliably in dynamic, real-world environments. This will require new architectures, representations, and learning paradigms that support end-to-end adaptability, transparency, explainability, safety, and continual improvement.
2. **Embodied AI.** True intelligence is grounded in physical experience. Advances in embodied AI—where agents learn, act, and adapt through interaction with the physical world—will be central to progress. Research in this area should bridge robotics, neuromotor control, sensorimotor learning, and simulation environments to set and achieve their objectives safely in uncertain dynamic environments.
3. **AI models of Animal and Human Intelligence.** AI offers a powerful lens through which to study natural intelligence. Models that replicate and help explain cognitive, emotional, and social aspects of human and animal behavior can drive breakthroughs in AI. This bidirectional research agenda—using AI to understand brains and using insights from biology to inform AI—will help uncover the principles underlying general intelligence and adaptive behavior.

4. **Human-centered AI.** As AI becomes deeply embedded in human lives, its design must prioritize human values, needs, and experiences. Human-centered AI research must address accountability, transparency, interpretability, and trust. This includes advancing user-aligned objectives, participatory design methods, adaptive interfaces, and safeguards against misuse. Human-AI collaboration must be engineered for augmentation—not replacement—of human capabilities.
5. **AI and Society.** AI systems influence the fabric of society—from labor markets to political discourse to education and healthcare. A responsible AI future requires cross-disciplinary inquiry into AI's societal impact, including ethical, legal, economic, and cultural dimensions. Research must explore mechanisms for democratic governance of AI technologies, equitable access, and proactive mitigation of systemic biases and harms.
6. **AI for Science.** AI is becoming an indispensable partner in scientific discovery. Next-generation AI systems must be designed to augment scientific reasoning, automate hypothesis generation and testing, interpret multimodal data, and design experiments in fields ranging from physics to biology to material science and agriculture. Investment in AI as a tool for accelerating the pace and breadth of scientific breakthroughs will have transformative impact across disciplines.
7. **AI for Humanities and Creative Arts.** Artificial Intelligence offers powerful new tools for exploring, interpreting, and expanding the boundaries of human creativity and cultural understanding. In the humanities, AI can analyze vast archives, uncover hidden patterns in language, art, and history, and support new modes of scholarly inquiry. In the creative arts, AI can act as collaborator, catalyst, and medium—augmenting human expression across music, literature, visual art, performance, and design.
8. **Translational Research in AI.** To maximize the societal benefits of AI research, strong pipelines are needed to translate fundamental advances into deployable, impactful technologies. This includes support for testbeds, open platforms, standards, regulatory pathways, and public-private partnerships that bridge academia, industry, and government. Translational research must ensure that AI innovations are robust, safe, and aligned with public interest.
9. **Core Computer Science Research for AI.** Core computer science research is fundamental to advancing the frontiers of AI. Breakthroughs in algorithms, data structures, programming languages, formal methods, computer architecture, distributed systems, and computational complexity provide the theoretical and practical foundations upon which intelligent systems are built. From enabling scalable learning architectures to ensuring provable guarantees of correctness, robustness, and efficiency, core CS research drives innovation in AI capabilities, their sustainable development and deployment, and safeguards their reliability.

Together, these focus areas represent a bold, multidisciplinary agenda for AI that is not only powerful and intelligent—but also responsible, resilient, and aligned with human values.

- **Intelligent Agents and Multi-Agent Systems**

We envision a new generation of AI systems composed of intelligent agents and multi-agent collectives that seamlessly integrate core facets of intelligence—including perception, learning, reasoning, decision-making, action, interaction, coordination, and communication. Historically, these capabilities have often been studied in isolation. However, achieving robust, general-purpose intelligence requires a unified approach that enables agents to understand their environments, adapt to new information, collaborate with others, and act autonomously in dynamic, unpredictable settings.

Future intelligent agents must go beyond static models or task-specific behaviors. They should continuously learn from experience, reason under uncertainty, align with human values, and make context-sensitive decisions in real time. Multi-agent systems add further complexity and opportunity: agents must cooperate or compete, negotiate shared goals, and coordinate distributed activities—whether in digital marketplaces, autonomous vehicle fleets, or planetary exploration.

Realizing this vision demands foundational advances in AI architectures, knowledge representations, learning frameworks, and communication protocols. It also calls for systems that are transparent, interpretable, and verifiably safe—particularly when deployed in high-stakes environments. Research must address how agents can build models of themselves and others, anticipate the behavior of teammates or adversaries, and resolve conflicts in complex social and organizational contexts.

This line of inquiry will not only push the boundaries of artificial intelligence but also illuminate deeper questions about autonomy, collective intelligence, and the principles that govern individual as well as collective intelligent behavior.

- **Embodied AI**

True intelligence is fundamentally grounded in physical experience. For artificial agents to exhibit robust, adaptive, and context-aware behavior, they must engage with the world not merely as observers but as active participants. Embodied AI represents a paradigm shift from disembodied models trained solely on static data to agents that learn, act, and adapt through real-time interaction with dynamic, often unpredictable environments.

At the heart of embodied AI is the integration of perception, motor control, learning, and decision-making in a closed feedback loop with the physical world. Whether navigating cluttered rooms, manipulating objects with dexterous hands, or collaborating with humans in shared spaces, embodied agents must continuously interpret noisy sensory data, infer intentions, plan actions, and respond to changing circumstances—all while learning from experience and adapting over time.

Advancing embodied AI requires breakthroughs across several foundational areas. Robotics provides the physical substrate, including hardware for mobility, manipulation, and sensing. Neuromotor control and biomechanics offer inspiration for low-level coordination and real-time responsiveness. Sensorimotor learning enables agents to

build internal models of their bodies and environments, allowing for generalization, anticipation, and transfer of learned behaviors. High-fidelity simulation environments serve as essential tools for safe, scalable, and accelerated experimentation, enabling agents to acquire foundational skills before deployment in the real world.

Embodied AI research also has broader implications for our understanding of intelligence. By grounding cognition in action and physicality, it provides insights into the developmental processes by which biological agents—infants, animals, and humans—acquire knowledge, social skills, and problem-solving abilities. It opens the door to more intuitive, interactive, and assistive AI systems that can collaborate with humans in physical spaces, from homes and hospitals to factories and disaster zones.

To realize the promise of embodied AI, we must foster interdisciplinary collaboration across AI, robotics, neuroscience, biomechanics, cognitive science, and human-computer interaction. Building embodied agents that are safe, efficient, and aligned with human values will be a cornerstone of the next frontier in artificial intelligence—one that brings us closer to creating machines that can truly understand and act in the world as humans and animals do.

- **AI Models of Animal and Human Intelligence**

Artificial Intelligence offers a transformative lens for studying the nature of intelligence itself. By developing computational models that replicate and help explain cognitive, emotional, and social aspects of animal and human behavior, AI can serve both as a scientific tool and a conceptual bridge between disciplines. This bidirectional research agenda—using AI to understand natural intelligence and using biological insights to inspire artificial systems—holds extraordinary promise for advancing our understanding of minds, both natural and synthetic.

In the study of human and animal cognition, AI models can simulate learning, perception, memory, decision-making, language, and even emotional regulation. These models allow researchers to test hypotheses about brain function, behavioral adaptation, and developmental learning in controlled, scalable ways that complement empirical studies. From modeling neural circuits involved in vision and motor control to exploring how agents acquire theory of mind or social norms, AI can help illuminate the mechanisms underlying intelligence and adaptive behavior.

Conversely, biology and neuroscience provide deep inspiration for AI. The architectures, learning strategies, and representational efficiencies evolved in animals offer powerful blueprints for more generalizable, robust, and energy-efficient AI systems. For example, insights into hippocampal navigation, reinforcement learning in dopamine systems, or hierarchical planning in the prefrontal cortex are already informing the design of advanced AI models. By bridging cognitive science, ethology, and machine learning, we can develop AI that is not only more intelligent but also more

aligned with the way humans and other organisms understand and interact with the world.

This line of inquiry also opens up new questions about the nature of consciousness, creativity, empathy, and autonomy—questions that are central to both the future of AI and the enduring philosophical quest to understand the mind. Progress will depend on tight collaboration across disciplines: neuroscience, psychology, computational modeling, robotics, and philosophy, among others.

- **Human-Centered AI**

As AI systems become increasingly pervasive in everyday life—shaping decisions in healthcare, education, finance, employment, transportation, and beyond—it is critical that their design and deployment place human values, needs, and well-being at the forefront. Human-centered AI (HCAI) seeks to ensure that technological advancement enhances human agency, dignity, and autonomy, rather than undermining them. This approach views AI not merely as a tool to optimize efficiency or scale decisions, but as a partner that must align with human goals and support the diversity of human experience.

Key to this vision is a robust research agenda focused on accountability, transparency, interpretability, and trust. Users must be able to understand, question, and—when necessary—challenge the decisions made or recommended by AI systems. This requires technical advances in explainable AI, value alignment, and algorithmic auditing, as well as legal and institutional mechanisms to ensure accountability. Trust must be earned, not assumed, and sustained through transparent design, consistent performance, and responsiveness to user feedback.

Human-centered AI also requires participatory and inclusive design methods. Systems must be co-designed with the people they affect, particularly those from historically marginalized or underrepresented communities. By incorporating human perspectives early and throughout the design process, researchers and developers can uncover hidden assumptions, prevent harm, and create AI systems that are not only functional but also equitable, usable, and respectful of cultural and contextual differences.

In addition, future AI systems should be engineered to augment, rather than replace, human capabilities. Human-AI collaboration must leverage the strengths of both—machines' ability to process vast information and detect patterns, and humans' judgment, empathy, creativity, and moral reasoning. Whether assisting doctors in diagnosis, supporting educators in personalized learning, or helping workers navigate complex tasks, AI should amplify human potential, not diminish it.

Ultimately, human-centered AI demands interdisciplinary collaboration across computer science, human-computer interaction, psychology, ethics, law, and the social sciences. It also requires a broader cultural shift in how we think about technology—not as

something that acts on us, but as something we shape deliberately, collectively, and ethically in service of human flourishing.

- **AI and Society**

AI does not function in a vacuum —it is rapidly becoming embedded in the core institutions and social systems that shape everyday life. From transforming labor markets and automating public services to influencing political discourse and restructuring education and healthcare, AI technologies are reshaping the fabric of society. As these systems increasingly mediate access to resources, opportunities, and information, their design and deployment raise profound ethical, legal, economic, and cultural questions that demand urgent attention.

A responsible AI future cannot be driven by technical innovation alone. It requires a broad, cross-disciplinary research agenda that critically examines the societal impact of AI. This includes understanding how algorithmic systems reinforce or mitigate existing inequalities, how AI influences decision-making in high-stakes domains, and how it affects public trust, democratic institutions, and civic life. Research must move beyond retrospective harm analysis to develop anticipatory frameworks that proactively identify and address risks—especially those affecting vulnerable or historically marginalized populations.

Democratic governance of AI is central to ensuring that these technologies serve the public interest. Mechanisms such as participatory policymaking, algorithmic transparency, community oversight, and public impact assessments can help ensure that the deployment of AI is aligned with democratic values, including accountability, fairness, and inclusivity. At the same time, regulatory and policy frameworks must keep pace with technological developments, balancing innovation with protections against misuse and abuse.

Equitable access to the benefits of AI is another critical challenge. Without deliberate action, the advantages of AI risk being concentrated in the hands of a few, exacerbating global disparities in wealth, power, and opportunity. Research must explore models for open infrastructure, inclusive innovation ecosystems, and global cooperation to ensure that the economic and social gains from AI are broadly shared.

Ultimately, the intersection of AI and society calls for a new social contract—one that recognizes the transformative potential of intelligent systems but insists on embedding them within structures of human responsibility, institutional accountability, and moral reflection. It is only through such a holistic, inclusive, and values-driven approach that AI can genuinely serve as a tool for societal progress.

- **AI for Science**

Artificial Intelligence is emerging as a transformative force in the practice of science. No longer limited to automating routine tasks, AI is becoming an active partner in the

scientific process—augmenting human reasoning, accelerating discovery, and opening new frontiers of knowledge. From predicting protein structures to simulating climate systems, AI is already reshaping how research is conducted across a wide range of disciplines, including physics, biology, chemistry, material science, agriculture, and environmental science.

To fully realize its potential, the next generation of AI systems must go beyond pattern recognition. They should be designed to support and enhance core elements of scientific inquiry—formulating hypotheses, designing experiments, interpreting complex, multimodal datasets, and even uncovering previously unrecognized phenomena. These systems must be capable of operating under uncertainty, integrating data from diverse sources, and adapting to evolving lines of inquiry. In doing so, AI can help scientists navigate the ever-growing volume and complexity of scientific data that would otherwise be overwhelming.

Crucially, AI can also help democratize access to scientific expertise and tools. By enabling more researchers to leverage computational modeling, predictive analytics, and intelligent lab assistants, AI can reduce barriers to entry and foster innovation in under-resourced institutions and regions. In agriculture, for example, AI can aid in precision farming and crop resilience; in biomedical research, it can accelerate drug discovery and personalized medicine; and in climate science, it can improve models for forecasting and mitigation.

This vision requires sustained and strategic investment in research and infrastructure. It entails creating interoperable platforms, shared datasets, domain-specific AI models, and collaborative environments where scientists and AI systems can iteratively formulate and test hypotheses, design and conduct experiments, analyze and interpret data, and co-create new knowledge. Importantly, it also requires a new generation of researchers fluent in both the scientific and computational domains—hybrid thinkers capable of bridging the gap between disciplines.

Ultimately, investing in AI for science is not just about increasing the speed of discovery. It is about transforming the very way we do science—making it more integrative, exploratory, and capable of tackling complex, urgent challenges that traditional methods alone cannot address. In short, AI could offer the most powerful tools humans have ever had, cognitive analogs of telescopes and microscopes, for understanding the world around us.

- **AI for Humanities and the Creative Arts**

AI is opening up powerful new possibilities for the humanities and creative arts, enabling novel forms of inquiry, expression, and engagement with human culture. By combining computational power with human imagination, AI is helping scholars and artists alike reimagine what it means to analyze, interpret, and create. In doing so, it is not only expanding the boundaries of artistic practice and humanities research, but also

enriching our understanding of history, language, and identity in an increasingly digital world.

In the humanities, AI systems can process and analyze vast archives of texts, images, audio, and video—uncovering patterns and connections that might elude even the most seasoned scholars. Natural language processing tools can illuminate shifts in discourse over time, trace cultural influences across geographies, and assist in the preservation and analysis of endangered languages. Computer vision and machine learning can help art historians detect stylistic patterns or identify forgeries. These technologies enable new forms of “distant reading,” while also supporting deeper close analysis by augmenting human intuition with large-scale data insight.

In the creative arts, AI functions not as a replacement for human creativity, but as a collaborator, catalyst, and medium. Artists and designers are increasingly using AI to generate music, compose poetry, synthesize visual art, choreograph performances, and design immersive experiences. These tools allow creators to experiment with form, narrative, and interactivity in ways previously unimaginable—while raising new aesthetic, philosophical, and ethical questions about authorship, originality, and the nature of creative agency.

At the same time, AI invites a deeper exploration of creativity itself. By studying how AI generates art and interprets cultural content, researchers can gain insight into the cognitive and social processes underlying human creativity. This opens the door to interdisciplinary collaborations between technologists, artists, and humanists that are both critically reflective and forward-looking—helping shape AI tools that are not only technically advanced, but also culturally sensitive and ethically grounded.

To support this emerging frontier, there is a need to foster interdisciplinary research and education at the interface of the AI and the humanities and arts. Equally important is ensuring that the use of AI in these domains accommodates cultural diversity, artistic autonomy, and the pluralism of human expression. Infusion of AI into arts and the humanities can deepen cultural understanding, inspire artistic innovation, and broaden access to the humanities and arts for future generations.

- **Translational AI Research**

Translational research plays a critical role in ensuring that the rapid advances in artificial intelligence move beyond the laboratory and into the real world in ways that are impactful, responsible, and equitable. As AI technologies increasingly influence healthcare, education, manufacturing, agriculture, transportation, and public services, it is essential to establish robust pipelines that can transform foundational research into scalable, trustworthy applications that serve societal needs.

To achieve this, greater investment is needed in infrastructure that supports experimentation, validation, and deployment. This includes developing large-scale, interdisciplinary testbeds where AI systems can be evaluated in realistic, high-stakes

settings—such as autonomous driving, clinical diagnostics, and emergency response. Open platforms, shared datasets, and reproducible benchmarks are equally important to foster collaboration, interoperability, and innovation across research communities. These platforms must be designed to reflect diverse environments, users, and contexts, ensuring that AI technologies perform reliably outside of controlled laboratory conditions.

Standardization and regulatory frameworks are another essential component of effective translation. Establishing clear, evidence-based standards for safety, robustness, fairness, and transparency can help guide both innovation and oversight. Simultaneously, streamlined regulatory pathways can facilitate the ethical and timely adoption of AI technologies, especially in domains such as medicine and public infrastructure, where approval processes can otherwise create bottlenecks.

Public-private partnerships have a crucial role to play in bridging the gap between academic research, industrial development, and public-sector implementation. These partnerships can accelerate the scaling of emerging technologies, enable joint investment in high-risk, high-reward innovation, and ensure that the benefits of AI reach a broad range of sectors and communities. Government agencies can further catalyze translational research by supporting interdisciplinary grant programs, innovation hubs, and technology transfer initiatives that connect researchers with practitioners and end users.

Ultimately, the goal of translational research in AI is not just to deploy advanced systems, but to do so in a manner that is aligned with public values and interests. This requires a proactive focus on ethics, equity, and long-term societal impact throughout the innovation pipeline—from basic research to real-world deployment. By strengthening these pathways, we can unlock the full promise of AI as a force for public good, economic resilience, and inclusive technological progress.

- **Core Computer Science Research for Advancing AI**

Advancing AI requires sustained progress in the foundational areas of computer science that underpin the design, development, and deployment of intelligent systems. Research in algorithms, data structures, programming languages, formal verification, distributed and parallel computing, quantum computing, and computational theory forms the critical backbone of AI capabilities. These core areas provide the mathematical rigor, efficiency, and reliability needed to support increasingly complex and high-performance AI systems.

As AI models scale in size and complexity, algorithmic efficiency becomes paramount. Innovations in optimization, approximation, and parallel algorithms can drastically reduce training times, energy consumption, and cost. Research into new data structures supports more efficient data access, storage, and retrieval—key to enabling real-time AI

applications and managing massive, heterogeneous datasets. Similarly, advances in formal methods and verification are essential for building trustworthy AI systems with provable guarantees on safety, correctness, and robustness—especially in high-stakes domains such as healthcare, transportation, and defense.

Programming languages and compilers tailored for AI workloads are another critical frontier. Domain-specific languages and toolchains can streamline the development of reliable and performant AI systems, while enabling better interpretability and debuggability. At the same time, research in distributed and cloud-scale systems is vital for scaling AI across federated devices, edge networks, and large data centers.

Finally, theoretical computer science offers deep insights into the fundamental limits and possibilities of AI. Understanding the computational complexity of learning, reasoning, and planning tasks not only sharpens the design of practical systems but also helps identify where new paradigms or abstractions are needed.

To fully realize the potential of AI, the research community must prioritize investment in core computer science disciplines—not only as enablers of performance and scalability, but as essential safeguards for transparency, reliability, and societal alignment. Integrating core CS research more deeply into the AI research agenda will help build intelligent systems that are not only powerful, but principled and resilient.

A.3 Emerging Challenges

While the vision for advancing artificial intelligence across domains is ambitious and inspiring, realizing it entails navigating a complex landscape of scientific, technical, institutional, and societal challenges.

Realizing powerful intelligent agents and multi-agent systems present new design challenges in interpretability, coordination, human-AI interaction, and alignment. Multi-agent systems require negotiation, teamwork, goal-setting and planning, ethical reasoning, while avoiding unsafe, unreliable, and unintended behaviors. Future research must support the development of robust, transparent architectures for LLM-driven agents capable of safe, interpretable collaboration in complex environments.

One of the foremost challenges lies in integrating traditionally siloed areas of AI research. Despite growing recognition of the need to unify perception, learning, reasoning, and action, most current systems remain fragmented, optimized for narrow tasks, and brittle when transferred to real-world conditions. Building general-purpose agents, embodied AI systems, and models of natural intelligence will require breakthroughs in algorithmic design, training efficiency, and evaluation methodologies—many of which remain open research problems. Achieving trustworthy, explainable, and continually adaptive AI also demands new frameworks that can reconcile symbolic and statistical reasoning, support multi-modal learning, and operate safely in dynamic environments.

For AI to meaningfully contribute to science, it must move beyond regularity detection and predictive modeling to support hypothesis generation, causal inference, and experimental design—tasks that demand deep integration of reasoning and domain knowledge. Scientific data are often sparse, noisy, multi-modal, heterogeneous, and span multiple spatial and temporal scales. Many scientific problems also require interpretability, causal or mechanistic explanations, and rigorous uncertainty quantification—areas where current AI systems often fall short. Furthermore, integrating AI into scientific workflows demands cross-disciplinary collaboration and trust between domain experts and AI systems. Looking ahead, the future of AI for scientific discovery will hinge on developing models that can reason with limited data, incorporate physical laws and prior knowledge, and actively collaborate with human scientists. Advances in human-AI collaboration in formulating and testing hypothesis, designing and conducting experiments, constructing predictive, causal and mechanistic models, interpreting and explaining data and models offer the promise of systems that not only accelerate research across the entire research lifecycle but also help reimagine the very process of scientific inquiry.

Realizing embodied AI calls for cognitive architectures that enable agents to learn and adapt in real-time, under a variety of constraints. Incomplete knowledge, partial observability, and uncertainty. Yet existing AI models are often ill-suited for these challenges. Embodied and cognitive AI needs innovations in architectures that simulate and extend human learning capabilities, enabling agents to navigate physical and simulated environments with cognitive flexibility. Cross-disciplinary research integrating neuroscience, psychology, and robotics will be critical in advancing this goal.

Public perception of AI frequently diverges from its actual capabilities, leading to mistrust, misuse, and misinformation. Closing this gap calls for demystifying AI, its capabilities, its potential as well as its pitfalls, to empower all stakeholders to shape the future of AI. The escalating energy and resource demands of state-of-the-art AI systems raise concerns about sustainability and equitable access, especially in light of global environmental goals. Further advances in some of the resource-intensive AI technologies calls for innovations across all areas of computing.

To ensure that AI advances benefit the society at large AI research must involve all stakeholders from the outset, through participatory design and evaluation to tackle pressing societal challenges. Furthermore, realizing a human-centered and socially responsible AI ecosystem requires not only technical safeguards but also institutional reform and cross-disciplinary collaboration. Embedding values such as fairness, accountability, and transparency into AI systems demands ongoing engagement with ethicists, legal scholars, social scientists, and affected communities—many of whom have historically been excluded from AI development processes. Building the capacity for participatory design, anticipatory governance, and ethical oversight across sectors will be critical but is still underdeveloped.

Data, compute, and access disparities present another major barrier. Many of the most powerful AI systems today are developed and deployed by a small number of well-resourced industry labs. Academic researchers, public institutions, and underserved

communities often lack access to the massive datasets, high-performance compute infrastructure, and scalable deployment platforms needed to pursue frontier research or participate meaningfully in AI innovation. Without deliberate action, this asymmetry threatens to exacerbate global and regional inequalities in research capacity, technological adoption, and economic benefit.

Finally, translational and interdisciplinary collaboration—a key pillar of this vision—is often hindered by misaligned incentives, disciplinary silos, and fragmented funding streams. Academic research is frequently decoupled from real-world testing and deployment, while industrial priorities may emphasize short-term commercial goals over long-term societal value. Bridging this gap will require new funding models, shared testbeds, data trust frameworks, and agile regulatory pathways that promote collaboration while safeguarding public interest.

Overcoming these challenges will require bold leadership, sustained investment, and inclusive coordination across sectors. But doing so is essential to ensure that AI advances not only in capability, but in service of the public good—expanding knowledge, enriching culture, strengthening democracy, and improving lives around the world.

A.4 Systemic Challenges that Might Undermine AI's Trajectory

Despite unprecedented momentum in AI development and deployment, several systemic challenges now pose significant risks to sustained progress and public trust. These obstacles, spanning trust, equity, sustainability, and global coordination, must be addressed to ensure that the benefits of AI R&D are both enduring and widely shared:

- **Lack of AI Literacy:** A lack of AI literacy across the public, workforce, and decision-makers presents a systemic challenge to realizing the full promise of AI. Without a foundational understanding of how AI systems work, their limitations, and their societal implications, individuals are less equipped to engage with AI critically, use it responsibly, or make informed choices about its adoption. This gap undermines trust, fuels misinformation, and hampers efforts to develop inclusive, participatory governance. Bridging the AI literacy divide is essential to ensure equitable access to AI's benefits and to foster a society capable of shaping its development in the public interest.
- **Shortage of AI-Savvy Workforce:** The shortage of AI-savvy scientists, technologists, scholars, and professionals presents a critical systemic barrier to advancing the field and fully realizing its transformative potential. As AI becomes integral to sectors ranging from healthcare and education to national security, the demand for talent with deep technical expertise, interdisciplinary fluency, and ethical awareness far exceeds supply. This talent gap slows innovation, limits the diversity of perspectives shaping AI development, and constrains the ability of institutions to deploy AI responsibly and effectively. Addressing this challenge

requires sustained investment in education, training, and career pathways that cultivate a broad, inclusive, and well-prepared AI-capable workforce.

- **Suboptimal organization of academic AI research and Education Programs:** Institutional silos within academic departments and colleges pose significant systemic hurdles to advancing truly interdisciplinary AI research and education. Most innovative AI advances and AI's and its most pressing challenges—span fields such as computer science, information sciences, engineering, behavioral, cognitive and brain sciences, the social sciences, law, humanities, and the arts. AI's most impactful applications span virtually all areas of human endeavor from healthcare to education, agriculture, and national defense. Yet rigid departmental boundaries often impede collaboration, limit cross-training opportunities for students, and create fragmented funding and incentive structures. Overcoming these barriers requires rethinking organizational models, fostering cross-cutting programs, and creating institutional cultures that reward interdisciplinary inquiry, team science, and shared educational innovation.
- **Trust and safety deficits**: The lack of public trust in AI, compounded by the absence of strong safety guarantees, significantly hinders the realization of AI's full promise to improve human life and wellbeing. Without clear assurances that AI systems are reliable, transparent, and aligned with human values, users are less likely to adopt them—particularly in high-stakes domains such as healthcare, education, and public services. Moreover, real-world failures or harmful outcomes can erode confidence and stall innovation. Building trustworthy AI requires not only technical robustness and explainability, but also regulatory frameworks, oversight mechanisms, and a commitment to ethical design that prioritizes human dignity and safety.
- **Environmental costs**: The carbon footprint of model training is soaring. GPT-4 emitted over 5,000 tons of carbon emissions during training—orders of magnitude higher than early models. The growing environmental costs of AI—driven by the energy demands of training and deploying large-scale models—pose a significant barrier to realizing its full benefits. As AI systems become more powerful, their carbon footprint and resource consumption increase, raising concerns about sustainability, equity, and long-term scalability. These impacts can undermine public support, strain infrastructure, and exacerbate global environmental challenges. To ensure that AI contributes positively to human wellbeing, the field must prioritize energy-efficient architectures, green computing practices, and policy incentives that ensure that AI advances are realized in environmentally sustainable ways,
- **Access gaps**: Gaps in access to high-quality data, advanced computing infrastructure, and industry-scale state-of-the-art AI models—especially among small research labs, universities, and institutions in developing countries—create profound inequities that hinder the global realization of AI's potential. These disparities limit who can contribute to cutting-edge research, slow innovation, and risk concentrating AI capabilities in the hands of a few dominant actors. Without democratized access to essential AI resources, much of the world remains excluded from shaping or benefiting from AI advancements. Bridging these gaps

is essential to foster innovation, unlock creativity, and ensure that AI serves broad societal interests.

- **Global competition and fragmentation**: Intensifying global competition in AI talent, infrastructure, and applications poses a significant challenge to sustained U.S. leadership in the field. Countries around the world are investing heavily in national AI strategies, expanding research ecosystems, and attracting top talent through aggressive funding and immigration policies. China now leads in total AI publications and patents, and the quality gap with U.S. models is rapidly closing. New players from Latin America, Southeast Asia, and the Middle East are contributing to a competitive but fragmented landscape, raising questions about alignment and global normsAs a result, the U.S. faces growing pressure to retain its competitive edge amid a shifting geopolitical landscape. Without coordinated efforts to strengthen domestic AI education, research infrastructure, workforce development, and international collaboration, the U.S. risks falling behind in both innovation and influence—jeopardizing its ability to shape the future of AI in alignment with democratic values.
- **Regulatory lag**: The United States currently lags behind regions like Europe and parts of Asia in establishing comprehensive regulatory frameworks for socially responsible and ethical development and use of AI technologies in critical sectors. While the U.S. has made progress through agency-led initiatives, it lacks a cohesive national policy that provides clear standards for safety, transparency, accountability, and ethical use. In contrast, the European Union has advanced landmark legislation like the AI Act, and countries in Asia are actively developing governance structures to promote trust and public confidence. This regulatory gap creates uncertainty for developers and users, slows responsible innovation, and weakens international influence in setting global norms. Without timely, thoughtful regulation, the U.S. risks ceding leadership in both AI governance and market competitiveness to more proactive regions.

B. Strategic Directions for National AI R&D

As mentioned earlier, the AI landscape has transformed dramatically, characterized by advances in large-scale language models, emerging agentic systems, and growing societal reliance on AI technologies. To secure U.S. leadership in foundational, trustworthy, and ethically aligned AI innovation—while proactively navigating emerging risks and opportunities in a rapidly evolving global landscape—we identify the following strategic **directions.**

B.1 Long-term foundational AI research

Since 2023, there has been remarkable progress in large language models (LLMs), neuro-symbolic systems, and hybrid architectures that combine data-driven learning with formal reasoning. These models have demonstrated capabilities that were once considered science fiction. However, the lack of guarantees in their reasoning outputs

has raised concerns in high-stakes domains. Research into formal methods, logical reasoning systems, and probabilistic graphical models continues to provide critical foundations for trustworthy AI. The state of the art now includes techniques that attempt to integrate symbolic representations with neural networks, enabling more interpretable and verifiable reasoning processes.

Challenges remain in bridging the gap between the statistical nature of modern AI models including LLMs and the rigor required for verifiable outputs. Deep learning systems often lack transparency and accountability, posing risks when deployed in autonomous or safety-critical environments. To address these concerns, federal investments should prioritize long-term research on hybrid neuro-symbolic approaches, formal verification methods, and foundational AI science that is independent of commercial imperatives.

Long-term foundational AI research must focus on uncovering the principles, architectures, and learning mechanisms necessary for building systems with broad generalization capabilities, robust reasoning, and sustained adaptability. A key strategic direction involves developing AI systems that can autonomously acquire and refine knowledge over time—learning from limited data, transferring knowledge across domains, and continually updating their internal models in response to novel situations. This requires advances in lifelong and self-supervised learning, meta-learning, and neuro-symbolic integration that combine the flexibility of statistical methods with the structure and interpretability of symbolic reasoning. Understanding how to build systems that can explain their behavior, justify their decisions, and gracefully recover from failures remains an open and critical research problem.

Ensuring that AI systems produce accurate and reliable outputs has become a central concern. Despite recent efforts to increase the reliability of AI systems, e.g., LLM, to generate factually accurate outputs, hallucinations, model brittleness, and context drift continue to undermine trust in generative AI systems. Current benchmarks reveal substantial gaps in performance, even among leading models. To advance this area, future investments should support the development of high-quality public datasets, robust evaluation frameworks, and models that can explicitly quantify uncertainty and explain their reasoning. Emphasis should also be placed on context-aware trust frameworks and dynamic knowledge updating.

Another long-term challenge lies in building AI systems that possess *generalizable reasoning* and *commonsense understanding*—core to human intelligence but still elusive for machines. Research must explore how to endow AI with the ability to reason causally, handle uncertainty, and construct abstract representations of the world. Foundational work is also needed in formal verification, value alignment, and the study of emergent behavior in large-scale systems. Additionally, understanding the theoretical limits of learning, representation, and decision-making in complex environments can inform the design of more efficient and reliable AI. These challenges are not merely technical; they intersect with deep philosophical and cognitive questions, requiring interdisciplinary engagement across computer science, neuroscience, cognitive science, and ethics.

Strategic, long-horizon investments in these foundational areas are essential to build AI that is not only powerful, but principled, safe, and aligned with long-term human values.

Universities are uniquely positioned to focus on the fundamental aspects of AI that underpin its long-term evolution. While industry often focuses on large-scale development and application of AI technologies, basic, curiosity-driven research carried out in academic settings is crucial for major conceptual breakthroughs. For example, the recent development of large language models has benefited from decades of academic research on natural language processing, machine learning, high-performance computing, and related areas. Further advances in AI will require advances across many subfields of AI, including techniques for ensuring the safety, robustness, reliability and explainability of AI systems, development of AI agents that can coordinate to perform complex tasks, etc.

1. Intelligent Agents and Multi-Agent Systems. Strategic research in intelligent agents and multi-agent systems must focus on developing unified, adaptive architectures that integrate core components of intelligence—perception, learning, reasoning, decision-making, communication, and coordinated action. These systems should be capable of operating autonomously and collaboratively in complex, uncertain, and dynamic real-world environments. Current AI systems often address these facets in isolation, but building general-purpose agents requires holistic models that can seamlessly synthesize sensory data, learn from experience, infer goals, plan and execute actions, and interact effectively with other agents and humans. Achieving this integration demands advances in multi-modal learning, hierarchical representations, continual learning, and explainable decision-making frameworks.

In multi-agent settings, additional layers of complexity arise from the need for coordination, negotiation, competition, and cooperation among autonomous entities. Research must explore protocols and algorithms for decentralized decision-making, shared knowledge representation, and dynamic team formation. Agents must be capable of modeling others' intentions, adapting strategies in real time, and aligning actions with shared objectives or societal norms. Safety, transparency, and robustness are especially critical in settings where agents interact with humans or other critical systems. Strategic directions include developing benchmarks for emergent behavior, frameworks for multi-agent value alignment, and methods for verifying collective behavior. These advancements will be foundational for applications such as autonomous vehicles, distributed robotics, intelligent assistants, and collaborative problem-solving in scientific, industrial, and social domains.

2. Embodied AI. Embodied AI represents a foundational shift in artificial intelligence—one that grounds intelligence in physical experience and real-world interaction. Strategic research in this area must focus on developing agents that perceive, learn, and act through continuous engagement with their environments. Unlike purely data-driven models trained on static inputs, embodied agents must process sensory feedback, adjust to changing conditions, and make context-sensitive decisions in real time. This requires the integration of robotics, neuromotor control, reinforcement learning, and cognitive

modeling to create systems capable of purposeful, adaptive, and safe behavior in uncertain, dynamic settings.

A key direction is the development of advanced sensorimotor learning algorithms that allow agents to acquire skills through trial and error, imitation, or goal-directed exploration—not unlike how humans and animals learn. High-fidelity simulation environments will play a critical role in training and testing these systems efficiently and safely before real-world deployment. Research must also address challenges in transferring learned behaviors from simulation to reality, ensuring robustness to unstructured environments, and enabling embodied agents to collaborate with humans. Applications span from autonomous service robots and assistive technologies to embodied educational tools and interactive AI companions. Realizing the full promise of embodied AI will require interdisciplinary collaboration across AI, neuroscience, biomechanics, and human-computer interaction, alongside ethical frameworks that ensure these agents operate transparently and responsibly in environments that include human actors.

3. AI models of human and animal intelligence. AI models inspired by human and animal intelligence offer a promising path toward understanding the principles that underlie general intelligence and adaptive behavior. Strategic research in this area involves building computational models that replicate cognitive, emotional, and social functions observed in natural organisms—from perception, memory, and decision-making to empathy, cooperation, and communication. These models not only help explain how brains function but also provide valuable insights into how to design AI systems that are more robust, flexible, and capable of operating in real-world, uncertain environments. Incorporating biological constraints and developmental learning processes into AI architectures can lead to systems that learn more efficiently, generalize more effectively, and adapt in open-ended ways.

This bidirectional agenda requires tight integration between AI research and fields such as neuroscience, psychology, cognitive science, and ethology. One strategic direction is the use of AI to simulate and test theories of brain function, offering scalable and testable platforms for investigating complex neural and behavioral phenomena. Conversely, biologically inspired models—such as those based on hippocampal and cortical architectures, hyperdimensional computing, or social learning in animals—can inform new AI architectures capable of developmental, continual and lifelong learning, real-time reasoning, dynamic interaction and adaptation. Another crucial research focus is understanding and modeling social intelligence: how agents infer others' beliefs, intentions, and emotions, and how such mechanisms contribute to cooperation and competition. This line of research not only advances AI but also has profound implications for understanding human cognition, enhancing mental health tools, and building AI systems that interact more naturally and ethically with people.

4. Human-Centered AI. Advancements in AI-human interaction technologies have transformed the way people engage with AI systems. Virtual assistants, code copilots, and dialog agents are increasingly used across professional and personal domains.

These systems rely heavily on models capable of understanding context, interpreting intent, and engaging in multi-turn conversations. Human-AI collaboration now involves shared decision-making, adaptive interfaces, and interactive learning.

Despite these developments, critical challenges persist. AI systems often fail to understand user goals, exhibit brittleness in dynamic environments, and lack mechanisms for transparent co-reasoning. The next phase of research must focus on building systems that support mutual understanding, adjustable autonomy, and trust. Investments are needed in cross-disciplinary research integrating cognitive science, human-computer interaction, and AI, as well as standardized evaluation protocols for collaborative performance.

Human-Centered AI (HCAI) research must focus on the principles and practice of placing human values, needs, and well-being at the core of AI system design, deployment, and governance. A key strategic research direction is the development of AI systems that are interpretable, transparent, and accountable—allowing users to understand how decisions are made and to challenge or correct them when necessary. This includes advancing explainable AI techniques, participatory design, and robust evaluation frameworks that reflect real-world contexts and diverse user perspectives. Research must also address algorithmic fairness, privacy preservation, and bias mitigation, particularly in applications that have high societal impact, such as education, healthcare, and criminal justice.

Another critical direction is enabling effective and equitable human-AI collaboration. This involves designing systems that can adapt to individual users' goals, preferences, and cognitive models while supporting shared decision-making and trust over time. Strategic investments are needed in participatory design methodologies, user-aligned reward functions, adaptive interfaces, and interactive learning environments that empower users rather than displace them. Interdisciplinary research that draws on psychology, sociology, ethics, design, and education is essential to understand how people interact with AI and how systems can be made more inclusive, accessible, and responsive. Ultimately, Human-Centered AI should aim not just to optimize performance metrics, but to augment human potential and promote societal well-being.

5. AI for society. A responsible AI future requires cross-disciplinary inquiry into AI's societal impact, including ethical, legal, economic, and cultural dimensions. Research must explore mechanisms for democratic governance of AI technologies, equitable access, and proactive mitigation of systemic biases and harms.

Strategic research in AI for society must prioritize understanding and shaping the complex ways in which AI technologies influence human institutions, social norms, and collective well-being. This includes studying the ethical, legal, economic, and cultural dimensions of AI deployment, particularly in contexts where algorithmic systems mediate access to resources, opportunities, and decision-making. Key research directions include identifying and mitigating algorithmic bias, ensuring transparency and accountability in automated systems, and understanding the broader societal impacts of AI on labor markets, public health, education, media, and democratic participation.

Social science-informed AI research is critical to uncover how AI technologies interact with historical inequalities, political structures, and cultural values.
The convergence of short-term harms and long-term risks has reshaped the discourse on AI ethics and safety. At the same time, the theoretical risks posed by misaligned superintelligent systems are no longer considered purely speculative. The emergence of AI-enabled cybercrime, deepfakes, and autonomous weapons has heightened the urgency of research on safety of AI systems. An important strategic direction has to do with development of formal methods to support safety by design as well as rigorous testing, evaluation, safety certification, and regulation frameworks.

Equally important is the design and evaluation of governance frameworks that ensure AI technologies are aligned with democratic principles and the public interest. This includes developing mechanisms for participatory oversight, impact assessment, and regulatory innovation that keep pace with rapid technological change. Research should explore models for equitable access to AI infrastructure and capabilities, particularly in underserved or marginalized communities, both domestically and globally. Strategic collaboration among technologists, policymakers, civil society organizations, and affected communities is essential to ensure that AI systems reflect diverse perspectives and are deployed in ways that support social justice, institutional trust, and human dignity. Through cross-disciplinary engagement and inclusive design, research in AI for society can guide the development of technologies that not only avoid harm but actively contribute to a more just and resilient world.

6. AI for Science. AI for science represents one of the most promising frontiers in accelerating discovery and expanding the boundaries of human knowledge.

A key strategic research direction is the development of AI systems that can assist in all stages of the scientific process—from hypothesis generation and experimental design to data analysis and interpretation. This involves creating AI models capable of integrating multimodal scientific data, reasoning across complex systems, and supporting causal inference, uncertainty quantification, and theory construction. In disciplines such as materials science, biology, chemistry, AI can dramatically reduce the time and cost associated with experimentation by predicting outcomes, optimizing simulations, and automating laboratory workflows.

Equally important is building AI systems that function as collaborative scientific partners—designed not to replace researchers but to augment their creativity, intuition, and expertise. This requires human-AI interfaces that support interactive exploration, adaptive feedback, and domain-informed learning. AI systems must be interpretable and transparent to earn the trust of scientists, and must support reproducibility, knowledge transfer, and integration with existing tools and workflows. Strategic investments should also target the development of domain-specific foundation models, open-access scientific datasets, and scalable computational infrastructure. Interdisciplinary collaboration between AI researchers and domain scientists is essential to ensure that AI methods are tailored to the unique challenges of each scientific field

and that their outputs are aligned with scientific norms, rigor, and support curiosity-driven inquiry.

7. **AI for Humanities and Creative Arts.** In the humanities, AI can analyze vast archives, uncover hidden patterns in language, art, and history, and support new modes of scholarly inquiry. In the creative arts, AI can act as collaborator, catalyst, and medium—augmenting human expression across music, literature, visual art, performance, and design.

Strategic research in AI for the humanities and creative arts should focus on developing systems that can support nuanced cultural interpretation, critical analysis, and creative collaboration. In the humanities, AI can help scholars analyze massive archives of texts, images, and audiovisual materials to uncover historical patterns, trace cultural influence, and generate new research questions. Research must address how to design AI models that are sensitive to context, ambiguity, and cultural diversity—going beyond surface-level pattern recognition to support interpretive depth. This includes creating tools for comparative analysis across languages and media, developing algorithms that can model narrative structure and metaphor, and enabling scholars to explore historical and cultural data in immersive, interactive ways. Ensuring transparency, provenance tracking, and methodological rigor in AI-assisted humanities research is essential to maintain scholarly standards and trust.

In the creative arts, strategic directions should focus on developing AI systems that empower, rather than constrain, human creativity. This includes designing generative models that can co-create with artists across disciplines—suggesting alternatives, transforming styles, or responding to real-time input in performance or design. Open research problems include how to encode aesthetic principles, personal style, and emotional nuance into AI systems, and how to support collaborative creativity while preserving artistic agency and authorship. Research should also explore how AI can be used as a medium in itself—expanding the possibilities of digital art, interactive installations, algorithmic composition, and hybrid human-machine performances. Cross-disciplinary collaboration between AI researchers, artists, humanists, and ethicists will be crucial to ensure that these tools are culturally aware, ethically grounded, and accessible to a broad range of creators and communities.

8. **Translational Research in AI.** Strategic translational research in AI must focus on bridging the gap between foundational innovations and their real-world deployment to ensure that AI technologies deliver broad societal value. This requires the development of end-to-end pipelines that support iterative testing, validation, and refinement of AI systems in realistic, complex environments. A key direction is the creation of shared testbeds and open platforms for experimentation—particularly in high-impact domains such as healthcare, education, agriculture, and critical infrastructure. These platforms should support rigorous evaluation of safety, accountability, usability, enabling researchers and practitioners to identify failure modes, improve robustness, and adapt systems to diverse operational contexts.

Another priority is the advancement of governance tools and regulatory-ready frameworks that can accelerate responsible deployment. This includes designing AI systems with built-in transparency, traceability, and value alignment, as well as establishing interoperable standards for data, models, and performance metrics. Public-private partnerships will play a vital role in translational research by fostering collaboration across academia, industry, and government, pooling resources, and aligning incentives. Equally important is the cultivation of interdisciplinary teams—including domain experts, ethicists, and end-users—to ensure that AI systems are designed with public needs, ethical considerations, and long-term societal impact in mind. By strengthening these translational pathways, AI research can more effectively realize its potential to enhance human well-being, economic opportunity, and institutional resilience.

9. Core Computer Science Research for advancing AI. Core computer science (CS) research forms the bedrock upon which all advances in artificial intelligence rest. As AI systems grow in complexity, scale, and societal importance, foundational research in algorithms, data structures, programming languages, formal methods, computational complexity, computer architecture, and distributed systems becomes increasingly critical. Strategic research in these areas enables scalable learning systems, improves computational efficiency, and supports the design of algorithms that are provably correct, fair, and robust. For instance, innovations in optimization algorithms and data management underpin the performance of modern deep learning systems, while breakthroughs in complexity theory help delineate the theoretical limits of what AI can achieve.

Equally vital is the development of programming languages and formal verification methods tailored to AI workloads. As AI applications become embedded in safety-critical domains—from autonomous vehicles to medical diagnostics—guarantees of correctness, interpretability, and safety are paramount. Research into distributed and parallel systems is also essential to support the training and deployment of large-scale AI models, especially under real-world constraints of latency, energy use, and data privacy. Advancing computer architecture to support AI workloads more efficiently—through specialized accelerators, memory systems, and hardware/software co-design—is another key frontier. Sustained investment in core CS research not only drives innovation but also ensures that AI systems are reliable, trustworthy, and sustainable at scale—laying the groundwork for future discoveries and applications across every domain AI touches.

B.2 Strategic Priorities in AI Research Infrastructure and Ecosystem

1. Access to Datasets, Compute, and Tools. Access to high-quality datasets and computational resources remains a major barrier for academic and public-interest AI research. Although open-source models and datasets have proliferated, much of the innovation remains gated behind proprietary infrastructure. The disparity in compute

access between industry and academia threatens to marginalize independent research and slow down progress in areas of public concern.

Recent developments in synthetic data generation, federated learning, and shared model architectures offer partial solutions, but a national strategy is needed. Federal investments should support public computing infrastructure, curated open datasets for socially relevant domains, and incentives for open-source tooling. Investments are needed to provide universities and colleges with access to a range of data sets and computational infrastructure to support AI research and education. AI research, particularly in machine learning and deep learning, demands significant computing power, often requiring access to GPU clusters, high-performance computing (HPC) systems, and cloud-based resources. Scalable computational infrastructures will enable researchers to tackle complex problems, such as training large models and to rigorously evaluate AI models. Also needed are investments in staff scientists with expertise in AI and/or in specific application domains to work with interdisciplinary teams of faculty on complex research projects that leverage AI to advance scientific, engineering or other applications of AI. Access to top-tier infrastructure will allow universities to remain competitive with industry leaders and attract high-caliber researchers who rely on cutting-edge tools to push the boundaries of AI development

2. Evaluation Infrastructure. Robust evaluation of AI systems is essential for responsible deployment. While benchmark-driven evaluation remains standard practice, it suffers from limitations including overfitting, lack of real-world generalizability, and poor reflection of human-AI interaction dynamics. New efforts have emerged to assess usability, robustness, fairness, and adherence to ethical standards. Nevertheless, there is no unified science of evaluation that meets the complexity of modern AI systems.

Federal initiatives should support the development of a comprehensive evaluation science. This includes longitudinal studies of system behavior post-deployment, multi-dimensional evaluation frameworks, and standardized protocols for human-centered assessments. Investments are also needed in red-teaming infrastructure, simulation environments, and reproducibility frameworks that allow for independent auditing of model performance and behavior.

3. Inclusive and Interdisciplinary AI Talent and Workforce

To build a truly inclusive and interdisciplinary workforce, federal efforts must expand access to AI education and research opportunities. This includes certificate programs for domain experts, and support for AI+X graduate training initiatives. Investments in K–12 education and community college partnerships are also critical for building future capacity.

The development of world-class AI talent is one of the most effective ways for universities to contribute to the field's progress. Strategic investments that support junior researchers through grants, research fellowships, and mentorship programs are necessary to nurture

the next generation of AI researchers and educators. Federal investments are needed to develop a healthy pipeline of talent across all areas of AI.

The AI workforce has grown significantly, with increased participation from diverse academic disciplines. Programs that bridge AI with medicine, education, law, and agriculture, the arts and the humanities, have become more prominent. Still, barriers persist in training, retention, and inclusion, particularly for underrepresented groups and non-technical disciplines.

4. Public-Private and Academic-Industry Partnerships

Public-private partnerships have yielded promising collaborations in safety, evaluation, and open science. Nonetheless, tensions around intellectual property, transparency, and alignment of interests remain. There is a risk that foundational research becomes overly dependent on commercial priorities, potentially limiting its societal value.

Going forward, federal programs should incentivize equitable partnerships that prioritize openness, reproducibility, and public interest outcomes. Mechanisms for fair IP sharing, benefit alignment, and dual-use assessment must be embedded into partnership frameworks. Federally funded joint research centers can serve as neutral grounds for long-term, high-impact projects involving academia, industry, and civil society.

5. International Cooperation

AI innovation is increasingly shaped by global dynamics, with competition and cooperation unfolding simultaneously. The past two years have seen a proliferation of international frameworks for AI governance, safety, and ethics. While promising, these efforts remain fragmented and uneven across regions.

The United States should lead in establishing global norms for AI safety, openness, and equity. Investments are needed in international research partnerships, multilateral capacity-building programs, and diplomatic channels for AI governance. Shared testbeds, model audits, and evaluation protocols can foster mutual trust and accountability in a geopolitically sensitive landscape.

C. Conclusion

This strategic plan recognizes the rapid and multi-dimensional evolution of AI technologies and their societal implications. It calls for a coordinated national response rooted in foundational research, responsible innovation, and broad-based collaboration. By aligning investments with the challenges and opportunities outlined here, the United States can continue to lead in the development of AI systems that are powerful, ethical, safe, and beneficial to all. To navigate this evolving AI landscape, the United States must

adopt a forward-looking, comprehensive national AI strategy grounded in the following pillars:

1. Fund Fundamental and Responsible AI Research: Prioritize sustained investment in foundational research aimed at advancing robust and generalizable AI systems. This includes support for long-term, open exploration of core AI capabilities such as reasoning, planning, decision-making, learning, communication, interaction, and coordination. Also critical is support for research in embodied AI – where intelligence is grounded in interaction with the physical world; bidirectional research agenda focused on understanding human and animal intelligence through the lens of AI models and the use of the resulting insights to improve AI systems; and research on AI to enhance human creativity. It is critical to invest in research on how to enhance the are trustworthiness and safety of AI systems, and their alignment with societal values. Also critical are investments in interdisciplinary research that embeds ethics, safety, and accountability into the design process from the outset, ensuring that responsible AI principles guide development at every level—from algorithm design to deployment in real-world, high-stakes domains.

2. Close Infrastructure and Compute Gaps: Build and maintain a robust, national-scale public compute infrastructure that empowers academic institutions, nonprofits, and public-interest researchers to participate fully in the development and oversight of frontier AI. This includes provisioning access to high-performance computing clusters, scalable cloud resources, and advanced hardware accelerators. Establish open model and dataset repositories with transparent licensing and governance to facilitate reproducibility, collaborative benchmarking, and responsible innovation. By broadening access to the tools and infrastructure that currently concentrate power in a few industrial actors, we can foster a more inclusive, diverse, and accountable AI research ecosystem.

3. Build a Sustainable AI Ecosystem: Promote a transformative shift toward environmentally responsible AI by embedding sustainability into every layer of the AI development lifecycle. Advance research in green AI, including energy-efficient algorithms, adaptive model scaling, and carbon-aware training and deployment strategies that optimize performance without environmental compromise. Catalyze breakthroughs in low-power hardware, neuromorphic computing, and photonic processors that radically reduce the energy footprint of AI systems. Require comprehensive, transparent lifecycle carbon impact assessments for large-scale AI models—from data acquisition and model training to deployment and retirement. Beyond technical solutions, support cross-sector collaboration between AI researchers, environmental scientists, and policymakers to create standards, incentives, and regulatory frameworks that align AI innovation with planetary health. A sustainable AI future is not just about efficiency—it is about responsibility, resilience, and ensuring that the power of AI serves both humanity and the Earth.

4. Foster Inclusive AI Education and Workforce Development: Build a diverse and future-ready AI workforce by expanding access to high-quality AI education across the entire learning spectrum—from K–12 to post-secondary and lifelong learning. Prioritize outreach and investment in underserved and historically marginalized communities to

ensure equitable participation in the AI-driven economy. Equip educators with robust, adaptable AI curricula, professional development resources, and tools to teach both technical skills and ethical reasoning. Launch AI+X initiatives that empower students to apply AI across disciplines—medicine, law, environmental science, humanities, and the arts—cultivating interdisciplinary fluency and real-world problem-solving. Partner with schools, community colleges, industry, and public institutions to create scalable apprenticeship pathways, inclusive competitions, and credentialing programs. By fostering inclusive, interdisciplinary, and socially aware AI education, we can cultivate the next generation of innovators, leaders, and citizens who will shape AI in the public interest.

5. Modernize Regulatory and Ethical Frameworks: Accelerate the development of agile, forward-looking governance structures that can keep pace with the rapid evolution of AI technologies. Coordinate federal initiatives to complement state-level innovation while actively engaging in global forums to harmonize international standards and promote shared values around safety, fairness, and human rights. Establish robust legal and institutional mechanisms to ensure accountability for the deployment of foundation models, including mandatory transparency reporting on data provenance, training procedures, model capabilities, and known risks. Implement enforceable safeguards against algorithmic discrimination, misinformation amplification, and systemic safety failures—especially in high-stakes domains such as healthcare, education, employment, and national security. Support the creation of independent auditing bodies, participatory oversight processes, and ethical review boards to ensure that the development and deployment of AI systems are aligned with democratic values and societal well-being. A modern AI governance framework must not only mitigate harm but actively promote justice, equity, and public trust.

6. Strengthen Global Leadership and Alliances: Lead efforts to create interoperable, enforceable international frameworks that govern the responsible development, deployment, and oversight of AI technologies—ensuring alignment on safety standards, data privacy, ethical use, and cross-border accountability. Expand strategic participation in multilateral institutions such as the OECD, G7, UN, and emerging global AI compacts, while fostering new alliances that amplify the voices of low- and middle-income countries in AI governance. Fund and coordinate joint R&D initiatives with like-minded nations to accelerate progress in areas of global importance, including climate resilience, health equity, and cybersecurity. By uniting democratic partners around a shared vision for trustworthy AI, we can counter authoritarian models of technological control and build an inclusive, secure, and values-driven global AI ecosystem.

7. Elevate Science and Public Health through AI: The cross-disciplinary collaborations to drive transformative AI-enabled advances in critical fields such as materials innovation, healthcare, manufacturing, agriculture, and public health. By fostering these partnerships, we will unlock transformative insights and innovative solutions that address urgent global challenges, driving progress toward a healthier planet and society.