I-140 Immigrant Petition

Petitioner and Beneficiary: X

Classification Sought: INA §203(b)(2)(B)(i)

This petition is respectfully submitted in support of Dr. X's application for classification as a member of the professions holding an advanced degree under the employment-based second preference (EB-2) category, and a National Interest Waiver (NIW) pursuant to *Section 203(b)(2)(B)(i) of the Immigration and Nationality Act (INA).*

In accordance with 8 C.F.R. § 204.5(k)(3)(i), Dr. X qualifies as a member of the professions holding an advanced degree, with a doctoral degree in <computer science degree> from the University X and a Master's degree in <social science degree> from the University Y (see ***Section 1***).

Situated at the intersection of Social Computing and Artificial Intelligence (AI), Dr. X's proposed endeavor seeks to advance cost-effective, domain-specific computational and AI mdoels and systems through the systematic integration of specialized domain knowledge into their architecture and operation, enabling AI technologies to more effectively meet the nuanced demands of specific domains (see ***Section 2***).

In support of this petition, Dr. X provides comprehensive evidence satisfying the three prongs set forth in *Matter of Dhanasar, 26 I&N Dec. 884 (AAO 2016)*:

1. **Substantial Merit and National Importance:** The proposed endeavor has the potential to enhance the accessibility, efficiency, and effectiveness of AI technologies across critical scientific, industrial, and societal sectors in the United States, thereby generating significant technological, economic, and societal value (see ***Section 3***).
2. **Well Positioned to Advance the Endeavor:** Dr. X's interdisciplinary background, industry and research experience, and record of publications, patents, and professional

recognition demonstrate that he is well positioned to pursue his proposed endeavor (see ***Section 4***).

3. **Benefit from Labor Certification Waiver**: On balance, it would be in the national interest to waive the job offer and labor certification requirements, as such requirements could limit Dr. X's ability to fully apply his expertise and advance his proposed work (see ***Section 5***).

Based on the evidence presented, Dr. X respectfully requests favorable consideration of this petition and approval of his National Interest Waiver under the EB-2 category.

The structure of this petition is organized as follows:

1. **Advanced Degrees and Professions**
2. **Proposed Endeavor**
3. **Substantial Merits and National Importance of the Endeavor (Prong 1)**
4. **Ability to Achieve the Endeavor (Prong 2)**
5. **Justifications for the Job Offer and Labor Certification Waiver (Prong 3)**
6. **List of Exhibits**

1 Advanced Degrees and Professions

1.1 Advanced Degrees

Dr. X holds a doctoral degree in <computer science degree> from the University X, and a master's degree in <social science degree> with a <computational certificate> from the University Y [*Exhibit 2 - Diplomas and Transcripts*]. Both institutions are highly regarded for academic excellence in their respective disciplines. According to U.S. News & World Report, X ranks No. 1 nationally in <Discipline X> and among the top five

programs in Computer Science, while the University Y is ranked among the top ten graduate programs in multiple disciplines in Social Sciences [*Exhibit 20 - Graduate School Ranking*]. His education background provides him with a rare interdisciplinary foundation combining technical, scientific, and societal knowledge.

1.2 Academic Membership and Service

Dr. X is an active member of the Association for Computational Linguistics (ACL) and the Association for Computing Machinery (ACM) [*Exhibit 3 - Membership Cards*]. He serves as a regular reviewer for the ACL Rolling Review (ARR), the unified evaluation platform supporting the most prestigious conferences in Natural Language Processing (NLP), and also reviewed research papers in other interdisciplinary conferences and journals spanning both AI and interdisciplinary fields [*Exhibit 22 - Review Records*].

1.3 Industry Engagement and Patents

Dr. X has also been a <Position X> specializing in conversational AI at <Company X> for nearly three years [*Exhibit 5 - Proof of Work*]. Specifically, he previously completed two summer internships with the conversational AI team in 2022 and 2023, followed by a one-year part-time contract from 2022 to 2023. He then joined <Company X> as a full-time employee in July 2024. His work in the conversational AI focuses on developing cost-effective AI systems that provide accurate, user-friendly, and safe X service to support customers in making informed X decisions. The company has filed **two patents** with him as the named inventor [*Exhibit 7 - Proof of Patents; See details in* ***Section 4.4.1***].

2 Proposed Endeavor

2.1 Endeavor Overview

Dr. X's proposed endeavor centers on the development of cost-effective, domain-specific computational and AI models and systems by integrating specialized domain knowledge into their design and functionality. Drawing upon his interdisciplinary expertise in computer sciences, information sciences, and social sciences, Dr. X seeks to overcome key limitations of current AI technologies, notably their inadequate handling of specialized, domain-specific tasks, high operational costs, and limited adaptability to specialized contexts (see ***Section 2.2***). This targeted approach prioritizes domain specificity, efficiency, and affordability, ensuring that AI solutions are optimized for nuanced, context-dependent tasks (see ***Section 2.3***). The endeavor will deliver several key contributions and create substantial national impact (see ***Section 3***): By embedding expert domain knowledge within AI systems, the endeavor aims to enhance their accuracy and practical utility, significantly broadening AI accessibility and application across scientific, societal, and industrial sectors. Importantly, Dr. X's endeavor frames the relationship between humans and AI as complementary rather than substitutive, fostering a productive integration of AI capabilities with professional knowledge. Consequently, this initiative advances national interests by significantly enhancing resource efficiency and accuracy of computational and AI models and systems. Furthermore, it improves AI accessibility in sectors currently underserved due to resource constraints and limited domain-specific adaptability.

2.2 Limitations of Current AI and Computational Technologies

Despite the rapid advancement and widespread use of current AI technologies, significant limitations still exist that restrict their full potential and applicability in the society. The inter-related key barriers and shortcomings that the proposed endeavor aims to address include:

- **Lack of Domain-Specific Sensitivity:** Current AI models and systems usually

prioritize general knowledge and struggle to capture the domain-specific details necessary for effectively addressing tasks within particular fields. For example, a 2025 study [*Exhibit 25.1 - Research Article*], authored by a broad team of researchers from leading U.S. universities and corporate research laboratories, emphasized that, "directly applying LLMs to solve sophisticated problems in specific domains meets many hurdles, caused by the heterogeneity of domain data, the sophistication of domain knowledge, the uniqueness of domain objectives, and the diversity of the constraints (e.g., various social norms, cultural conformity, religious beliefs, and ethical standards in the domain applications)." Therefore, advancing domain-specific and adaptable AI modeling, or effectively adapting general-purpose models to domain-specific tasks, is of critical importance for unlocking the full potential of AI to benefit society, accelerate scientific discovery, and enhance industry productivity.

- **Hallucinations and Factual Inaccuracy:** AI models (such as ChatGPT) often generate responses that sound convincing but are factually incorrect or entirely fabricated, a phenomenon known as "hallucination" or "Inaccuracy." This problem poses serious risks across multiple domains. For example, in decision-making, hallucinated content can mislead policymakers, legal professionals, and even individuals from evidence-based decisions [*Exhibits 23.6; 24.2 - DOJ and NCSC Reports*]. In e-commerce, hallucinations can erode consumer trust, damage brand reputation, and even trigger regulatory or compliance violations if chatbots provide inaccurate information about products or services [*Exhibit 6 - Supporting Letter from Company X*]. The causes of hallucinations stem from two main sources. First, as the first challenge mentions, general-purpose models rely on generic knowledge and reasoning, which are often insufficient for specialized tasks and lead to errors due to the absence of domain expertise. Second, limitations within the models themselves (particularly in inference and reasoning capabilities) further contribute to these inaccuracies. The solutions to these issues determine whether AI can be reliably and safely integrated into critical real-world applications, ultimately influencing its potential for substantial positive societal and economic impacts. The solution to this issue have been one of the top considerations of the government [*Exhibits*

23.2; 23.6; 24.2 - Executive Order 14110 and DOJ & NCSC Reports].

- **High Computational and Financial Costs:** Developing, adapting, and using modern AI models for domain-specific tasks still faces high computational and financial costs, which requires enormous resources, including large volumes of human-labeled data, data center infrastructure, and massive electricity [*Exhibits 23.1; 25.2 - Executive Order 14141 and Research Articles*]. These demands create a significant barrier for smaller research groups, nonprofit organizations, and small- to medium-sized companies, which often lack the funding and computational resources to develop models tailored to their specific needs. A report from National Science Foundation to the president and the Congress [*Exhibit 23.5 - NSF Report*] states that "the opportunities to pursue cutting-edge AI research and apply AI to new domains and challenges are currently not accessible by all of America's incredible talent nor harnessed by the public sectorm, are often unavailable to researchers beyond those at well-resourced technology companies and universities." Furthurmore, even directly using large language models (LLMs) can be prohibitively expensive, as it generally involves either deploying and running models on cloud servers with high-end GPUs and substantial bandwidth or relying on costly commercial Application Programming Interfaces (APIs), such as those offered by OpenAI. Importantly, these expenses can pose a financial burden not only for smaller entities but also for larger companies and established research organizations. Consequently, developing cost-effective methods to build or adapt domain-specific models remains a critical priority in order to unlock the full potential of AI.

2.3 Proposed Research in the Endeavor

To directly address these challenges, **Dr. X's proposed endeavor introduces three interconnected strategies:** (1) integrating specialized domain knowledge into AI systems and computational models; (2) employing automated methods to efficiently collect and expand domain-specific resources for training, adapting, and deploying AI; and (3) developing cost-effective computational and AI models and systems optimized to perform effectively even with limited computing resources.

- **Domain Knowledge Incorporation into AI Systems and Modeling:** Human-in-the-loop approaches are increasingly popular for enhancing AI performance. For example, ChatGPT, the widely used large language model (LLM), employs Reinforcement Learning from Human Feedback (RLHF), a technique in which human input is critical for guiding model behavior and improving accuracy. Similarly, in domain-specific tasks, incorporating specialized domain knowledge can significantly enhance the accuracy, relevance, and reliability of AI systems, especially when dealing with content and concepts that differ substantially from everyday or general-use contexts. However, effectively bridging the gap between AI technical development, specialized domain expertise, and model behavior remains challenging and requires focused research. Technical experts in AI modeling and domain specialists often have different understandings and priorities, leading to discrepancies and conflicts between specialized domain requirements and AI systems. Addressing this gap, particularly in a cost-effective and scalable manner, is essential to fully realize AI's potential within specialized domains.

- **Automated Curation of Domain-Specific Resources:** Domain-specific resources, such as specialized dictionaries, data with domain specific content, and structured knowledge about concepts, terms, and their meanings are extremely valuable for creating and using accurate and effective AI systems. These resources help AI systems better understand specialized topics and concepts, produce more accurate and reliable results, and correctly interpret or improve their outputs, both when developing the AI models and when applying them in real-world tasks. Traditionally, preparing these resources requires extensive human effort, making the process costly, slow, and inefficient. To address this, Dr. X's research focuses on developing automated methods to quickly gather, organize, and expand specialized resources from available materials, such as online platforms, academic articles, technical reports, organization documents, and other digital documents. By automatically selecting relevant information and organizing it clearly for AI use, this approach significantly reduces costs and saves time. Ultimately, it improves AI's accuracy, usability, and accessibility in specific domains. When these resources are built using publicly accessible data, they can also be shared with other

researchers, organizations, or stakeholders, significantly amplifying their value and impact. Importantly, while the previously discussed strategy focuses on methods to incorporate expert knowledge directly into AI models, this resource-curation strategy creates essential materials that can facilitate such knowledge incorporation. Together, these two interconnected strategies enable a more efficient, cost-effective, and reliable integration of domain-specific expertise into AI systems, benefiting critical scientific and societal fields.

- **Cost-Effective Models and Methods:** The third focus of this endeavor emphasizes cost-effectiveness across both the usage and development phases of AI systems. This focus consists of both AI use and development phases. First, during the usage phase, AI models and systems must be cost-effective in carrying out domain-specific tasks. Deploying and maintaining large, complex AI models requires substantial infrastructure, including high-performance computing resources, storage capacity, and ongoing technical support, all of which generate significant financial costs. Relying on commercial APIs to access such models also incurs considerable expenses, since API usage is typically priced according to model size and request volume. However, larger and more complex models for open domains are not always necessary, as domain-specific applications often involve narrower tasks that can be addressed more efficiently. In such cases, cost-effective strategies may include using smaller models tailored to the domain, regardless of their performance on general-purpose tasks, or combining smaller domain-specific models with larger models to address different subtasks. These approaches significantly reduce costs, whether the models are deployed independently or accessed through APIs. Second, cost reduction is also achieved during model development by leveraging the other two strategies outlined above: domain knowledge incorporation and automated curation of domain-specific resources. Embedding specialized expertise directly into AI systems and automating the creation of high-quality resources both reduce reliance on extensive manual data collection and large volumes of training data for model adaption or training. These efficiencies offset significant portions of the financial and computational burden typically associated with training and adapting advanced models.

In summary, this proposed endeavor directly addresses three critical limitations of current AI technologies: the lack of domain-specific sensitivity, the risks of factual inaccuracies and hallucinations, and the high computational and financial costs that restrict accessibility. By integrating specialized knowledge into AI systems, automating the creation of domain-specific resources, and developing cost-effective models and methods, this endeavor establishes a coherent and practical framework for advancing AI beyond its current constraints. The combined effect of these strategies is to create AI systems that are accurate, efficient, and adaptable to real-world contexts.

3 Substantial Merit and National Importance

Dr. X's proposed endeavor advances computational and AI models and systems toward broader accessibility, precise domain adaptation, and practical utility. By reducing barriers to adoption and enabling reliable AI systems for specialized contexts, the endeavor expands the reach and impact of AI across scientific research, industrial innovation, and societal applications. It aligns with strategic U.S. objectives of fostering innovation, developing a skilled workforce, and sustaining economic competitiveness, reinforcing the nation's global leadership in responsible and inclusive AI.

3.1 Expanding the Reach of AI

A central benefit of the proposed endeavor is to expand the reach of AI into a broader set of sectors and fields that are currently unable to leverage its potential.

3.1.1 Lower costs and address computing resource demand

U.S. government vision: The U.S. government has explicitly acknowledged that access to advanced AI development remains concentrated among a limited number of well-resourced technology companies and research universities, creating an "access divide" that constrains innovation and limits the nation's ability to leverage AI for pressing societal challenges [*Exhibit*

23.5 - NSF Report]. To bridge this divide, initiatives such as the National AI Research Resource (NAIRR) from NSF have been launched to provide computing infrastructure and shared datasets for researchers in under-resourced institutions. Complementing these efforts, the 2025 Executive Order on Artificial Intelligence directed the Department of Energy to construct "advanced computing clusters needed to train AI models and the energy infrastructure needed to power this work" [*Exhibits 23.1 - Executive Order 14141*]. These actions underscore a national commitment to expanding AI accessibility in terms of computing resource demand.

Dr. X's contribution and impact: Dr. X's proposed endeavor directly advances this national goal by addressing the access divide through technological efficiency rather than infrastructure expansion. His prior research and proposed endeavor focus on developing cost-effective, domain-specific AI models and systems that minimize computational demands while maintaining high performance. By reducing the need for extensive GPU infrastructure, massive datasets, or costly commercial APIs, his approach makes cutting-edge AI development feasible for smaller research groups, nonprofit organizations, and local agencies. These organizations are vital to addressing issues such as public health, environmental sustainability, and community services but often remain excluded from AI innovation. For instance, Dr. X has developed AI models and systems for low-resource fields that operate efficiently on standard CPUs or a single GPU with minimal labeled data, avoiding the need for costly large-scale models that depend on multiple high-end GPUs or large annotated data. His emphasis on cost-effective AI design has benefited both academic fields with limited computing resources [e.g., *Exhibit 8.1; 8.4* - Research Publications], as well as the X industry, where his methods help substantially reduce computing and deployment costs [*Exhibit 6 - Supporting Letter from Company X*].

By lowering entry costs and computational barriers, Dr. X's endeavor broadens participation in AI research and application, amplifies the return on public and private investments in AI, and ensures that the benefits of technological progress are distributed more equitably across American society.

3.1.2 Domain adaptation

Equally critical to the national importance of Dr. X's endeavor is its emphasis on domain adaptation, which ensures that artificial intelligence serves highly specialized applications rather than remaining confined to generic tasks. Current general-purpose AI systems—such as large language models (LLMs) like ChatGPT—often struggle to capture the nuanced reasoning, factual precision, and contextual understanding required for rigorous scientific research, targeted policy analysis, and domain-specific decision-making.

Recognition of this issue by the U.S. government and other organizations: The risks posed by factual inaccuracy and hallucination in AI systems have drawn significant attention across governmental and professional fields, including but not limited to juridical judgment [*Exhibit 24.2 - NCSC Report*], law enforcement [*Exhibit 23.6 - DOJ Report*], medical practice [*Exhibit 24.1 - FSMB Report*], and e-commerce [*Exhibit 6 - Supporting Letter from Company X*], etc. Collectively, the cross-field concerns demonstrate a broad consensus that domain awareness and factual reliability are essential for the safe and effective use of AI in critical U.S. sectors.

Dr. X's contribution and impact: Dr. X's proposed endeavor directly addresses these challenges by embedding specialized domain knowledge into AI models and computational systems, thereby enhancing their contextual accuracy and reliability. This approach enables scientists, professionals, and policymakers to integrate AI into their workflows with greater confidence, ensuring that AI systems support—rather than compromise—expert decision-making. The endeavor aligns with the 2023 Executive Order on Safe, Secure, and Trustworthy AI [*Exhibit 23.2 – Executive Order 14110*], which emphasizes the responsible deployment of AI technologies in high-stakes domains. Dr. X's previous work has also applied this approach in multiple fields, including biodiversity conservation research [*Exhibits 8.1; 8.4* - Research Publications], impact assessment of funded projects [*Exhibits 8.7; 8.8* - Research Publications], and X industry [*Exhibits 6; 11; 8.6 - Supporting letter, Commercial Applications, and Research Article*].

Summary As AI becomes both more affordable and more contextually intelligent, its

adoption will extend across a broader range of U.S. fields and industries. This expansion strengthens the nation's research ecosystem, promotes evidence-based policy and innovation, and enhances industrial competitiveness. Equally important, it advances inclusive technological progress by enabling smaller, resource-constrained organizations—often closest to critical societal challenges—to develop and benefit from domain-adapted AI tools. Through this work, Dr. X's endeavor advances U.S. priorities of ensuring that artificial intelligence remains accurate, equitable, and beneficial across all sectors of society.

3.1.3 Methodological Generalizability and Cross-Disciplinary Influence

Although Dr. X's research projects focus on particular domains, the computational methods and frameworks he develops possess broad applicability across diverse fields, demonstrating strong methodological generalizability [*Exhibit 21* - *Citation Records*]. For instance, his expert-in-the-loop framework for X research—originally designed for biodiversity conservation data [*Exhibit 8.1* - *Research Publication*]—has been cited, adapted, or extended in multiple disciplines, including spatial data science, cognitive and mental health research, educational technology, quantitative science evaluation, and digital library indexing. His patent for X industry filed under Company X [*Exhibit 7* - *Proof of Patents*] has also been cited by the Bank of America.

This cross-disciplinary influence underscores that Dr. X's proposed endeavor, like his prior work, is not limited to a single scientific or industrial application. The underlying principles of domain knowledge integration, resource efficiency, and model adaptability can be applied across a wide spectrum of research and professional contexts. Consequently, the benefits of his approach—lower computational costs, improved factual reliability, and enhanced domain relevance—extend well beyond individual projects or datasets. By providing generalizable frameworks that different disciplines can adopt or tailor to their specific needs, Dr. X's research fosters methodological innovation across the U.S. research and development ecosystem. This amplifies the broader national impact of his proposed endeavor, reinforcing its contribution to scientific advancement, industrial competitiveness, and the responsible integration of AI throughout society.

3.2 AI as a Complement, Not a Substitute, for Skilled Professionals

A further dimension of national importance in Dr. X's endeavor lies in its explicit positioning of artificial intelligence as a complement to human expertise, rather than a replacement for it. His research advances a model of human–AI collaboration that preserves the central role of skilled professionals in guiding, validating, and improving AI systems.

U.S. government vision: This principle directly aligns with the 2023 Executive Order on Safe, Secure, and Trustworthy AI [*Exhibit 23.2 – Executive Order 14110*], which affirms that "the responsible development and use of AI require a commitment to supporting American workers." The Order emphasizes that as AI creates new jobs and industries, "all workers need a seat at the table" to ensure they share in the benefits of technological progress. Dr. X's work embodies this national vision by designing AI systems that enhance, rather than displace, the expertise of U.S. professionals.

Recognition of expertise across sectors: Governmental and professional organizations across critical sectors—such as the judiciary [*Exhibit 24.2 - NCSC Report*], law enforcement [*Exhibit 23.6 - DOJ Report*], healthcare [*Exhibit 24.1 - FSMB Report*], and education [*Exhibit 23.4 – Superintendent Report*]—have consistently emphasized the necessity of human involvement in AI applications. In these domains, experts are expected to validate and interpret AI-generated outputs rather than accept them passively, ensuring accuracy, fairness, and ethical compliance. These guidelines reflect a national consensus that domain expertise is indispensable for the responsible use of AI.

Dr. X's contribution and impact: Dr. X's proposed endeavor focuses on systematically incorporating expert knowledge into AI systems during the development phase, converting human expertise into forms that computational models can understand and learn from. This approach deepens collaboration between technical specialists and domain professionals, ensuring that AI systems are accurate, contextually relevant, and aligned with established professional standards. Far from diminishing the role of human experts, his methods underscore their importance: AI systems cannot function effectively in specialized contexts without the conceptual frameworks, curated data, and evaluative insights provided by domain

professionals. Dr. X's prior interdisciplinary collaborations and publications also underscore his recognition of the critical role that non-technical domain experts play in effective AI development. In these projects, he has consistently designed methodologies that integrate data mining, domain expertise, and machine learning to enhance the accuracy and applicability of AI models. A notable example of this approach is presented in his publication "Title" [*Exhibits 8.1 - Research Publication*]

Summary: By positioning AI as a tool that amplifies human expertise, Dr. X's endeavor promotes innovation while safeguarding the integrity of professional judgment. It also supports U.S. workforce development by fostering interdisciplinary collaboration and training, equipping researchers, policymakers, and practitioners to thrive in AI-enhanced environments, rather than replacing them. In doing so, his work contributes to a future in which AI expands the nation's productivity and scientific capacity while reinforcing—rather than eroding—the skilled workforce that underpins the U.S. economy.

3.3 Advancing Informed Decision-Making in Resource Allocation

A central focus of Dr. X's past and ongoing research is the development of cost-effective AI tools that automatically clean, index, categorize, summarize, and retrieve relevant information from large collections of domain-specific documents, such as project reports, research articles, and online discussions. These technologies enable government agencies, administrative staff, and researchers to efficiently extract and synthesize key information from massive unstructured data to understand public needs, identify completed work, assess outcomes, and evaluate impact within domain-specific frameworks. By automating the identification and organization of useful information, Dr. X's tools empower more effective and strategic resource allocation decisions. This approach strengthens the nation's capacity to assess the technological and societal impacts of public and private investments and supports federal and philanthropic priorities emphasizing accountability, transparency, and the optimal use of resources to achieve measurable, evidence-based benefits for society.

U.S. government vision: Ensuring that investment and resource allocation decisions gen-

erate tangible societal benefits has become a core priority for both federal agencies and major philanthropic organizations. The National Science Foundation (NSF), for example, evaluates proposals not only for their intellectual merit but also for their broader impacts—specifically, their potential "to benefit society and contribute to the achievement of specific, desired societal outcomes" [*Exhibit 23.3 – NSF Definition of Broader Impact*]. Similarly, leading private foundations such as the X Foundation promote impact investments that pursue measurable social and environmental outcomes alongside financial returns. These initiatives reflect a shared national commitment to developing robust impact assessment methods that guide resource allocation, evaluate outcomes, and maximize the societal value of public and private investments. However, because manually processing and analyzing large volumes of documents is time-consuming and labor-intensive, AI-assisted approaches to impact assessment are becoming a more effective way for ensuring that resources are deployed efficiently to advance technological innovation and serve the public good.

Dr. X's contribution and impact: Dr. X's work in this field would benefit the following sectors.

Academic research: Dr. X's previous and ongoing research directly contributes to this national vision by developing computational and AI-based models for large-scale assessment of projects in science, technology, and social-benefit domains [*Exhibits 8.7; 8.8; 9.2; 18 - Research Publications and Dissertation*]. Supported by <US major foundation> and <a government department in Europe> in two projects, he has collaborated with U.S. and international teams to evaluate the outcomes of initiatives in biodiversity conservation and scientific research funding. These interdisciplinary projects demonstrate his ability to apply AI and data-driven methods to improve the accountability and impact of global research and policy investments.

Private foundation: Beyond academic research, Dr. X creates practical AI tools that assist decision-makers in analyzing complex funding landscapes. For example, by developing models for information retrieval, text summarization, and data cleaning, he built a large-scale database mapping which organizations fund specific areas of conservation research. This system enabled the <US major foundation> to identify underfunded and oversaturated

research areas, allowing for more strategic and equitable funding decisions [*Exhibit 10 - Private Database*].

Government agency: Dr. X was also invited to contribute to a project supported by the <department of a state government>, which sought to use AI tools to understand residents' needs for X and improve resource allocation [*Exhibit 15 - Contract Invitation from a State Government Agency*]. Although work authorization restrictions prevented him from accepting the contract, the invitation underscores the practical relevance and policy value of his expertise.

Summary: Building on this foundation, Dr. X's proposed endeavor aims to develop domain-adapted AI systems capable of processing large volumes of domain-specific documents and data to provide policymakers, funding agencies, and research institutions with timely and efficient insights into the status and impact of their investment. These systems will help decision-makers plan future investments more strategically while ensuring that the outcomes of federally and privately funded programs remain transparent, measurable, and optimized for societal benefit. As discussed earlier, this approach is not intended to replace domain experts but rather to free them from the tedious, mechanical, and time-consuming tasks of locating and extracting useful information from vast collections of documents, allowing them to focus on higher-level analysis and decision-making. Through these contributions, Dr. X's endeavor advances the goal of maximizing the return on investments and resource allocation while promoting accountability, efficiency, and public trust in the nation's funding ecosystem.

3.4 Industrial and Economic Competitiveness

U.S. government vision: The competitiveness of the U.S. economy increasingly depends on its ability to develop and deploy accurate, efficient, and trustworthy AI systems across diverse industrial sectors. Federal agencies have explicitly recognized this imperative. A memo to the president from the Department of Commerce notes that "AI is poised to deliver significant productivity gains across U.S. industries" and calls for policies that strengthen the nation's industrial base for techno-industrial competition in AI [*Exhibit 23.7 - Department of*

Commerce Memo]. These priorities underscore the federal commitment to ensuring that AI innovation enhances productivity, sustainability, and global competitiveness.

Private-sector perspective: U.S. companies likewise view AI as a catalyst for innovation, efficiency, and market growth. <One of the largest companies> has publicly emphasized the strategic role of AI in "work in Industry X" [*Exhibit 24.3* - *Company Announcement*] The market's response to such initiatives has been strongly positive: following Company X's announcement of its AI program X in which Dr. X actively contributes—Company X's stock has reached an all-time high (exceeding xxx as of November 30, 2025)

Dr. X's contribution and impact: Dr. X's ongoing work at Company X directly advances U.S. competitiveness in both the industry X and artificial intelligence sectors.

Retail industry and national economy: As the nation's largest <industry x> and one of its largest employers, Company X plays a central role in the U.S. economy. By developing cost-effective, domain-tailored AI systems for X, Dr. X strengthens Company X's technological edge in areas such as xxx, yyy, and zzz [*Exhibit 6* - *Supporting Letter from Company X*]. Enhancing the competitiveness of such a cornerstone company yields broad economic benefits, such as improving efficiency across the xxx, encouraging innovation throughout the industry X, and contributing to national productivity and economic stability.

AI industry: Building on this foundation, Dr. X's work also supports the broader advancement of the U.S. AI ecosystem through strategic collaborations with leading technology partners, including OpenAI and NVIDIA [*Exhibit 6* - *Supporting Letter from Company X*]. In these partnerships, Dr. X helps articulate and translate the unique needs of the x sector into actionable directions for AI model design and deployment. This feedback loop not only accelerates the development of X-specific AI applications but also informs the design of next-generation, domain-adapted AI systems. By bridging applied industry contexts with frontier AI research, his work contributes to a more robust and diversified national AI landscape—one that prioritizes practical, trustworthy, and industry-specific innovation over generalized, one-size-fits-all models.

Workforce stability: Importantly, as described earlier, Dr. X's approach positions AI as a complement to human expertise rather than a substitute. His systems are designed to enhance

cross-team collaboration among product, safety, personalization, and customer service groups, allowing associates to focus on creative and strategic tasks while routine data processing is automated. This approach helps maintain workforce stability, reduces the need for large-scale layoffs, and demonstrates how AI can promote labor efficiency and upskilling rather than displacement—an outcome consistent with federal priorities for inclusive technological growth.

Intellectual properties and cross-sector industrial influence: Dr. X's research has resulted in two patents under Company X [*Exhibit 7 - Proof of Patents*], expanding the U.S. intellectual property base in applied artificial intelligence. One of these patents has already been cited by the Bank of America [*Exhibit 21 - Citation Report*], demonstrating the broad applicability of his innovations beyond the x sector. This cross-industry recognition highlights the far-reaching potential of his work to advance technological progress across diverse sectors, independent of any single employer or institutional affiliation. It also ensures that his proposed endeavor will continue to strengthen U.S. technological leadership and economic growth regardless of future professional transitions.

Summary: Collectively, Dr. X's work enhances Company X's competitiveness, strengthens the U.S. x sector and national economic stability, accelerates innovation within the national AI ecosystem, and expands the country's intellectual property portfolio. His contributions directly advance the U.S. national interest by fostering technological leadership, industrial modernization, and long-term economic resilience in the era of artificial intelligence.

4 Ability to Advance the Proposed Endeavor

4.1 Interdisciplinary Academic Training

Dr. X is well positioned to advance the proposed endeavor because his interdisciplinary academic training provides a rare, robust foundation for developing domain-specific, cost-effective computational and AI models with a human- and society-centered focus. He holds a Ph.D. in <cs degree> from the University X, a master's degree in <social science degree> from the University Y, and a bachelor's degree in philosophy from University Z[*Exhibit 2 - Diplomas*

and Transcripts].

This progression reflects a deliberate intellectual trajectory: (1) grounding in human reasoning and societal values, (2) methodological expertise for analyzing social needs, and (3) technical depth in AI and data science for building models and systems. His doctoral research at University X strengthened his command of machine learning and computational methods, while earlier training at University Y and Z developed sensitivity to the ethical, social, and cultural contexts in which technology operates. Accordingly, Dr. X approaches AI as a socio-technical system, which must integrate expert domain knowledge and consider public benefit. This formation directly equips him to design domain-specific models and systems that embed specialized knowledge, improve factual reliability, and reduce computational and financial costs, making advanced AI accessible to research institutions, industry, and nonprofit organizations. Taken together, this evidence shows that Dr. X's academic preparation uniquely positions him to advance the proposed endeavor in the national interest while maintaining both efficiency and technical rigor.

4.2 Interdisciplinary Collaboration Experience beyond Technical Fields and Academia

Dr. X's proposed endeavor to develop cost-effective, domain-adapted computational and AI models depends fundamentally on collaboration among technical experts, domain specialists, and organizational stakeholders. His record demonstrates extensive experience contributing to such interdisciplinary teams across academia, philanthropy, international research, and industry, effectively bridging computational innovation with domain-specific application.

Academic and Philanthropic Collaboration: In the X-supported project "Title" [*Exhibit 19.1 - Project Webpages*], Dr. X worked with colleagues from computer and information science, environmental and natural-resource sciences, and policy studies across U.S. and international research institutions, as well as partner organizations including the Nature Conservancy and the Gordon and Betty Moore Foundation. He contributed computational and AI expertise to design automated methods for processing conservation-related documents

and assessing the social and ecological impacts of philanthropic investments in this field. The project produced joint publications and conference presentations spanning conservation and information sciences, demonstrating his ability to integrate human-domain expertise with technical AI modeling to generate actionable insights. He further utilized AI-based models to process a large corpus of prior conservation publications and synthesize the results into a structured database that documented which conservation projects had been supported by specific research or philanthropic funding. This resource enabled the X Foundation and its partners to make more informed, data-driven decisions about future investments in biodiversity conservation and to evaluate the long-term impacts of their prior funding initiatives.

Academic and Industrial Research Collaboration: Dr. X also contributed to the project: "Title" [*Exhibit 19.2* - *Project Webpages*], funded by <a European government>, unites researchers from the <a European research institutes>, <the University X - my university in the US>, and the industrial partner <a European company>. The project's goal is to develop machine learning methods to analyze scientific texts and identify measurable economic, social, political-legal, and ecological impacts of research. Dr. X's role involved developing scalable computational frameworks that integrated linguistic knowledge with machine learning algorithms, helping to optimize the prototype system for cross-institutional use cases. This collaboration exemplifies his capacity to coordinate across international research and industry partners and to apply AI techniques to socially and scientifically relevant challenges.

Industrial Collaboration In his role as a Position X at Company X, Dr. X works closely with engineering, product, legal, personalization, and user-experience teams to develop large-scale conversational-AI systems to service tens of millions Company X's customers. These efforts require aligning technical modeling with consumer-experience goals, corporate policy, and compliance frameworks. He has played a key role in designing lightweight, cost-efficient AI models and systems that deliver accurate, user-friendly, and responsible conversational and interactions. His collaborative work has resulted in patent filings and real-world applications [*Exhibits 7; 11* - *Proof of Patents and Applications* (see details in the next section)]. The

company experience demonstrates his ability to translate interdisciplinary communication into commercially and socially valuable AI solutions.

Summary: Across academic, philanthropic, and industrial contexts, Dr. X has repeatedly shown the ability to connect computational modeling with domain-specific expertise to achieve results of tangible public and commercial value. These experiences confirm that he possesses not only advanced technical knowledge but also the collaborative capacity and cross-sector communication skills required to realize his proposed endeavor. Accordingly, the evidence establishes that Dr. X is well positioned to advance the proposed endeavor through effective interdisciplinary and multi-sector engagement.

4.3 Industry Experience

Since 2022, Dr. X has served as a Position X (intern, contractor, and then full-time employee) at Company X, where he has participated in applied research and development on cost-effective, X-adapted conversational-AI systems that assist millions of U.S. customers daily. His work translates academic innovation into large-scale, real-world implementation, demonstrating both technical mastery and the practical feasibility of his research agenda.

At Company X, Dr. X has contributed to question-answering and dialogue-management models that power the company's in-app customer-support chatbots [*Exhibit 11 – Chatbot Apps*]. These systems handle product-information queries, price and feature comparisons, and order-support questions, improving both accuracy and efficiency. He collaborated with engineering, product, legal-compliance, and personalization teams to ensure that the conversational systems deliver information that is not only relevant and factually correct but also aligned with consumer-protection and privacy requirements. This cross-functional engagement reflects his ability to integrate domain expertise and ethical oversight into technical design—an essential component of his proposed endeavor.

Dr. X also contributed to Company X's AI-First Experiences initiative, a flagship program conducted in partnership with OpenAI [*Exhibit 6 – Supporting Letter from Company X*]. In this collaboration, he helped build lightweight, domain-specific model architectures that provide

accurate, context-aware recommendations and conversational assistance while substantially reducing computational costs. His focus on efficiency and domain adaptation directly aligns with the central aims of his proposed endeavor to create AI systems that are both technically robust and economically sustainable.

These developments have already resulted in two patent filings naming Dr. X as inventor (see details in next section). Through these industry accomplishments, Dr. X has demonstrated his ability to lead and deliver high-impact AI innovations that advance both corporate and societal interests. His work shows that domain-adapted, cost-efficient AI models can operate successfully at enterprise scale, benefiting U.S. consumers and strengthening the country's technological leadership. Accordingly, the evidence establishes that Dr. X is well positioned to advance the proposed endeavor through continued research and deployment of scalable, human-centered AI systems.

4.4 Patent, Publications, and Applications

4.4.1 Patent

Dr. X's inventive work has resulted in two patent filings with Company X [*7 - Proof of Patent Filling*], directly supporting his proposed endeavor to create domain-adapted, cost-effective conversational-AI systems. These filings exemplify his ability to transform research concepts into scalable technologies that improve efficiency, accuracy, and accessibility of AI for millions of U.S. consumers.

Patent name 1

(U.S. Patent Application Publication No. xxx; inventors: Dr X, other people)

This invention describes an architecture for domain-specific question-answering within chatbot systems. It enables xxx, while minimizing computational cost through optimized model selection and adaptive inference. The approach directly advances Dr. X's central research goal of integrating specialized knowledge into AI models to enhance factual reliability and cost performance.

Patent name 2

(U.S. Patent Application No. yyy; inventors: yyy)

This invention presents a lightweight, cost-effective decision-assistant framework for X. Recognizing that large language models (LLMs) often contain memorized but outdated or inconsistent knowledge, Dr. X and his co-inventors developed a system that integrates xxx. The framework demonstrated competitive performance in answer quality while using far fewer computational resources. This patent exemplifies Dr. X's ability to create practical, domain-specific AI systems that balance accuracy, interpretability, and cost efficiency, directly advancing his proposed endeavor to make AI more accessible and reliable across U.S. industries.

4.4.2 Research Publications and its Resulting Applications

Publications and their topics: Dr. X has published seven papers, including four as first or co-first author, which have collectively received xxx citations [*Exhibit 4 - Google Scholar Profile*]. Three additional papers, all with Dr. X as first or co-first author, are currently under review. His publications span multiple interdisciplinary fields [*Exhibit 8 - Research Articles*]. The main research topics include:

- Integrating domain expertise into AI systems: Developing methods to incorporate expert knowledge into AI models, with applications in biodiversity conservation, scientific topic mining, social impact assessment, and e-commerce.
- Cost-effective AI development: Proposing efficient techniques for AI model training using small datasets, lightweight large language models, and AI-assisted methods to reduce reliance on costly human annotation.
- Automated resource and data curation: Building AI-driven systems that automatically collect and organize domain-specific data to support model training, enhance AI performance, and enable deeper domain analysis.

These research papers have also been presented at several peer-reviewed conferences or workshops, including xxx, yyy, zzz [*Exhibit 12 - Proof of Presentations*].

Cross-Disciplinary Influence: As noted earlier (see ***Section 3.1.3***), these citations extend beyond the immediate research domains of his papers to fields such as spatial data science, cognitive and mental health research, educational technology, quantitative science evaluation, and digital library indexing, underscoring the broad relevance and interdisciplinary impact of his work.

Practical applications: These research efforts have also produced several public and private datasets and models with practical applications:

- A public hierarchical concept dictionary that enables AI models to classify, index, and analyze both established and emerging topics in computer science research. This resource has been downloaded more than 1,400 times [*Exhibit 9.1*].
- A public dataset of cleaned and annotated project reports with xxx for xxx. It has been downloaded more than 500 times [*Exhibit 9.2*].
- An internal database mapping which field X studies were funded by specific organizations, developed to assist Foundation X in making informed decisions about future resource allocations [*Exhibit 10*].
- Conversational AI models that has been deployed in the Company X apps to enhance user interaction and customer service [*Exhibit 11*].

Summary: Collectively, Dr. X's publications and its resulting applications demonstrate his exceptional ability to translate advanced AI methodologies into impactful, real-world solutions. His peer-reviewed publications establish a strong theoretical and technical foundation, while the associated datasets, models, and tools illustrate his capacity to deliver tangible benefits to both the research community and major institutions. The wide adoption and cross-sector relevance of these outputs—from academic use to deployment in large-scale industry environments—show that Dr. X possesses the expertise, credibility, and practical experience necessary to successfully execute his proposed endeavor. These accomplishments position him uniquely well to continue advancing AI technologies that enhance efficiency, transparency, and societal benefit across U.S. industries.

4.5 Academic Reviewing Service

Dr. X has provided peer-review service for 31 academic papers in artificial intelligence and interdisciplinary domains directly related to his proposed endeavor [*Exhibit 22 – Review Record*]. His reviewing activities span top-tier conferences and journals, demonstrating recognition of his expertise by leading scholarly communities. His record includes:

22 papers reviewed for the ACL Rolling Review, the unified evaluation platform supporting the most prestigious NLP conferences.

2 papers for the ACL Student Research Workshop.

1 paper for the Workshop on Large Language Model Memorization.

2 papers for the International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

1 paper for the International Conference on Complex Networks and Their Applications.

1 paper for the journal Information Processing and Management.

1 paper for the journal Conservation Science and Practice.

1 paper for the journal Computers & Industrial Engineering.

His service record further evidences his standing as a trusted expert in the field and his active participation in shaping the direction of AI and computational research globally and interdisciplinarily.

4.6 Invited Talk, Book Chapter, Contract, and Service

Dr. X's professional recognition extends beyond publications, patents and apps to invitations from academic, governmental, and industry partners to contribute his expertise in applied artificial intelligence and computational modeling. These engagements provide further evidence that he is well positioned to advance his proposed endeavor through continued interdisciplinary collaboration and knowledge dissemination.

Invited book chapters: Dr. X was invited by the editorial board of the Encyclopedia of Social Network Analysis and Mining (3rd Edition) to contribute one or more chapters [*Exhibit 13 – Proof of Invitation*].

Invited talk: He was invited to serve as a speaker at an international conference on X [*Exhibit 14 – Proof of Invitation*].

Government-related contractor invitation: Dr. X was invited by a professor at a South Dakota state university to collaborate as a research contractor for a project from <state government department> [*Exhibit 15 - Proof of Invitation*]. The project seeks to employ AI technologies to synthesize information, identify public needs related to X, and support data-driven decision-making. This invitation evidences the governmental demand for his expertise and the public value of his research in practical policy contexts.

Industry-related contractor invitation: After completing his first internship at Company X, Dr. X was invited to continue contributing as a remote data scientist contractor [*Exhibit 16 – Proof of Invitation*], even while not completing his Ph.D. studies. This invitation reflects the company's recognition of his technical expertise and the direct value of his research to ongoing industrial AI initiatives.

Dissertation committee invitation: Dr. X was also invited by a doctoral student to serve on his dissertation committee [*Exhibit 17 – Proof of Invitation*]. Such invitations demonstrate that peers within academia regard Dr. X as a qualified expert capable of guiding original research in AI and computational modeling.

Summary: Collectively, these invitations confirm that Dr. X's expertise is both recognized and sought after across academic, governmental, and professional communities. They further demonstrate that he is well positioned to advance his proposed endeavor, as institutions continue to engage his insight and collaboration in projects of scientific, educational, and public importance.

5 Justifications for the Job Offer and Labor Certification Waiver

Under Matter of Dhanasar, the final prong requires demonstrating that waiving the job offer and labor certification requirements would, on balance, better serve the national interest. While the petitioner could theoretically pursue employer-sponsored permanent residence, the structure and purpose of the PERM system are incompatible with the petitioner's interdisciplinary, cross-sector, and public-serving contributions. The national interest is best advanced by granting the petitioner the flexibility necessary to collaborate across institutional boundaries, participate in public-interest research, and respond to evolving priorities in AI.

5.1 The Proposed Endeavor Requires Mobility Across Institutions and Sectors

A central reason the job offer and labor certification requirements should be waived is that the petitioner's proposed endeavor depends on continuous collaboration across academia, industry, public-interest organizations, and government-affiliated partners. AI research—especially in areas involving domain adaptation, model alignment, evaluation, and public-interest applications—advances through multi-institutional cooperation rather than the work of a single employer or laboratory. The petitioner's contributions to domain-specific model and system design, domain-specific knowledge curation, and socially beneficial AI applications require involvement in a distributed ecosystem of stakeholders, each providing different resources, data contexts, and domain expertise. The flexibility to work with these diverse collaborators is essential for maximizing the national benefit of the petitioner's work.

Because PERM binds a beneficiary to one employer and one permanent job description, it cannot accommodate the level of cross-sector engagement that the petitioner's endeavor requires. Without the flexibility to collaborate freely across institutions, many of the petitioner's contributions would be curtailed, diminishing the United States' ability to leverage Dr X's expertise that advances cost-effective, domain-specific AI systems and benefits multiple

sectors simultaneously.

5.2 The Interdisciplinary Nature of the Work Exceeds the Scope of a Single Employer

The petitioner's work is inherently interdisciplinary, spanning computer science, information science, social science, and human-centered computing. These areas reside in different academic and practical domains and must be synthesized continuously as the field evolves. As detailed in the publications and projects described in earlier sections, the petitioner has consistently conducted work that bridges multiple sectors, including academic research, nonprofit initiatives, and industry. This established record demonstrates that the petitioner's contributions naturally arise at the intersection of technical, social, and applied domains and require collaboration across institutional boundaries. The petitioner's past accomplishments therefore provide concrete evidence that advancing the proposed endeavor demands professional flexibility not achievable within a single employer's structure.

PERM, however, certifies a beneficiary for a single set of permanent job duties within a narrowly defined occupational category. Such a structure does not always allow for evolving, interdisciplinary roles that require shifting between interdisciplinary fields and tasks. The petitioner's proposed endeavor cannot be meaningfully advanced within a single employer-defined job classification, making a national interest waiver more appropriate than employer-bound sponsorship.

5.3 Enabling the Petitioner's Contributions to Both Commercial and Public-Interest Sectors Through a National Interest Waiver

Many of the organizations that benefit from the petitioner's expertise cannot feasibly support employer sponsored immigration petitions. Nonprofits, small research groups, foundations, and state affiliated public interest initiatives often operate on grant based or project driven

timelines and lack the administrative capacity, financial resources, or institutional mandate to sponsor long term, employer specific PERM applications.

At the same time, large technology-driven employers, while understandably oriented toward commercial objectives, play an essential role in advancing the frontiers of AI by providing substantial job opportunities, funding, compute resources, and technical infrastructure that enable rapid experimentation and the development of cutting-edge research.

The petitioner's interdisciplinary background, which integrates technical modeling, social science informed evaluation, and public impact research, positions him to contribute meaningfully within both commercially oriented settings that drive frontier innovation and public serving contexts where AI can generate broad societal value. A national interest waiver would allow the petitioner to engage fully across this entire spectrum, ensuring that his expertise can be deployed wherever it most benefits the United States, including in sectors that cannot sponsor employer based immigration petitions as well as in well resourced research environments that accelerate scientific and technological progress

5.4 Illustrative Example of Constraints Imposed by Certain Work Restrictions

The limitations of employer-tied authorization are not theoretical. While Dr. X was a graduate student on an F-1 visa, he was invited to serve as a research contractor for X, a state agency seeking to apply AI techniques to synthesize public feedback, identify community needs related to X, and support evidence-based resource management [*Exhibit 15 - Proof of Invitation*]. This invitation demonstrates both the governmental demand for Dr. X's interdisciplinary expertise and the practical policy value of his work in data-driven public decision-making. However, because F-1 regulations restrict graduate students to a maximum of 20 hours per week of combined authorized employment during the academic term, and Dr. X was already engaged in part-time work through his university and a separate authorized contractor role at Company X, he could not accept the additional position. This lost opportunity illustrates how visa-based employment limitations can prevent qualified experts

from contributing to public-serving, government-affiliated initiatives.

Under the PERM framework, similar barriers would persist, as the beneficiary is tied to one sponsoring employer and possibly unable to engage flexibly with state agencies, nonprofits, and other public-interest institutions that frequently depend on short-term or project-based collaboration. Granting a National Interest Waiver would remove these structural impediments and enable Dr. X to collaborate freely with the diverse range of institutions advancing responsible AI development and data-informed public policy in the United States.

5.5 Conclusion

Taken together, these factors demonstrate that the petitioner's ability to serve the national interest depends on professional flexibility that the PERM process cannot provide. Waiving the job offer and labor certification requirements would enable the petitioner to collaborate across institutions, contribute to public-benefit initiatives, respond rapidly to emerging challenges in AI, and continue advancing work that the United States has identified as strategically important. Because employer-sponsored immigration would constrain, rather than facilitate, the petitioner's nationally beneficial contributions, granting a national interest waiver would, on balance, better serve the interests of the United States.

List of Exhibits

Academic, Industrial, and Professional Background

Exhibit 1: **Curriculum Vitae**

Exhibit 2: **Degree Diplomas and Transcripts**

Exhibit 3: **Membership Cards for Professional Associations**

Exhibit 4: **Google Scholar Profile**

Exhibit 5: **Proof of Work**

Supporting Letter

Exhibit 6: **Supporting Letter from Company X Manager**

Patents

Exhibit 7: **Proof of Published and Filed Patents**

Scientific Publications and Articles

Exhibit 8: **First Few Pages of Research Articles Authored by Dr. X**

First or co-first authorship - ** indicates co-first authorship.*

8.1 xxx

8.2 xxx

8.3 xxx

8.4 xxx

8.5 xxx

8.6 xxx

8.7 xxx

Other authorship

8.8 xxx

8.9 xxx

8.10 xxx

Developed Data, Apps, and Resources

Exhibit 9: **Publicly accessible resources**

9.1 xxx

9.2 xxx

Exhibit 10: **Private Database for Funding Decision-Making.**

This is a private database designed for Foundation X to explore X and then to make better funding decisions in the future. The attached are the database demo and the email communications with staffers from Foundation X.

Exhibit 11: **Apps that Dr. X is working on at Company X**

Peer-Reviewed Conference Presentations

Exhibit 12: **Proof of Conference/Workshop Presentations**

Academic and Professional Invitations

Exhibit 13: **Invited Book Chapters** of the 3rd Edition of the Encyclopedia on Social Network Analysis and Mining.

Exhibit 14: **Invited Talk** from xxx

Exhibit 15: **Contract Invitation - State Government Agency**

Exhibit 16: **Industrial Contract Invitation**

Exhibit 17: **Invitation to a PhD Dissertation Committee**

Other Materials

Exhibit 18: **Dissertation** (first 7 pages)

Exhibit 19: **Project Description**

19.1: xxx

19.2: xxx

Exhibit 20: **Graduate School rankings**

Exhibit 21: **Cross-disciplinary Citations**

Dr. X's paper that applies AI into x research has been cited, adapted, or extended in multiple disciplines, including spatial data science, cognitive and mental health research, educational technology, quantitative science evaluation, and digital library indexing, etc. And his patent for e-commerce has also been cited by the Bank of America.

The Patent filed by Company X has been cited by the Bank of America.

The AI application in x research has been cited by papers in a variety of domains.

Exhibit 22: **Proof of 30 Academic Manuscript Reviews**

Cited Documents

Exhibit 23: **Cited Government Documents**

23.1: Executive Order on Advancing United States Leadership in Artificial Intelligence Infrastructure (first two sections)

Federal Register
Vol. 90, No. 11
Friday, January 17, 2025

Presidential Documents

Title 3—

The President

Executive Order 14141 of January 14, 2025

Advancing United States Leadership in Artificial Intelligence Infrastructure

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. *Purpose*. Artificial intelligence (AI) is a defining technology of our era. Recent advancements in AI demonstrate its rapidly growing relevance to national security, including with respect to logistics, military capabilities, intelligence analysis, and cybersecurity. Building AI in the United States will help prevent adversaries from gaining access to, and using, powerful future systems to the detriment of our military and national security. It will also enable the United States Government to continue harnessing AI in service of national-security missions while preventing the United States from becoming dependent on other countries' infrastructure to develop and operate powerful AI tools.

Advances at the frontier of AI will also have significant implications for United States economic competitiveness. These imperatives require building AI infrastructure in the United States on the time frame needed to ensure United States leadership over competitors who, already, are racing to take the lead in AI development and adoption. Building AI in the United States requires enormous private-sector investments in infrastructure, especially for the advanced computing clusters needed to train AI models and the energy infrastructure needed to power this work. Already, AI's electricity and computational needs are vast, and they are set to surge in the years ahead. This work also requires secure, reliable supply chains for critical components needed to build AI infrastructure, from construction materials to advanced electronics.

This order sets our Nation on the path to ensure that future frontier AI can, and will, continue to be built here in the United States. In building domestic AI infrastructure, our Nation will also advance its leadership in the clean energy technologies needed to power the future economy, including geothermal, solar, wind, and nuclear energy; foster a vibrant, competitive, and open technology ecosystem in the United States, in which small companies can compete alongside large ones; maintain low consumer electricity prices; and help ensure that the development of AI infrastructure benefits the workers building it and communities near it.

With this order, I provide a plan for protecting national security, preserving our economic competitiveness, revitalizing our energy infrastructure, and ensuring United States leadership in AI.

Sec. 2. *Policy*. It is the policy of the United States to enable the development and operation of AI infrastructure, including data centers, in the United States in accordance with five guiding principles. When undertaking the actions set forth in this order, executive departments and agencies (agencies) shall adhere to these principles, as appropriate and consistent with applicable law:

(a) The development of AI infrastructure should advance United States national security and leadership in AI. Meeting this goal will require steps by the Federal Government, in collaboration with the private sector, to advance AI development and use AI for future national-security missions, including through the work described in National Security Memorandum 25 of October 24, 2024 (Advancing the United States' Leadership in Artificial

Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence) (NSM–25). It will also require the use of safeguards to improve the cyber, supply-chain, and physical security of the laboratories at which powerful AI is developed, stored, and used. Additionally, protecting United States national security will require further work to evaluate and manage risks related to the powerful capabilities that future frontier AI may possess.

(b) The development of AI infrastructure should advance United States economic competitiveness, including by fostering a vibrant technology ecosystem. Already, AI is creating new jobs and industries, and its effects are being felt in sectors across the economy. The Federal Government must ensure that the United States remains competitive in the global economy, including through harnessing the benefits of this technology for all Americans. It must also promote a fair, open, and competitive AI ecosystem so that small developers and entrepreneurs can continue to drive innovation—a priority highlighted in both Executive Order 14110 of October 30, 2023 (Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence), and NSM–25—as well as to support secure, reliable supply-chain infrastructure for AI activities.

(c) The United States can and should lead the world in operating the next generation of AI data centers with clean power. Meeting this goal will require building on recent successes to modernize our Nation's energy infrastructure; improve permitting processes; and support investments in, and expeditious development of, both currently available and emerging clean energy technologies, such as geothermal energy, nuclear energy, and long-duration energy storage used to store clean energy, as well as relevant supply chains. The United States must not be surpassed in its support for the development, commercialization, and operation of clean energy technologies at home and abroad, and the rapid buildout of AI infrastructure offers another vital opportunity to accelerate and deploy these energy technologies. To help ensure that new data center electricity demand does not take clean power away from other end users, result in resource adequacy issues, or increase grid emissions, the construction of AI infrastructure must be matched with new, clean electricity generation resources.

(d) The development of AI infrastructure should proceed without raising energy costs for American consumers and businesses, and it should have strong community support. The companies developing, commercializing, and deploying AI must finance the cost of building the infrastructure needed for AI operations, including the development of next-generation power infrastructure built for these operations.

(e) The development of AI infrastructure should benefit those working to build it. Meeting this goal will require high labor standards and safeguards for the buildout of AI infrastructure, consultation and close collaboration with communities affected by this infrastructure's development and operation, and continuous work to mitigate risks and potential harms. The American people more broadly must safely enjoy the gains and opportunities from technological innovation in the AI ecosystem.

Sec. 3. *Definitions*. For purposes of this order:

(a) The term ''agency'' means each agency described in 44 U.S.C. 3502(1), except for the independent regulatory agencies described in 44 U.S.C. 3502(5).

(b) The term ''AI data center'' means a data center used primarily with respect to developing or operating AI.

(c) The term ''AI infrastructure'' refers collectively to AI data centers, generation and storage resources procured to deliver electrical energy to data centers, and transmission facilities developed or upgraded for the same purpose.

23.2: Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (first two sections)

Federal Register
Vol. 88, No. 210
Wednesday, November 1, 2023

Presidential Documents

Title 3—

The President

Executive Order 14110 of October 30, 2023

Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. *Purpose.* Artificial intelligence (AI) holds extraordinary potential for both promise and peril. Responsible AI use has the potential to help solve urgent challenges while making our world more prosperous, productive, innovative, and secure. At the same time, irresponsible use could exacerbate societal harms such as fraud, discrimination, bias, and disinformation; displace and disempower workers; stifle competition; and pose risks to national security. Harnessing AI for good and realizing its myriad benefits requires mitigating its substantial risks. This endeavor demands a society-wide effort that includes government, the private sector, academia, and civil society.

My Administration places the highest urgency on governing the development and use of AI safely and responsibly, and is therefore advancing a coordinated, Federal Government-wide approach to doing so. The rapid speed at which AI capabilities are advancing compels the United States to lead in this moment for the sake of our security, economy, and society.

In the end, AI reflects the principles of the people who build it, the people who use it, and the data upon which it is built. I firmly believe that the power of our ideals; the foundations of our society; and the creativity, diversity, and decency of our people are the reasons that America thrived in past eras of rapid change. They are the reasons we will succeed again in this moment. We are more than capable of harnessing AI for justice, security, and opportunity for all.

Sec. 2. *Policy and Principles.* It is the policy of my Administration to advance and govern the development and use of AI in accordance with eight guiding principles and priorities. When undertaking the actions set forth in this order, executive departments and agencies (agencies) shall, as appropriate and consistent with applicable law, adhere to these principles, while, as feasible, taking into account the views of other agencies, industry, members of academia, civil society, labor unions, international allies and partners, and other relevant organizations:

(a) Artificial Intelligence must be safe and secure. Meeting this goal requires robust, reliable, repeatable, and standardized evaluations of AI systems, as well as policies, institutions, and, as appropriate, other mechanisms to test, understand, and mitigate risks from these systems before they are put to use. It also requires addressing AI systems' most pressing security risks—including with respect to biotechnology, cybersecurity, critical infrastructure, and other national security dangers—while navigating AI's opacity and complexity. Testing and evaluations, including post-deployment performance monitoring, will help ensure that AI systems function as intended, are resilient against misuse or dangerous modifications, are ethically developed and operated in a secure manner, and are compliant with applicable Federal laws and policies. Finally, my Administration will help develop effective labeling and content provenance mechanisms, so that Americans are able to determine when content is generated using AI and when it is not. These actions will provide a vital foundation for an approach that addresses AI's risks without unduly reducing its benefits.

(b) Promoting responsible innovation, competition, and collaboration will allow the United States to lead in AI and unlock the technology's potential to solve some of society's most difficult challenges. This effort requires investments in AI-related education, training, development, research, and capacity, while simultaneously tackling novel intellectual property (IP) questions and other problems to protect inventors and creators. Across the Federal Government, my Administration will support programs to provide Americans the skills they need for the age of AI and attract the world's AI talent to our shores—not just to study, but to stay—so that the companies and technologies of the future are made in America. The Federal Government will promote a fair, open, and competitive ecosystem and marketplace for AI and related technologies so that small developers and entrepreneurs can continue to drive innovation. Doing so requires stopping unlawful collusion and addressing risks from dominant firms' use of key assets such as semiconductors, computing power, cloud storage, and data to disadvantage competitors, and it requires supporting a marketplace that harnesses the benefits of AI to provide new opportunities for small businesses, workers, and entrepreneurs.

(c) The responsible development and use of AI require a commitment to supporting American workers. As AI creates new jobs and industries, all workers need a seat at the table, including through collective bargaining, to ensure that they benefit from these opportunities. My Administration will seek to adapt job training and education to support a diverse workforce and help provide access to opportunities that AI creates. In the workplace itself, AI should not be deployed in ways that undermine rights, worsen job quality, encourage undue worker surveillance, lessen market competition, introduce new health and safety risks, or cause harmful labor-force disruptions. The critical next steps in AI development should be built on the views of workers, labor unions, educators, and employers to support responsible uses of AI that improve workers' lives, positively augment human work, and help all people safely enjoy the gains and opportunities from technological innovation.

(d) Artificial Intelligence policies must be consistent with my Administration's dedication to advancing equity and civil rights. My Administration cannot—and will not—tolerate the use of AI to disadvantage those who are already too often denied equal opportunity and justice. From hiring to housing to healthcare, we have seen what happens when AI use deepens discrimination and bias, rather than improving quality of life. Artificial Intelligence systems deployed irresponsibly have reproduced and intensified existing inequities, caused new types of harmful discrimination, and exacerbated online and physical harms. My Administration will build on the important steps that have already been taken—such as issuing the Blueprint for an AI Bill of Rights, the AI Risk Management Framework, and Executive Order 14091 of February 16, 2023 (Further Advancing Racial Equity and Support for Underserved Communities Through the Federal Government)—in seeking to ensure that AI complies with all Federal laws and to promote robust technical evaluations, careful oversight, engagement with affected communities, and rigorous regulation. It is necessary to hold those developing and deploying AI accountable to standards that protect against unlawful discrimination and abuse, including in the justice system and the Federal Government. Only then can Americans trust AI to advance civil rights, civil liberties, equity, and justice for all.

(e) The interests of Americans who increasingly use, interact with, or purchase AI and AI-enabled products in their daily lives must be protected. Use of new technologies, such as AI, does not excuse organizations from their legal obligations, and hard-won consumer protections are more important than ever in moments of technological change. The Federal Government will enforce existing consumer protection laws and principles and enact appropriate safeguards against fraud, unintended bias, discrimination, infringements on privacy, and other harms from AI. Such protections are

especially important in critical fields like healthcare, financial services, education, housing, law, and transportation, where mistakes by or misuse of AI could harm patients, cost consumers or small businesses, or jeopardize safety or rights. At the same time, my Administration will promote responsible uses of AI that protect consumers, raise the quality of goods and services, lower their prices, or expand selection and availability.

(f) Americans' privacy and civil liberties must be protected as AI continues advancing. Artificial Intelligence is making it easier to extract, re-identify, link, infer, and act on sensitive information about people's identities, locations, habits, and desires. Artificial Intelligence's capabilities in these areas can increase the risk that personal data could be exploited and exposed. To combat this risk, the Federal Government will ensure that the collection, use, and retention of data is lawful, is secure, and mitigates privacy and confidentiality risks. Agencies shall use available policy and technical tools, including privacy-enhancing technologies (PETs) where appropriate, to protect privacy and to combat the broader legal and societal risks—including the chilling of First Amendment rights—that result from the improper collection and use of people's data.

(g) It is important to manage the risks from the Federal Government's own use of AI and increase its internal capacity to regulate, govern, and support responsible use of AI to deliver better results for Americans. These efforts start with people, our Nation's greatest asset. My Administration will take steps to attract, retain, and develop public service-oriented AI professionals, including from underserved communities, across disciplines—including technology, policy, managerial, procurement, regulatory, ethical, governance, and legal fields—and ease AI professionals' path into the Federal Government to help harness and govern AI. The Federal Government will work to ensure that all members of its workforce receive adequate training to understand the benefits, risks, and limitations of AI for their job functions, and to modernize Federal Government information technology infrastructure, remove bureaucratic obstacles, and ensure that safe and rights-respecting AI is adopted, deployed, and used.

(h) The Federal Government should lead the way to global societal, economic, and technological progress, as the United States has in previous eras of disruptive innovation and change. This leadership is not measured solely by the technological advancements our country makes. Effective leadership also means pioneering those systems and safeguards needed to deploy technology responsibly—and building and promoting those safeguards with the rest of the world. My Administration will engage with international allies and partners in developing a framework to manage AI's risks, unlock AI's potential for good, and promote common approaches to shared challenges. The Federal Government will seek to promote responsible AI safety and security principles and actions with other nations, including our competitors, while leading key global conversations and collaborations to ensure that AI benefits the whole world, rather than exacerbating inequities, threatening human rights, and causing other harms.

Sec. 3. *Definitions.* For purposes of this order:

(a) The term ''agency'' means each agency described in 44 U.S.C. 3502(1), except for the independent regulatory agencies described in 44 U.S.C. 3502(5).

(b) The term ''artificial intelligence'' or ''AI'' has the meaning set forth in 15 U.S.C. 9401(3): a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action.

(c) The term ''AI model'' means a component of an information system that implements AI technology and uses computational, statistical, or machine-learning techniques to produce outputs from a given set of inputs.

23.3: National Science Foundation on Broader Impact

U.S. National Science Foundation

Updates to NSF Research Security Policies
On July 10, 2025, NSF issued an **Important Notice** providi...

The broader impacts discussion is a critical component of any proposal submitted to the U.S. National Science Foundation. It answers the following question: How does your research benefit society?

On this page

What are broader impacts?

NSF funds scientists and engineers to perform research that advances discovery and innovation. The agency also expects researchers' work to have *broader impacts*: the potential to benefit society and contribute to the achievement of specific, desired societal outcomes.

NSF does not want to be prescriptive about the societal outcomes a project addresses. Examples of desired outcomes **include, but aren't limited to**:

STEM education

Improving education and educator development — at any level — in science, technology, engineering and mathematics.

Public engagement

Increasing public scientific literacy and public engagement with STEM.

Societal well-being

Improving the well-being of individuals in society.

STEM workforce

Partnerships

Building partnerships between

National security

Developing a more diverse, globally competitive STEM workforce.

academia, industry and others.

Improving national security.

Economic competitiveness

Increasing the economic competitiveness of the U.S.

Infrastructure

Enhancing infrastructure for research and education.

Why does NSF focus on broader impacts?

By evaluating every proposal it receives according to its intellectual merit and its broader impacts, NSF ensures that publicly funded research has tangible benefits to society that go beyond increasing knowledge.

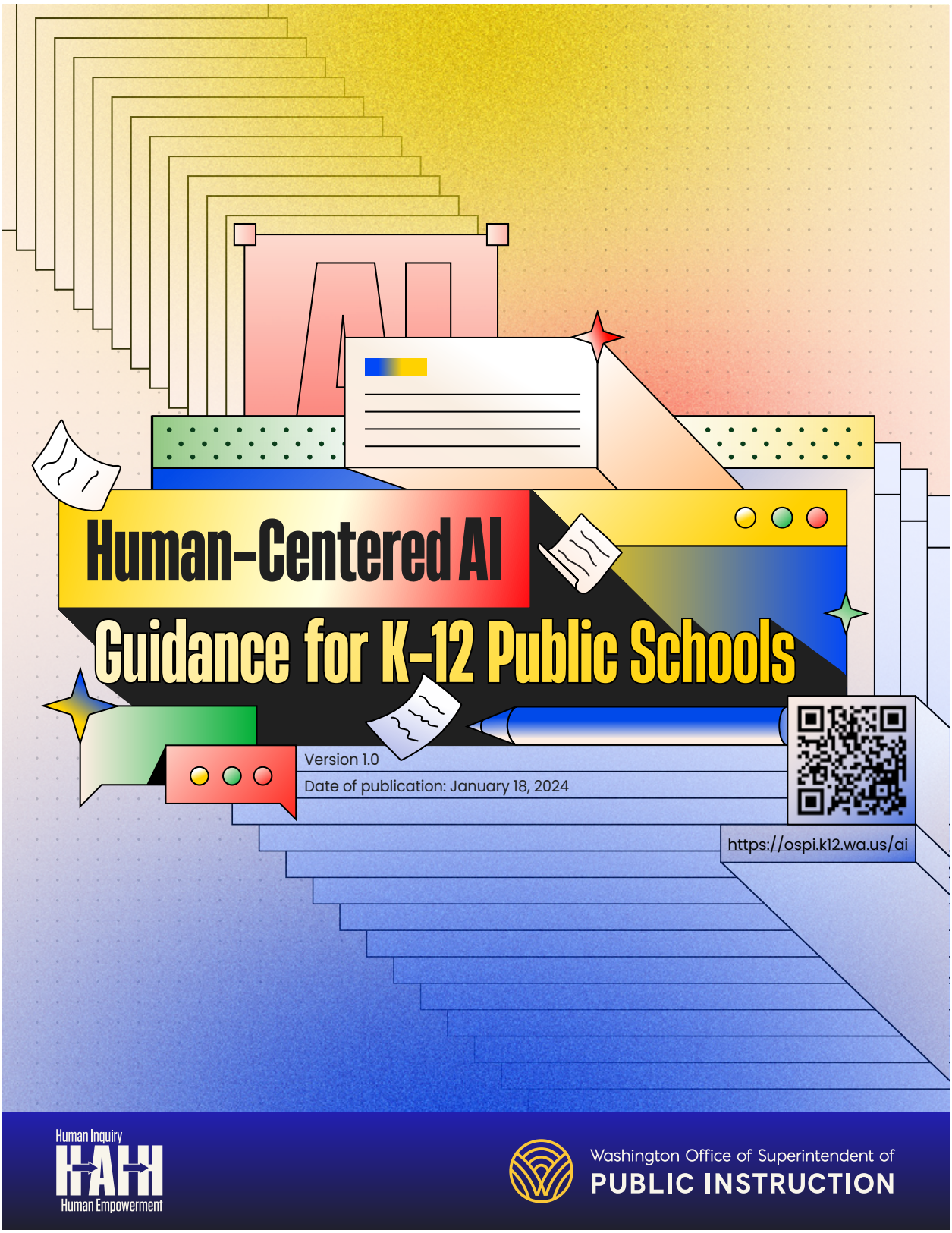
Explore examples of NSF-funded projects' broader impacts below:

Researchers work in the cell manufacturing laboratory of Krishnendu Roy at Georgia Tech. Shown, left to right, are NSF Graduate Research Fellow Joscelyn Mejias, Research Experience for Undergraduates (REU) program student Angela Jimenez, (background) postdoctoral fellow Randall Toy, Georgia Tech Research Institute TAG-Ed High School Intern Gita Balakirsky, and project ENGAGES high school intern Ayanna Prather.

Building STEM talent

From training the next generation of **cutting-edge materials researchers** to encouraging **entrepreneurial thinking**

23.4: Washington office of superintendent: Human-centered AI guidance for K-12 Public School (selected pages)



Human Inquiry
HAH
Human Empowerment

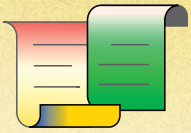
AI in Education: A Human-Centered Approach

A human-centered AI learning environment is one that prioritizes the needs, abilities, and experiences of students, educators, and administrators. An educational leader can support a human-centered learning environment by considering the following:

- **Developing students' AI literacy** by helping them understand the concepts, applications, and implications of AI in various domains, and empowering them to use AI as a tool for learning and problem-solving.
- **Ensuring ethical, equitable, and safe use of AI** by protecting the privacy and security of data, addressing potential biases and harms, and promoting digital citizenship and responsibility.
- **Providing professional development** and support for educators by helping them integrate AI into their pedagogy, curriculum, and assessment, and by facilitating their collaboration and innovation with AI.
- **Applying human-centered design principles** to the development and implementation of AI solutions, such as involving stakeholders in the design process, testing and iterating the solutions, and evaluating the impact and outcomes.
- **Aligning AI solutions with the best practices and principles of learning,** such as supporting learner agency, fostering collaboration, enhancing feedback, and promoting critical thinking.

Human Inquiry
HAH
Human Empowerment

Guidance



A Human-Centered Approach to AI

A human-centered AI learning environment always starts with human inputs and inquiry, and always concludes with human reflection and edits. It should prioritize the needs, abilities, and experiences of students, teachers, and administrators. An education leader can support a human-centered learning environment by considering the following:

- **Developing students' AI literacy** by helping them understand the concepts, applications, and implications of AI in various domains, and empowering them to use AI as a tool for learning and problem-solving.
- **Ensuring ethical, equitable, and safe use of AI** by protecting the privacy and security of data, addressing potential biases and harms, and promoting digital citizenship and responsibility.
- **Providing professional development** and support for teachers by helping them integrate AI into their pedagogy, curriculum, and assessment, and by facilitating their collaboration and innovation with AI.
- **Applying human-centered design principles** to the development and implementation of AI solutions, such as involving stakeholders in the design process, testing and iterating the solutions, and evaluating the impact and outcomes.
- **Aligning AI solutions with the best practices and principles of learning,** such as supporting student agency, fostering collaboration, enhancing feedback, and promoting critical thinking.
- **Avoiding sole reliance on the use of "AI detection" tools** in checking for student plagiarism. These tools often use data that is biased against students who are multilingual/English language learners.

Implementing AI in Student Learning

When integrating AI into student learning, it's important to empower students in how and to what degree AI is utilized in their learning journey. By doing so, students can actively participate in shaping their educational experience with AI.

- Co-create and share an AI decision-making rubric with students (example rubric).
- Support students in taking a human-centered approach to using AI.
- Empower students in leveraging AI in scaffolding understanding, feedback, and reflection.
- Support students in critically thinking about the role of AI within their learning journey and within their preparation for college, career, and life.
- Empower students receiving special education services to use AI to personalize and increase their access to learning.
- Integrate ethics and critical thinking activities that align with grade-level and subject-level instruction.
- Use AI for differentiation and assessment, including intelligent tutoring systems that allow text to speech, translation, personalized learning, and inquiry-based learning.
- Prepare students for jobs of the future, including those in career and technical education (CTE) programs, by partnering with industry to update and integrate learning standards.

Human Inquiry
HAH
Human Empowerment

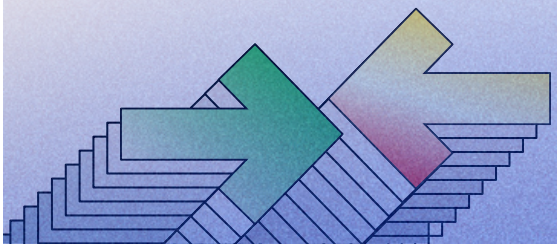
Policy

AI policies must promote equitable and inclusive access to AI. Education policymakers must focus on ensuring the use of AI increases the public good, with emphasis on empowering students who are members of communities that have been historically underserved. It is important that policies, by design, enhance a human-centered approach to pedagogy and learning, and respects ethical norms and standards. AI policy and use should be geared to improving learning for every student, empowering teachers, and strengthening learning management systems.

Building Human-Centered AI Policies

Incorporate the need for human intelligence and responsibility into AI usage policies. All AI use should start, and end, with human insight.

- **Responsible Use Policy:** Known as an acceptable use policy (AUP) or technology use policy, this describes what any person authorized to utilize the district's technology system may do and not do. It describes the terms and conditions for educational institutions and should be updated to include the safe and appropriate use of AI tools. (A separate AI AUP is not needed.)
- **AI Inquiry and Input Review:** Require human input of data with clear mandates that staff and students should never input personal, sensitive, or confidential data, including any data related to student education records, into any AI system without first ensuring that the system meets FERPA, COPPA, and CIPA requirements. Emphasize the need for review prior to finalizing any information into a system that learns from data entered.
- **Embrace the Use of Data and the Evaluation of AI Output:** Invest in systems that create streamlined opportunities for staff and students to enter information efficiently and safely, allow them to improve instruction, and draw connections to better understand student thinking and learning.
- **AI Output Review:** AI users should review and critically assess outputs from AI tools before sharing or publicizing results, including in the classroom. Staff and students should not rely exclusively on AI-generated content without fact-checking and evaluating results. Ultimately, it is up to human users to determine how AI information is shared and used.
    - **Bias and Misinformation:** AI-generated content is based on datasets or data models that may contain biases, false information, or other inaccuracies. AI systems do not have the ability to think or verify accuracy. Therefore, verifying AI results to ensure the source is credible must occur before considering an AI output in academic work.
    - **Safety and Respect:** Users must never use AI tools to create misleading or inappropriate content, take someone's likeness without permission, or harm humans or the community at large. (Note: This may also be added to a student code of conduct or bullying/cyberbullying/harassment policy.)



23.5: National Science Foundation on Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem

Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem:

An Implementation Plan for a National Artificial Intelligence Research Resource

National Artificial Intelligence Research Resource Task Force

January 2023

Dear Mr. President and Members of Congress,

Artificial Intelligence (AI) is changing our country and our world. From how citizens navigate their daily lives to how researchers drive discoveries in the lab to how manufacturers build products, AI is giving rise to new capabilities. New AI and AI-driven discoveries and capabilities hold the potential to drive practical solutions to address critical global challenges such as food production, climate change, poverty, and cancer. We have only started to scratch the surface of what is possible, and cannot afford to miss out on seizing the opportunity for leveraging AI to serve the public good.

However, the opportunities to pursue cutting-edge AI research and apply AI to new domains and challenges are currently not accessible by all of America's incredible talent nor harnessed by the public sector. Much of today's AI research relies on access to large volumes of data and advanced computational power, which are often unavailable to researchers beyond those at well-resourced technology companies and universities. This access divide limits the ability to leverage AI to tackle the big challenges in our society. It also constrains the diversity of researchers in the field and the breadth of ideas incorporated into AI innovations, contributing to embedded biases and other systemic inequalities found in AI systems today.

Recognizing this challenge, in the National AI Initiative Act of 2020, Congress directed the National Science Foundation (NSF), in consultation with the White House Office of Science and Technology Policy (OSTP), to establish a task force to create a roadmap for a National AI Research Resource (NAIRR)—a shared research infrastructure that would provide AI researchers and students with significantly expanded access to computational resources, high-quality data, educational tools, and user support.

This final report of the NAIRR Task Force presents a roadmap and implementation plan for a national cyberinfrastructure aimed at overcoming the access divide, reaping the benefits of greater brainpower and more diverse perspectives and experiences applied to developing the future of AI technology and its role in our society. Such a national cyberinfrastructure also presents a unique and critical opportunity to "design in" the standards for responsible AI research practices and governance processes that uphold our priority to develop and harness these groundbreaking technologies in a manner that reinforces our Nation's democratic values and Americans' personal freedoms.

OSTP and NSF formally launched the NAIRR Task Force in June 2021, appointing 12 leading experts equally representing academia, government, and private organizations. Over the course of its work, the Task Force held 11 public meetings, engaged with 65 experts on a wide range of aspects related to the design of the NAIRR, and considered responses from the public to two requests for information. We extend our gratitude to the members of the Task Force who have donated an extraordinary number of hours of their time to this effort, as well as to the many members of the public who have contributed their expertise and provided inputs to the Task Force. The result of the last one and one-half years of effort is this final report.

We see the NAIRR as a foundational investment that would amplify efforts across the Federal Government to cultivate AI innovation and advance trustworthy AI. Research, experimentation, and innovation are integral to our progress as a Nation, and it is imperative that we engage people from every zip code and every background to live up to America's unique promise of possibility and ensure our leadership on the world stage.

The work of the NAIRR Task Force and this report will be an invaluable resource as we work collaboratively across government and across sectors to drive this important work forward.

Sincerely,

Sethuraman Panchanathan
Director
National Science Foundation

Arati Prabhakar
Assistant to the President for
Science and Technology
Director, Office of Science and
Technology Policy

Executive Summary

Artificial Intelligence (AI) is an engine of innovation that is driving scientific discovery and economic growth. It is increasingly becoming an integral component of solutions that stand to impact everything from routine daily tasks to societal-level challenges, ultimately serving the public good. At the same time, there are also concerns that AI could have negative social and environmental consequences. To realize the positive and transformative potential of AI, it is imperative to harness all of America's ingenuity to advance the field in a manner that addresses societal challenges, works for all Americans, and upholds our democratic values.

Yet progress at the current frontiers of AI is often tied to access to large amounts of computational power and data. Such access today is too often limited to those in well-resourced organizations. This large and growing resource divide has the potential to limit and adversely skew our AI research ecosystem. The imbalance threatens our Nation's ability to cultivate an AI research community and workforce that reflect America's rich diversity and the ability to harness AI to advance the public good.

A widely accessible AI research cyberinfrastructure that brings together computational resources, data, testbeds, algorithms, software, services, networks, and expertise, as described in this report, would help to democratize the AI research and development (R&D) landscape in the United States for the benefit of all. It would help create pathways to broaden the range of researchers involved in AI, and to grow and diversify approaches to, and applications of, AI. This cyberinfrastructure can also help to open up new opportunities for progress across all scientific fields and disciplines, including in critical areas such as AI auditing, testing and evaluation, trustworthy AI, bias mitigation, and AI safety. Increased access and a diversity of perspectives can, in turn, lead to new ideas that would not otherwise materialize and set the conditions for developing AI systems that are inclusive by design.

As part of the National Artificial Intelligence Initiative Act of 2020, Congress established the National Artificial Intelligence Research Resource (NAIRR) Task Force to "investigate the feasibility and advisability of developing" the NAIRR as a national AI research cyberinfrastructure, and "to propose a roadmap detailing [how the NAIRR] should be established and sustained." The recent CHIPS and Science Act of 2022 reinforces the importance of democratizing access to a national AI research cyberinfrastructure, via investments that will accelerate development of advanced computing—from next-generation graphics processing units to high-density memory chips—as well as steps to actively engage broad and diverse U.S. talent in frontier science and engineering, including AI.

This final report is the culmination of the Task Force's 18-month effort to develop a vision and implementation plan for establishing the NAIRR. It builds on the findings and recommendations outlined in the Task Force's interim report released in May 2022, providing an implementation plan to achieve the objective of the NAIRR: to strengthen and democratize the U.S. AI innovation ecosystem in a way that protects privacy, civil rights, and civil liberties.

The NAIRR should be established with four measurable goals in mind, namely to (1) spur innovation, (2) increase diversity of talent, (3) improve capacity, and (4) advance trustworthy AI. The NAIRR should meet these goals by supporting the needs of researchers and students from diverse backgrounds who are pursuing foundational, use-inspired, and translational AI research. These users should be U.S.-based or affiliated with U.S. organizations, to include academic institutions, non-profit organizations, and startups or small businesses.

The NAIRR should comprise a federated set of computational, data, testbed, and software resources from a variety of providers, along with technical support and training, to meet the needs of this target user base. The specific design, implementation, and evaluation of the NAIRR should be centered around the four key goals, and should support the collection of data for assessment of key indicators of system performance and success in progress toward these goals.

The NAIRR administration and governance should follow a cooperative stewardship model, whereby a single Federal agency serves as the administrative home for NAIRR operations and a Steering Committee comprising principals from Federal agencies with equities in AI research drives the strategic direction of the NAIRR. A Program Management Office within the administrative home agency should provide funding and oversight for an independent Operating Entity that manages the day-to-day operations of the NAIRR. The Steering Committee, co-chaired by the National AI Initiative Office (NAIIO), would incorporate interests and perspectives from across Federal agencies in the governance of the NAIRR. These agencies should also directly support resource providers whose resources, in federation, would constitute the NAIRR. Diverse perspectives and expertise should be tapped to inform the NAIRR's operations through a User Committee, a Science Advisory Board, a Technology Advisory Board, and an Ethics Advisory Board that provide advice to the Operating Entity.

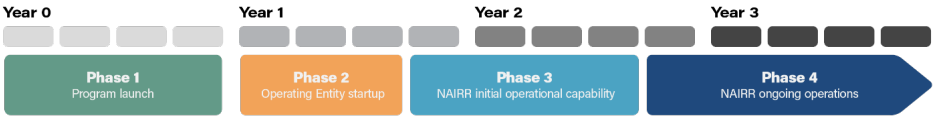
The NAIRR should provide access to a federated mix of computational and data resources, testbeds, software and testing tools, and user support services via an integrated portal. Computational resources should include conventional servers, computing clusters, high-performance computing, and cloud computing, and should support access to edge computing resources and testbeds for AI R&D. Open and protected data should be made available under tiered-access protocols and co-located with computational resources. The Operating Entity should not itself operate the totality of the computer hardware that composes the NAIRR; instead, computing, along with data, testing, and training resources, should be delivered as services by partner resource providers selected through Federal agency or multi-agency funding opportunities. When fully implemented, the NAIRR should address both the capacity (ability to support a large number of users) and capability (ability to train resource-intensive AI models) needs of the AI research community.

The NAIRR must be broadly accessible to a range of users and provide a platform that can be used for educational and community-building activities in order to lower the barriers to participation in the AI research ecosystem and increase the diversity of AI researchers. The NAIRR access portal and public website should provide catalogs and search and discovery tools to facilitate access to data, testbeds, and educational and training resources serving a range of experience levels.

The NAIRR should set the standard for responsible AI research through the design and implementation of its governance processes. The NAIRR must be proactive in addressing privacy, civil rights, and civil liberties issues by integrating appropriate technical controls, policies, and governance mechanisms from its outset. The Operating Entity should work with its Ethics Advisory Board to develop criteria and mechanisms for evaluating proposed research and resources for inclusion in the NAIRR from a privacy, civil rights, and civil liberties perspective. Regular training should be required to build NAIRR users' awareness about rights, responsibilities, and best practices related to privacy, civil rights, and civil liberties in AI research, in accordance with the Blueprint for an AI Bill of Rights published by the White House Office of Science and Technology Policy in October 2022.

The NAIRR should implement system safeguards in accordance with established guidelines. These guidelines include those developed by the National Institute of Standards and Technology (NIST) and the Five Safes framework: safe projects, safe people, safe settings, safe data, and safe outputs. The Operating Entity should design the NAIRR cyberinfrastructure to consist of multiple tiers, starting with two primary zones: an open science zone "NAIRR-Open" and a secure zone "NAIRR-Secure." Each zone should federate computational, network, and data resources operating in accordance with security and access-control policies that are uniform within the zone, but different between zones, reflecting the different priorities and needs of the users and resource operators. NAIRR-Open should adopt the best practices developed over two decades in the open science community; be consistent with Federal open data, open government, and research security policies; and manage access using single sign-on authentication and a resource allocation mechanism managed by the Operating Entity. NAIRR-Secure should consist of one or more secure enclaves adhering to a common set of security controls, and have the ability to support security requirements arising from legally protected data.

NAIRR implementation should occur over four phases, beginning immediately after the publication of this report. In phase one, Congress should authorize and appropriate funds to establish the NAIRR. The administrative home agency and the NAIIO should coordinate the formation of the Steering Committee and stand up a Program Management Office, which will then prepare the solicitation for the Operating Entity and manage the selection process.



Phased NAIRR Implementation Timeline

In phase two, the Operating Entity should establish its activities and oversee creation of the NAIRR portal and user interface, building in appropriate technical and policy controls. The architecture should support collection of key performance indicators for evaluation of NAIRR progress. Resource providers should be selected via coordinated, multi-agency funding opportunities ideally released within six months of the initial Operating Entity award.

In phase three, the NAIRR should achieve initial operational capability and the Operating Entity should also formalize the policies, processes, and initial technical resources to be made available to AI researchers. Initial capabilities include (1) a portal and user support resources, (2) a mix of computational resource providers, (3) an allocation and identity system, and (4) a data publication system. In phase four, activities should transition from building out the NAIRR to establishing steady-state operations, as well as the planned evolution of NAIRR resources in response to user uptake and demand.

Finally, the Task Force also presents a pilot option for the implementation of the NAIRR that would be initiated in parallel with the above phases to expedite the availability of NAIRR resources to the AI R&D community.

As envisioned, the impact of the NAIRR will be significant and far-reaching, enabling researchers to tackle problems that range from routine tasks to global challenges. In order to achieve its vision and goals, the Task Force estimates the budget for the NAIRR as $2.6 billion over an initial six-year period. The bulk of this investment ($2.25 billion) is to fund the resources to be made accessible via the NAIRR, through appropriations to multiple Federal agencies. The Task Force estimated this budget based on recent costs of advanced computing resources as well as data, training, and software resources; estimates of usage levels to meet the current needs of the AI R&D community; and expected growth of the AI R&D community. Resource providers should be brought online every two years with a six-year lifetime, so that a new $750 million investment is made every two years to ensure that the NAIRR resources remain state-of-the-art. The Operating Entity will require between $55 million and $65 million per year to support the coordination and management of NAIRR activities. An additional $5 million per year is budgeted for external evaluation of the Operating Entity and NAIRR performance.

The vision for the NAIRR laid out in this report is designed to meet the national need for increased access to the state-of-the-art resources that fuel AI innovation. The roadmap for achieving this vision builds on existing Federal investments; designs in protections for privacy, civil rights, and civil liberties; and promotes diversity and equitable access. **If successful, the National AI Research Resource would transform the U.S. national AI research ecosystem and facilitate the ability to address societal-level problems by strengthening and democratizing participation in foundational, use-inspired, and translational AI R&D in the United States.**

23.6: Department of Justice on Provide General Guidance on the Use of Generative Artificial Intelligence (AI) in EOIR Proceedings

OOD
PM 25-40
Effective: August 8, 2025

To: All of EOIR
From: Sirce E. Owen, Acting Director
Date: August 8, 2025

USE OF GENERATIVE ARTIFICIAL INTELLIGENCE IN EOIR PROCEEDINGS

PURPOSE:	Provide general guidance on the use of generative artificial intelligence (AI) in EOIR proceedings
OWNER:	Office of the Director
AUTHORITY:	8 C.F.R. § 1003.0(b)
CANCELLATION:	None

The use of generative AI has been one of the most significant—and potentially transformative—developments in the legal profession in recent years. As with many technological advancements, however, the increased use of generative AI offers both significant benefits and risks. As Chief Justice John Roberts noted at the end of 2023, "AI obviously has great potential to dramatically increase access to key information for lawyers and non-lawyers alike. . . . But any use of AI requires caution and humility." 2023 Year-End Report on the Federal Judiciary, U.S. Supreme Court, https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf. EOIR, too, should approach the use of generative AI with an open mind due to its potential benefits, but also with caution due to its potential risks.

Perhaps the greatest potential risk, as demonstrated by multiple media reports in recent years, is the use of hallucinated legal citations or arguments generated by AI and subsequently filed with a court, which causes significant damage to the legal system and may lead to possible sanctions in addition to reputational harm.[1] As the largest administrative court system by case volume in the

[1] The use of hallucinated citations is not necessarily limited to pleadings filed by attorneys. Recently, two separate federal district court judges withdrew opinions after attorneys raised questions regarding the accuracy of record citations and information in those opinions. Reuters, Mike Scarcella, *Two US judges withdraw rulings after attorneys question accuracy* (Jul. 29, 2025), https://www.reuters.com/legal/government/two-us-judges-withdraw-rulings-after-attorneys-question-accuracy-2025-07-29/. Although neither judge confirmed whether generative AI was responsible for the errors, the decisions bore indicia consistent with research or drafting done by generative AI. Although EOIR prohibits "the unauthorized use of AI services. . .on [Department of Justice] Government Furnished Equipment," EOIR Office of Information Technology, Notification: Updated - Unauthorized use of AI services on GFEs (May 22, 2025), the broader policy of the federal government is "to accelerate the Federal use of AI by focusing on three key priorities: innovation, governance, and public trust." Office of Management and Budget Memorandum M-25-21, *Accelerating Federal Use of AI through Innovation, Governance, and Public Trust*, at 2 (Apr. 3, 2025),

federal government, EOIR is particularly susceptible to the improper or problematic use of generative AI. Although several professional state bars have issued ethical guidance regarding the use of generative AI and many courts and judges at both the federal and state levels have established standing orders regarding the disclosure of the use of generative AI in pleadings, EOIR has largely lagged behind with guidance for its adjudicatory components. In January 2025 EOIR's Office of the Chief Administrative Hearing Officer (OCAHO) established general guidelines through adjudication regarding parties' use of generative AI in its proceedings. *See United States v. Wallcon, LLC*, 21 OCAHO no. 1630, 9-14 (2025). However, the Office of the Chief Immigration Judge and the Board of Immigration Appeals (BIA) have not—as of yet—established any uniform positions on the subject. EOIR may pursue rulemaking at a future date to provide further guidance on the use of generative AI, and the BIA may also establish guidelines through a published precedential decision. Moreover, the Department of Justice is expected to issue departmentwide guidance on the use of generative AI by the end of 2025. *See* OMB Memorandum M-25-21, at 12 (requiring each federal agency to "develop a policy that sets the terms for acceptable use of generative AI for their missions and establishes adequate safeguards and oversight mechanisms that allow generative AI to be used in the agency without posing undue risk" within 270 days, or by approximately December 29, 2025). Until further policy directives are issued, however, this Policy Memorandum (PM) provides general guidance for EOIR adjudicators to consider regarding parties' use of generative AI in immigration proceedings, particularly in cases before the Immigration Courts or the BIA.[2]

EOIR has neither a blanket prohibition on the use of generative AI in its proceedings nor a mandatory disclosure requirement regarding its use. Nevertheless, nothing in EOIR's rules prohibit individual adjudicators or courts from adopting standing orders, *see* PM 20-09, The Immigration Court Practice Manual and Orders (Feb. 13, 2020), or local operating procedures, *see* 8 C.F.R. § 1003.40, regulating the use and disclosure of generative AI contained in pleadings. Any such orders or procedures that are adopted remain subject to management approval requirements and will be posted in the appropriate location on EOIR's website. *See* PM 20-09 at 3. The BIA, too, may prescribe rules, with the approval of the Director, related to the use of generative AI in pleadings filed with it. *See* 8 C.F.R. § 1003.1(d)(4).

For attorneys who choose to use generative AI tools in the preparation of legal filings in any proceeding before EOIR, that use has the potential to implicate applicable rules of professional conduct, as well as associated ethics rules. *See*, *e.g.*, American Bar Association, Formal Opinion 512, Generative Artificial Intelligence Tools (Jul. 29, 2024) (ABA Formal Opinion 512), https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf (identifying multiple model rules of professional responsibility implicated by the use of generative AI). Moreover, practitioners appearing before EOIR's adjudicatory components are expected to act in a professional, ethical manner and in conformance with the applicable rules and standards of professional conduct, including the rules

https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf. Thus, to the extent that authorized generative AI use is approved in the near future, EOIR adjudicators are strongly cautioned to ensure that any authorized use of generative AI complies with all applicable ethical and professional responsibility obligations. Any improper use of generative AI by an EOIR adjudicator may result in corrective or disciplinary action.

[2] This PM is largely consistent with OCAHO's established guidance. To the extent this PM may conflict with OCAHO's policies established through adjudication, the OCAHO adjudicatory decisions would control.

of professional conduct of any relevant state bar. *See* 8 C.F.R. § 1003.101(a) (providing that a practitioner authorized to practice before the Board and the Immigration Courts may be subject to disciplinary sanctions "when such person has engaged in criminal, unethical, or unprofessional conduct, or in frivolous behavior"); 28 C.F.R. § 68.35 (providing that those appearing in proceedings before OCAHO "are expected to act with integrity, and in an ethical manner").

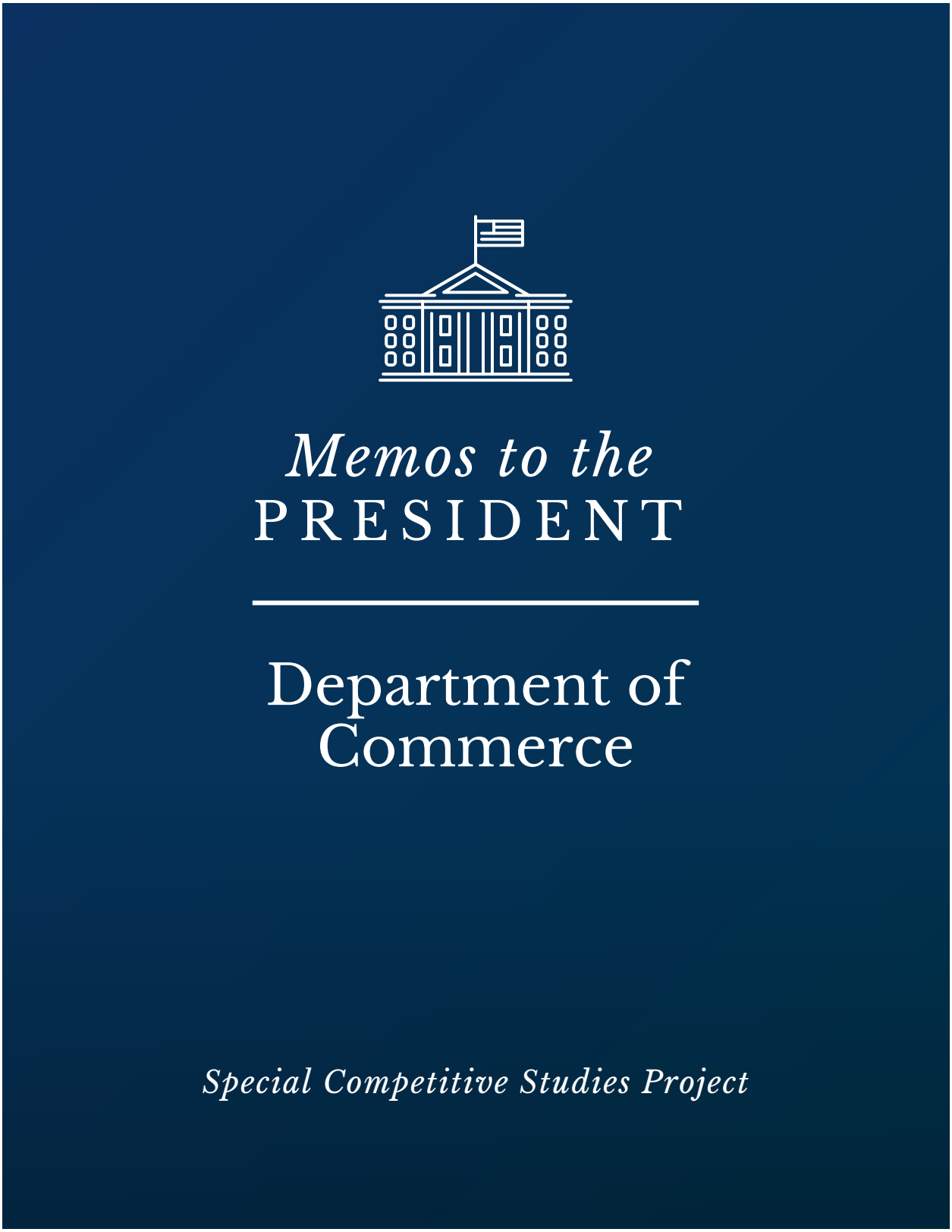
Thus, practitioners submitting hallucinated or erroneous AI-generated content in filings before EOIR likely violate professional conduct obligations, and attorneys who submit such content to an Immigration Court or the BIA may be subject to discipline for "knowingly or with reckless disregard offering false evidence," 8 C.F.R. § 1003.102(c). Depending on the particular posture and facts of a case and the impact of the filing, the use of such content may also implicate disciplinary rules regarding "[e]ngag[ing] in conduct that constitutes ineffective assistance of counsel," "[e]ngag[ing] in conduct that is prejudicial to the administration of justice or undermines the integrity of the adjudicative process," "[f]ail[ing] to provide competent representation to a client," failing to disclose adverse legal authority, and "[r]epeatedly draft[ing] notices, motions, briefs, or claims that are filed with. . .EOIR that reflect little or no attention to the specific factual or legal issues applicable to a client's case, but rather rely on boilerplate language indicative of a substantial failure to competently and diligently represent the client." 8 C.F.R. §§ 1003.102(k)-(o), (s), (u). Consequently, parties and attorneys before each of EOIR's adjudicatory components should take care to confirm the accuracy of any citations or other research or drafting conducted using generative AI tools. Moreover, in certain circumstances, attorneys may need to consult with their clients regarding the use of generative AI. *See* ABA Formal Opinion 512 at 8 (discussing situations in which the Model Rules of Professional Responsibility may require the disclosure of the use of generative AI to a client).

EOIR adjudicators should also be vigilant to ensure that decisions are not based on hallucinated or inaccurate case citations and information and that attorneys are not submitting pleadings with false information or non-existent legal citations. Accordingly, consistent with EOIR's core policy values to maintain the integrity of its immigration proceedings, EOIR adjudicators who suspect or discover the inappropriate use of generative AI should report those instances to EOIR's Attorney Discipline Program and, as appropriate, its Anti-Fraud Program.

This PM is not intended to, does not, and may not be relied upon to create, any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person. Nothing herein should be construed as mandating a particular outcome in any specific case. Nothing in this PM limits an adjudicator's independent judgment and discretion in adjudicating cases or an adjudicator's authority under applicable law.

Please contact your supervisor if you have any questions.

23.7: Memos to the President - Department of Commerce



SPECIAL COMPETITIVE STUDIES PROJECT

Memos to the
PRESIDENT

Subject: Modernizing the Department of Commerce for Strategic Competition

Purpose: Outline key recommendations for restructuring the Department of Commerce to advance U.S. economic and national security interests.

Objectives:
1. *Supercharge* the Department of Commerce to Lead America's AI Future
2. *Bolster* Techno-Economic Intelligence Capabilities
3. *Position* the U.S. Industrial Base for Techno-Industrial Competition
4. *Organize* the Bureau of Industry and Security for Tech Rivalry
5. *Retool* the International Trade Administration

Background

As the tip of the spear for advancing U.S. techno-economic competitiveness and security, the Department of Commerce (Commerce) must evolve to address the challenges and seize the transformative opportunities of AI and strategic competition. AI is poised to deliver massive economic benefits, driving productivity gains across industries and unleashing new sources of prosperity.[1] However, the geopolitical landscape has shifted, requiring Commerce to adapt to an era of strategic rivalry rather than unfettered globalization. China's state-driven economic model gives Beijing asymmetric advantages that the U.S. Government is not currently organized to counter effectively. To rise to this challenge while unlocking AI's potential, Commerce must be reorganized and strengthened. As it recalibrates for this new reality, Commerce must also sustain its core mission of fostering economic dynamism. The following reforms are essential to prepare the Department for these demands.

Recommendations

Objective 1: Supercharge Commerce to Lead America's AI Future

Traditionally focused on promoting business and trade, Commerce must adapt to capitalize on advances in AI while countering China's efforts to dominate critical technology sectors. By taking a more proactive role in accelerating technology innovation and adoption, strengthening partnerships with industry, and shaping international technology standards, Commerce can drive U.S. leadership in AI and other strategic industries.

[1] Henry Kissinger, et al., Genesis: Artificial Intelligence, Hope, and the Human Spirit, Little, Brown and Company at 137 (2024).

- **Launch an Agentic AI Acceleration Initiative (A3I).** Agentic AI systems are poised to transform entire industries, enabling the automated execution of complex tasks across sectors and varied use-cases.[2] The federal government, led by Commerce's National Institute of Standards and Technology (NIST), should join forces with leading AI companies to create dynamic testing environments that accelerate breakthroughs in agentic AI. These testbeds should combine government resources with private sector ingenuity, enabling rapid prototyping and scaling of agentic AI systems with applications in areas like autonomous infrastructure, advanced robotics, computational biology, and cyberdefense.[3] By bringing together the best minds from government and industry in state-of-the-art facilities, these partnerships could dramatically speed up the development cycle for agentic systems—and drive their adoption in government—turning visionary ideas into deployed solutions that drive progress across critical sectors of the economy.

- **Dominate International Standards Setting for Advanced Networks.** As the backbone of the digital economy, advanced networks are a critical enabler of AI infrastructure. The United States and its partners must counter China's coordinated push to dominate international technology standards, particularly in 6G, with an equally robust response. While China floods standards bodies with representatives and promotes frameworks that could entrench its technological dominance,[4] the United States—enabled by additional NIST staff—must mobilize a coalition of government agencies, such as the Department of State and Commerce, alongside tech companies, research institutions, and allies and partners. This coalition should deploy technical experts to standards organizations, advocating for standards that promote innovation, secure America's competitive edge, and prevent China from dictating the technical foundations of transformative technologies.

Objective 2: Bolster Techno-Economic Intelligence Capabilities

With advances in AI transforming data analysis and decision-making, and the emergence of increasingly capable AI agents that can work under human supervision, the time is right to modernize Commerce's approach to assessing and countering technological and economic threats. Currently, fragmented capabilities across bureaus lead to delayed responses and missed opportunities. Establishing a Techno-Economic Intelligence Unit (TEIU) would enable Commerce to anticipate and mitigate threats proactively, leveraging AI-enabled data processing to strengthen U.S. competitiveness in an increasingly dynamic global environment.

- **Bridge the Intelligence Divide.** The TEIU would function as a primary bridge facilitating data sharing and strategy alignment across Commerce, the Department of Defense, and the Intelligence Community. This would include establishing joint analytical teams, developing common assessment frameworks for emerging threats, and sharing best practices for integrating

[2] An agentic AI system is one that can interact with and utilize multiple external systems and services to accomplish complex tasks, like coordinating between airline booking platforms, calendars, and weather services to plan a trip. Agentic AI operates with a degree of autonomy—meaning it can perceive its environment, make decisions, and take actions to achieve specified goals without needing constant human direction. Some AI leaders are predicting that 2025 will be the year that AI agents "join the workforce" and begin changing companies' output significantly. See, e.g. Sam Altman, Reflections (2025). These systems often incorporate elements like goal-oriented planning, self-guided learning, and the ability to adapt to changing conditions.
[3] NIST has begun to develop various AI testbeds. See, for example, Assessing Risks and Impacts of AI, National Institute for Standards and Technology (last accessed 2025). Similar efforts, such as testbeds supported by the Department of Energy and the National Science Foundation, integrate government resources like high-performance computing and regulatory flexibility with private sector expertise to enable rapid prototyping and scaling of advanced AI systems. See, for example, Artificial Intelligence Testbeds at DOE, U.S. Department of Energy (last accessed 2025); NSF Announces New AI Testbed Initiative to Advance Safety and Security of AI Technology, National Science Foundation (2024).
[4] China is Writing the World's Technology Rules, The Economist (2024).

AI tools into the analytic process. The unit would ensure that timely analysis informs decision-making across export controls, industrial strategy, and supply chain efforts while reducing duplicative efforts.[5]

- **Adopt Advanced Commercial AI Systems.** Equip the TEIU with advanced commercial AI systems tailored to support key functions across Commerce's bureaus. These tools would enable detailed analysis for export control enforcement by identifying potential violations and high-risk transfers in real time. In supply chain monitoring, AI agents could track trade flows, flag idiosyncrasies, and support humans to develop mitigation actions. For strategic planning, AI-driven analytics would provide actionable insights into global market trends and technological developments.

Objective 3: Organize for Advanced Manufacturing Leadership

To bolster U.S. industrial competitiveness and capitalize on the transformative potential of AI, the Department of Commerce must prioritize strengthening industrial capabilities, enhancing resilience, and accelerating the adoption of advanced technologies across the manufacturing sector. This effort should focus on integrating AI into industrial processes, enabling U.S. manufacturers—particularly small and medium-sized enterprises—to modernize, innovate, and remain competitive in a rapidly evolving global economy. By aligning its policies and programs to support these objectives, Commerce can ensure the U.S. industrial base remains a cornerstone of economic growth and technological leadership.

- **Reorganize Manufacturing-Related Offices Under an Assistant Secretary for Manufacturing and Industrial Strategy.** Manufacturing-related offices are fractured throughout the Department, leading to a lack of policy coordination between programs. The Office of Advanced Manufacturing and the Manufacturing Extension Partnership Program should be moved from NIST and placed under an Assistant Secretary for Manufacturing and Industrial Strategy, along with the Department of Commerce's Supply Chain Office and CHIPS Program Office. These revitalized offices would help U.S. manufacturers adopt AI, robotics, and other advanced technologies, track the construction of factories of the future, and bolster supply chain resilience.

- **Modernize and Scale the Manufacturing USA and Manufacturing Extension Partnership (MEP) Programs.** Manufacturing USA and MEP are the nation's flagship programs supporting U.S. manufacturers. MEP is a national network that supports manufacturers in all 50 states, but the program is not organized to help small- and medium-sized firms adopt advanced technologies. These programs receive one tenth of the funding than programs in other industrialized nations.[6] Funding should be significantly increased to reflect the strategic importance of these programs.

Objective 4: Position the Bureau of Industry and Security for Tech Rivalry

The Bureau of Industry and Security (BIS) operates with outdated legacy systems and insufficient resources, limiting its ability to manage increasingly complex export controls on emerging technologies. Its current structure and mission reflect a post-Cold War focus on nonproliferation and law enforcement

[5] At present, the Supply Chain Center, housed in the Industry and Analysis unit within the International Trade Administration (ITA), aims to be "the analytic engine for supply chain resiliency policy for the U.S. Government." The TEIU could consolidate analytic functions across the Department. See The Decisive Decade: Advancing National Security at the Department of Commerce, U.S. Department of Commerce at 30 (2024).
[6] Sridhar Kota & Tom Mahoney, Reclaiming America's Leadership in Advanced Manufacturing, MForesight: Alliance for Manufacturing Foresight at 19 (2019).

rather than comprehensive technology security. Moreover, concerns have been raised that export control decisions are at times subject to undue influence from industry.[7] Modernizing BIS is essential to safeguarding U.S. technological advantages and preventing adversaries from accessing critical capabilities.

- **Implement AI-Enabled Export Control Monitoring.** As an example of how commercial AI platforms can enhance the work of the Department, BIS should receive funding for AI-enabled agentic systems to identify technology transfer risks and violations in real-time by analyzing export data, shipping records, and threat intelligence. These systems, supervised by human experts, would integrate predictive analytics and specialized tools to track both physical and intangible technology transfers.

- **Create a Whistleblower Program for Export Controls.** A whistleblower program modeled after the framework of the program at the Securities and Exchange Commission (SEC)[8] would provide a confidential mechanism for reporting potential export control violations and regulatory gaps. Employees, contractors, and members of the public could confidentially report potentially illicit behavior, with protections against retaliation and potential monetary awards for significant disclosures. This approach would deter wrongdoing, increase transparency, and help BIS more effectively protect national security interests. Rather than adding a new government office, the whistleblower program could be run by a nonprofit or public-private partnership.

Objective 5: Retool the International Trade Administration

The International Trade Administration (ITA) should shift its focus from global trade promotion that made sense in the era of hyperglobalization, to fostering strategic bilateral and plurilateral trade relationships with trustworthy allies and partners. The ITA should prioritize market economies over their autocratic, state-driven peers. The ITA's deep expertise and international network must be recalibrated to counter competitors like China, which has distorted the global trading system. This transformation would involve leveraging the ITA's industry relationships to align trade efforts with national security and competitiveness goals, including establishing U.S. leadership positions in AI and other areas like quantum computing and biotechnology, prioritizing collaboration with partners in critical sectors, and integrating with broader economic security initiatives. By doing so, the ITA can strengthen U.S. trade leadership while continuing to support exports and investment.

- **Launch a Tech Export Accelerator.** In collaboration with the Department of State, the ITA should launch a Tech Export Accelerator to support the development, structuring, and completion of strategic technology-related commercial transactions abroad by galvanizing U.S. financing, commercial promotion, and advocacy tools, in partnership with U.S. tech firms. The Accelerator would serve as a "one-stop shop" to liaise with U.S. embassies, tech firms, and economic agencies, including the Export-Import Bank (EXIM), the Development Finance Corporation (DFC), and the United States Trade and Development Agency (USTDA) to maintain a pipeline of opportunities for U.S. technology tenders and investment opportunities abroad, leveraging AI tools to maintain an up-to-date database of potential transactions.[9]

- **Forward Deploy Technology Security Units.** Commerce should create specialized units combining commercial, technology, and national security expertise, deployed in domestic tech hubs and at key diplomatic missions. These units would use AI-enabled tools to tackle economic security threats like IP theft, technology transfer risks, and digital trade barriers, liaising with and

[7] Ana Swanson, How U.S. Firms Battled a Government Crackdown to Keep Tech Sales to China, New York Times (2024).
[8] Whistleblower Program, U.S. Securities and Exchange Commission (last accessed 2025).
[9] Restoring the Sources of Techno-Economic Advantage, Special Competitive Studies Project (2022).

offering training to industry and supporting enforcement. By integrating TEIU analytics, collaborating with the Department of the Treasury, the United States Trade Representative (USTR), and the Department of State, and leveraging industry insights, the units would provide early warnings, support outreach to industry on tech threats, and develop strategies to address technology security challenges. The units would also reinforce the presence of Foreign Commercial Service officers and USPTO IP attachés in strategic markets while strengthening coordination between embassy personnel and domestic agencies.

Conclusion

The Department of Commerce must undergo significant modernization to meet the dual challenges of advancing U.S. leadership in AI while countering strategic competitors, particularly China. The proposed reforms—including launching an Agentic AI Acceleration Initiative, establishing a Techno-Economic Intelligence Unit, strengthening focus on advanced manufacturing, modernizing export controls, and retooling the international trade administration—would transform Commerce into an agency better equipped for technological rivalry. By implementing these changes while maintaining its core economic mission, Commerce can effectively promote U.S. technological leadership, protect critical capabilities, and foster strategic economic partnerships in an era of intense global competition.

Exhibit 24: **Cited Industrial and Professional Reports and Documents**

24.1: Federation of State Medical Boards on Navigating the Responsible and Ethical Incorporation of Artificial Intelligence into Clinical Practice

Navigating the Responsible and Ethical Incorporation of Artificial Intelligence into Clinical Practice

Adopted by FSMB House of Delegates, April 2024

EXECUTIVE SUMMARY

Artificial Intelligence (AI) holds tremendous potential to aid healthcare providers in diagnosis, treatment selection, clinical documentation, and other tasks to improve quality, access, and efficiency. However, these technologies introduce risks if deployed without proper "guardrails" and understanding which may impact considerations in clinical practice as well as regulatory processes of state medical boards. By taking a proactive and standardized governance approach anchored in ethical principles, state medical boards can promote safe and effective integration of AI, in its various forms, while prioritizing patient wellbeing.

This report summarizes expert opinion and proceedings to develop guidance from the FSMB Ethics and Professionalism Committee to aid physicians and state medical boards in navigating the responsible and ethical incorporation of AI centered on (1) education, (2) emphasizing human accountability, (3) ensuring informed consent and data privacy, (4) proactively addressing responsibility and liability concerns, (5) collaborating with experts, and (6) anchoring AI governance in ethical principles.

Clinical systems and processes making use of AI must be continually monitored and refined. This should not occur in a vacuum but should be the focus of collaborative efforts among physicians, health systems, data scientists, and regulatory agencies, *including state medical boards*. By thoughtfully addressing the opportunities and challenges posed by AI in healthcare, state medical boards can promote the safe, effective, and ethical use of AI as a tool to enhance, but generally not replace, human judgment and accountability in medical practice. In fulfilling their missions to ensure that patients benefit from and are not harmed by applications of AI in their care, it is essential that state medical boards avoid over-regulation and regulatory overreach by attempting to regulate that which is not in their purview. With focused efforts on the current and future state of the use of AI by licensees, state medical boards may sustain regulatory efficiency, achieve consistency across jurisdictions in the regulation of AI in clinical practice, help secure the benefits of AI, and proactively safeguard patients while upholding professional standards.

24.2: National Center for State Courts on Guidance for Use of AI and Generative AI in Courts (selected pages)

Introduction and Background

The Artificial Intelligence Rapid Response Team (AI RRT) is a project of the Conference of Chief Justices (CCJ) and the Conference of State Court Administrators (COSCA), and supported by the National Center for State Courts (NCSC).

The AI RRT was established to help courts plan for the impact that Generative Artificial Intelligence (GenAI) may have on the courts. GenAI is rapidly evolving and has the potential to change the practice of law and how courts operate. As with many new technologies, it is imperative that the courts become informed consumers of GenAI. The AI RRT has spent the past eight months examining this issue. As part of its work, the AI RRT has published seven (7) interim guidance documents for the courts and created a **resource center** that includes a landscape of court orders, rules, guidance and other initiatives of the state court community and the federal courts regarding AI or GenAI. The AI RRT conducted a survey and follow-up survey of state activities and published the results on **NCSC's AI website** (ncsc.org/ai). The information provided in this document is intended to help get courts started on their GenAI journey. State Court leaders are encouraged, if they have not already done so, to establish an internal work group to examine the impact of AI and GenAI on their courts and establish a plan moving forward.

Artificial Intelligence (AI) is an umbrella term and GenAI is a type of AI technology that is one of the most recognized by the public today. The term AI is used to refer to something as simple as spell check, predictive typing or asking Siri or Alexa the temperature, or as complex as computer based legal research, projections, facial recognition, or generating documents, videos, or audio.

AI Rapid Response Team Members

Chief Judge Anna Blackburne-Rigsby
Chief Judge, D.C. Court of Appeals/CCJ Chair
AI RRT Co-Chair

Justin Forkner
Chief Administrative Officer
Indiana Supreme Court/COSCA
AI RRT Co-Chair

Justice Beth Walker
Justice, Supreme Court of West Virginia

Chief Justice Matthew Fader
Chief Justice, Supreme Court of Maryland

Chief Justice Michael P. Boggs
Chief Justice, Supreme Court of Georgia

Judge Joseph A. Zayas
Chief Administrative Judge,
New York State Unified Court System/COSCA

Stacey Marz
Administrative Director,
Alaska Court System/JTC/COSCA

Sara Omundson
Administrative Director of the Courts,
Idaho Supreme Court/COSCA

NCSC Staff

Shay Cleary Cathy Zacharias David Sachar Miguel Trujillo Andrea Miller

Guidance for Using AI in Courts

Public Trust and Confidence

Public trust and confidence in the courts is integral to the credibility of the judicial branch. Courts and judicial officers are responsible to ensure that the use of GenAI and other AI tools does not erode the public's trust and confidence in courts due to errors or biases.

Court Rules: Courts should review their rules to determine whether they are sufficient to address expectations of lawyers and litigants concerning the responsible use of GenAI in court filings and proceedings or whether changes may be appropriate to clarify those expectations.

Ethical Guidelines: Education on the applicability of current ethical guidelines is vital to ensure that GenAI is used ethically by lawyers, litigants, and the courts. Courts should review their rules and comments to the rules to determine if they should be updated to clarify their applicability to new technological tools.

Education: Courts must ensure that judicial officer and court staff are educated on the benefits and risks of AI. Courts will need to be aware of how GenAI is used to create content that looks real, sometimes referred to as deepfakes, which will increasingly impact discovery and evidentiary issues in legal proceedings.

Understanding GenAI — What Courts Should Know

AI has the potential to streamline tasks within the courts, increasing efficiency and allowing staff to work on higher level tasks. AI also has the potential to be used to help create resources for self-represented litigants, expanding access to justice. But like any technology, AI is not infallible or without risks.

Limitations

With the proliferation of new GenAI tools being developed for the courts, lawyers, and self-represented litigants and the ease of their use, courts need be aware of the capabilities and potential limitations of GenAI tools such as ChatGPT, Gemini, and CoPilot (popular GenAI tools at the publication of this document).

GenAI is not a traditional search engine and most GenAI platforms are not designed to provide legal authority. The purpose of GenAI is to create content. Lawyers and self-represented litigants are already using GenAI in drafting legal documents and performing legal research, and courts must understand the capabilities of the tools they are using. This includes the benefits of time saving legal research, drafting assistance, and organizing large volumes of information. There are also significant concerns about lack of accuracy, bias, GenAI-enhanced evidence, and deepfakes. As discussed below, judicial officers should be aware of certain indicators that a document filed with a court was generated with GenAI.

Accuracy

Early GenAI tools have been known to create hallucinations, which means generating inaccurate or fictitious content, such as case citations to cases that do not exist. Multiple courts have now issued sanctions for lawyers submitting filings with fictitious citations generated by GenAI tools.

Attorneys and self-represented litigants are using these tools to create legal documents. Westlaw and Lexis now provide the capability of using GenAI for legal research. However, a recent Stanford paper revealed inaccuracies in the output generated by these legal research tools, despite the fact that they use closed training systems.[3] Courts should be aware of these issues with accuracy when reviewing legal documents.

The following are indications that GenAI may have been used to create a document:

- References to cases that do not sound familiar, cannot be found through traditional legal research, or have unfamiliar citation formats.
- At first read, AI text may sound impressive and well written, but there are often structural issues. AI content tends to be overly formulaic and lacks natural transitions between topics. Once you strike out all the words that are meaningless filler, there may not be a lot of substance left. AI is also not mindful of grammar rules or basic punctuation although that is improving.
- AI is designed to recognize patterns and replicate them as accurately as possible so look for repetitive patterns in the writing. Perhaps the most obvious sign of AI-generated content is the use of repeated words, phrases, or the same sentence structure used regularly in different paragraphs within the same document.
- Often AI generated content is written in the general sense, glossing over facts and figures and may be lacking details, unnatural phrasing, lack of natural transitions between topics, or errors that a human is less likely to make. It often uses alliteration to articulate an appealing word arrangement.
- The absence of relevant very recent on-point case citations may indicate the use of AI generated content. OpenAI models are trained on massive data sets that are not continually updated so if recent relevant cases are not cited, it may be due to the AI being trained on an earlier dataset.
- Humans use idioms and slang frequently. AI often uses these phrases and words incorrectly. If you spot an idiom that feels a bit off and seems forced into the text it is likely a sign it was created with GenAI.

---

[3] *Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools*, preprint study, https://dho.stanford.edu/wp-content/uploads/Legal_RAG_Hallucinations.pdf

Bias

Courts need to be aware of potential bias in the content produced by GenAI. It is important to understand the datasets used in training the model because if they are not diverse or contain incorrect data, the results could be biased or inaccurate. Examples include the initial version of Google's Gemini chatbot that created images of people who did not match the historical ethnic backgrounds, such as creating images of people of color wearing Nazi uniforms[4] and AirCanada's chatbot that gave wrong information about the policy on bereavement travel.[5]

Confidentiality

Any information entered into an open GenAI platform, including through a basic prompt, could become visible to the company operating the platform and other users. Court personnel should be educated to not enter confidential, sensitive or privileged information in a chatbot or GenAI system that uses an open training model. Open systems use the information entered to train the database and will retain the information in the system unless the terms of use for the system explicitly specify that it does not retain the information. If using a chatbot, disable the chat history if possible. Judicial officers and law clerks must avoid inputting confidential or non-public information, including draft decisions and opinions, when using tools that use open models.

Ethics

Judges and court staff need to learn to use GenAI ethically and responsibly and be aware of applicable ethical obligations under the judicial canons and rules of professional responsibility. *See* Ethics section below.

Security

Courts should continue to follow best practices related to cybersecurity in connection with GenAI. When using GenAI that is authorized by the court, court personnel should use court issued equipment and email software so that appropriate security protocols are in place.

---

[4] https://tech.co/news/list-ai-failures-mistakes-errors

[5] Id.

Select a Few Simple "Low Risk" Tasks

Select tasks to be performed with the assistance of GenAI tools that exclusively utilize public data or nonconfidential information and are easily verified for accuracy. Internal facing examples include drafting internal communications and policies, drafting performance evaluations and improvement plans (not including identifying information), drafting training plans for different positions, and conducting basic research. Outward facing examples include summarizing published Supreme Court opinions, drafting press releases about upcoming public events, or drafting committee meeting agendas and minutes. Get comfortable with using the different GenAI tools by starting with internal facing tasks and documents before using AI tools on external facing items.

Use a "Human-in-the-Loop" Approach

GenAI technologies and the use of them in courts are new, and therefore AI-generated output should not be relied upon until it has been reviewed by a human subject matter expert an approach called "Human-in-the-Loop". Presume the output will contain errors and likely bias. Carefully review AI-generated documents and output for accuracy, bias, and completeness. Once more comfortable with the technology (and depending on the task), and its reliability in terms of desired results, accuracy and bias, reevaluate to determine whether the documents and output can be periodically spot-checked by a human to ensure accuracy, instead of checking every document.

Note that the approach may vary with a closed model AI tool. from a reputable vendor having a model that was developed/trained for a specific purpose versus free or low-cost public tools.

Train Staff and Judges on AI Systems

To effectively utilize generative AI technologies, provide training and education to staff and judges on those technologies approved for court use. This helps them understand how to navigate the AI tool, interpret and successfully generate outputs, and effectively review and validate the AI-generated documents or results.

Prepare for Advanced Tasks

As court personnel become more comfortable with utilizing GenAI for basic tasks, consider how it can be used for more advanced tasks, such as data extraction and entry, external facing chatbots for customer service using court self-help and website content, or automated drafting of orders. Conduct pilot projects to test the feasibility and effectiveness of the technology in each specific context. This allows for a controlled testing environment where the technology's impact, benefits, and risks can be assessed.

Engage in Knowledge Sharing

Share what is learned with other courts that are also experimenting with GenAI. This allows for the exchange of experiences, best practices, and lessons learned, enabling courts to make informed decisions and avoid potential pitfalls.

Possible Uses of AI in the Courts

Potentially Useful Tasks

- AI tools are capable of summarizing large amounts of text. As with any summary, care needs to be taken to ensure the summary is accurate.
- AI tools can organize a large amount of information as directed.
- AI tools can find specific information in a large volume of data.
- AI tools can do an acceptable job of creating a first draft of something – a contract, a speech or remarks on a specified topic, job interview questions, position descriptions, performance evaluations, or policy provisions. However, it is essential to review, check, and refine the output and not treat it as a final product. Be aware that different prompts, even with only slightly different wording, will produce different results, so try several prompts to get closer to your desired result.
- AI tools can be used in writing presentations, e.g., to provide suggestions for topics to cover.
- Administrative tasks like composing emails and memoranda can be performed by AI.
- Generating images for presentations. Images often contain hallucinations or accuracies so make sure to closely review to make sure there aren't oddities included. Multiple prompts may be needed to get the desired outcome.

24.3: Company X News

Exhibit 25: **Cited Scholarly Publication** (first two pages)

25.1: *Ling et al. "Domain specialization as the key to make large language models disruptive: A comprehensive survey." ACM Computing Surveys 58, no. 3 (2025): 1-39.*

Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey

CHEN LING, Computer Science, Emory University, Atlanta, United States
XUJIANG ZHAO, NEC Laboratories America Inc, Princeton, United States
JIAYING LU, Computer Science, Emory University, Atlanta, United States
CHENGYUAN DENG, Rutgers The State University of New Jersey, New Brunswick, United States
CAN ZHENG, University of Pittsburgh, Pittsburgh, United States
JUNXIANG WANG, NEC Laboratories America Inc, Princeton, United States
TANMOY CHOWDHURY, George Mason University, Fairfax, United States
YUN LI, George Mason University, Fairfax, United States
HEJIE CUI, Computer Science, Emory University, Atlanta, United States
XUCHAO ZHANG, Microsoft Research, Redmond, United States
TIANJIAO ZHAO, BlackRock Inc, Atlanta, United States
AMIT PANALKAR, BlackRock Inc, Atlanta, United States
DHAGASH MEHTA, Black Rock Inc, Chiyoda, United States
STEFANO PASQUALI, BlackRock Inc, New York, United States
WEI CHENG, NEC Laboratories America Inc, Princeton, United States

Xujiang Zhao, Jiaying Lu, Chengyuan Deng, Can Zheng, Junxiang Wang and Tanmoy Chowdhury contributed equally to this research.
Authors' Contact Information: Chen Ling, Computer Science, Emory University, Atlanta, Georgia, United States; e-mail: chen.ling@emory.edu; Xujiang Zhao (corresponding author), NEC Laboratories America Inc, Princeton, New Jersey, United States; e-mail: xuzhao@neclabs.com; Jiaying Lu, Computer Science, Emory University, Atlanta, Georgia, United States; e-mail: jiaying.lu@emory.edu; Chengyuan Deng, Rutgers The State University of New Jersey, New Brunswick, New Jersey, United States; e-mail: cd751@rutgers.edu; Can Zheng, University of Pittsburgh, Pittsburgh, Pennsylvania, United States; e-mail: caz51@pitt.edu; Junxiang Wang, NEC Laboratories America Inc, Princeton, New Jersey, United States; e-mail: junwang@nec-labs.com; Tanmoy Chowdhury, George Mason University, Fairfax, Virginia, United States; e-mail: tchowdh6@gmu.edu; Yun Li, George Mason University, Fairfax, Virginia, United States; e-mail: yli38@gmu.edu; Hejie Cui, Computer Science, Emory University, Atlanta, Georgia, United States; e-mail: hejie.cui@emory.edu; Xuchao Zhang, Microsoft Research, Redmond, Washington, United States; e-mail: xuchaozhang@microsoft.com; Tianjiao Zhao, BlackRock Inc, Atlanta, Georgia, United States; e-mail: tina.zhao@blackrock.com; Amit Panalkar, BlackRock Inc, Atlanta, Georgia, United States; e-mail: Amit.panalkar@blackrock.com; Dhagash Mehta, Black Rock Inc, Chiyoda, United States; e-mail: dhagash.mehta@blackrock.com; Stefano Pasquali, BlackRock Inc, New York, New York, United States; e-mail: pacca.pasquali@gmail.com; Wei Cheng, NEC Laboratories America Inc, Princeton, New Jersey, United States; e-mail: weicheng@nec-labs.com; Haoyu Wang, NEC Laboratories America Inc, Princeton, New Jersey, United States; e-mail: haoyu@nec-labs.com; Yanchi Liu, NEC Laboratories America Inc, Princeton, New Jersey, United States; e-mail: yanchi@neclabs.com; Zhengzhang Chen, NEC Laboratories America Inc, Princeton, New Jersey, United States; e-mail: zchen@nec-labs.com; Haifeng Chen, NEC Laboratories America Inc, Princeton, New Jersey, United States; e-mail: Haifeng@nec-labs.com; Chris White, NEC Laboratories America Inc, Princeton, New Jersey, United States; e-mail: chris.white@nec-labs.com; Quanquan Gu, University of California Los Angeles, Los Angeles, California, United States; e-mail: qgu@ucla.edu; Jian Pei, Duke University, Durham, North Carolina, United States; e-mail: j.pei@duke.edu; Carl Yang, Computer Science, Emory University, Atlanta, Georgia, United States; e-mail: j.carlyang@emory.edu; Liang Zhao (corresponding author), Computer Science, Emory University, Atlanta, Georgia, Georgia; e-mail: liang.zhao@emory.edu.

ACM 0360-0300/2025/10-ART79
https://doi.org/10.1145/3764579

HAOYU WANG, NEC Laboratories America Inc, Princeton, United States
YANCHI LIU, NEC Laboratories America Inc, Princeton, United States
ZHENGZHANG CHEN, NEC Laboratories America Inc, Princeton, United States
HAIFENG CHEN, NEC Laboratories America Inc, Princeton, United States
CHRIS WHITE, NEC Laboratories America Inc, Princeton, United States
QUANQUAN GU, University of California Los Angeles, Los Angeles, United States
JIAN PEI, Duke University, Durham, United States
CARL YANG, Computer Science, Emory University, Atlanta, United States
LIANG ZHAO, Computer Science, Emory University, Atlanta, United States

Large language models (LLMs) have significantly advanced the field of natural language processing (NLP), providing a highly useful, task-agnostic foundation for a wide range of applications. However, directly applying LLMs to solve sophisticated problems in specific domains meets many hurdles, caused by the heterogeneity of domain data, the sophistication of domain knowledge, the uniqueness of domain objectives, and the diversity of the constraints (e.g., various social norms, cultural conformity, religious beliefs, and ethical standards in the domain applications). Domain specification techniques are key to making large language models disruptive in many applications. Specifically, to solve these hurdles, there has been a notable increase in research and practices conducted in recent years on the domain specialization of LLMs. This emerging field of study, with its substantial potential for impact, necessitates a comprehensive and systematic review to summarize better and guide ongoing work in this area. In this article, we present a comprehensive survey on domain specification techniques for large language models, an emerging direction critical for large language model applications. First, we propose a systematic taxonomy that categorizes the LLM domain-specialization techniques based on the accessibility to LLMs and summarizes the framework for all the subcategories as well as their relations and differences to each other. Second, we present an extensive taxonomy of critical application domains that can benefit dramatically from specialized LLMs, discussing their practical significance and open challenges. Last, we offer our insights into the current research status and future trends in this area.

CCS Concepts: • **Computing methodologies** → **Natural language processing**; *Planning and scheduling;*

Additional Key Words and Phrases: Large language models, natural language processing, domain specialization

ACM Reference Format:
Chen Ling, Xujiang Zhao, Jiaying Lu, Chengyuan Deng, Can Zheng, Junxiang Wang, Tanmoy Chowdhury, Yun Li, Hejie Cui, Xuchao Zhang, Tianjiao Zhao, Amit Panalkar, Dhagash Mehta, Stefano Pasquali, Wei Cheng, Haoyu Wang, Yanchi Liu, Zhengzhang Chen, Haifeng Chen, Chris White, Quanquan Gu, Jian Pei, Carl Yang, and Liang Zhao. 2025. Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey. *ACM Comput. Surv.* 58, 3, Article 79 (October 2025), 39 pages. https://doi.org/10.1145/3764579

1 Introduction

The evolution of **natural language processing** (**NLP**) and **artificial intelligence** (**AI**) models has witnessed a remarkable trajectory, beginning with the rule-based systems of the 1950s and 1960s, transitioning to statistical models in the 1990s, followed by the emergence of neural networks in the 2010s. Owing to the success of self-attention and Transformer-based neural network architecture [167], **Pre-trained Language Models** (**PLMs**) emerged and swiftly gained popularity in the late 2010s due to their ability to learn universal language representations from large-scale data in an unsupervised manner, which can be beneficial for many downstream NLP tasks such as commonsense reasoning [192], multiple-choice question answering [141], and story generation [14], while avoiding training new models from scratch. In the last few years, with the fast growth of large corpus and hardware capacities, researchers have found scaling up model and training data

25.2: *Cottier et al. "The rising costs of training frontier AI models." arXiv preprint arXiv:2405.21015 (2024).*

Luccioni et al. "Power hungry processing: Watts driving the cost of AI deployment?." In Proceedings of the 2024 ACM conference on fairness, accountability, and transparency, pp. 85-99. 2024.

THE RISING COSTS OF TRAINING FRONTIER AI MODELS

Ben Cottier[1] Robi Rahman[1,2]

Loredana Fattorini[2] Nestor Maslej[2] Tamay Besiroglu[1] David Owen[1]

ABSTRACT

The costs of training frontier AI models have grown dramatically in recent years, but there is limited public data on the magnitude and growth of these expenses. This paper develops a detailed cost model to address this gap, estimating training costs using three approaches that account for hardware, energy, cloud rental, and staff expenses. The analysis reveals that the amortized cost to train the most compute-intensive models has grown precipitously at a rate of 2.4× per year since 2016 (90% CI: 2.0× to 2.9×). For key frontier models, such as GPT-4 and Gemini, the most significant expenses are AI accelerator chips and staff costs, each costing tens of millions of dollars. Other notable costs include server components (15-22%), cluster-level interconnect (9-13%), and energy consumption (2-6%). If the trend of growing development costs continues, the largest training runs will cost more than a billion dollars by 2027, meaning that only the most well-funded organizations will be able to finance frontier AI models.

1 Introduction

The large and growing cost of training state-of-the-art AI models has become an important issue in the field of AI [1]. Improving AI capabilities demand exponential increases in computing power, as evidenced by both economic analysis [2] and the discovery of empirical scaling laws, which show that model performance improves with more parameters and training data [3, 4]. Dario Amodei, CEO of the AI lab Anthropic, has stated that frontier AI developers are likely to spend close to a billion dollars on a single training run this year, and up to ten billion-dollar training runs in the next two years [5]. Given this trend, some innovations, particularly those requiring large-scale training, may become inaccessible to all but the most well-funded organizations.

Although it is widely known that training the largest ML models is expensive, until recently there were few concrete estimates of training costs in the public domain. In collaboration with Epoch AI, the 2024 AI Index presented one of the most comprehensive datasets to date, estimating the costs of training runs based on cloud rental prices [6]. We build on that work with a more in-depth account of hardware, energy and R&D staff costs for both training runs and experiments, as well as a more detailed analysis of how costs are increasing over time. To our knowledge, our study is the most thorough analysis of model development costs to date.

Our methods are built upon a comprehensive database of notable machine learning models [7], and informed by interviews with industry experts. We consider three complementary approaches to measuring the cost of frontier models. The first approach estimates the hardware capital expenses (CapEx) amortized over the final training run, along with the cost of hardware energy consumption. By considering AI accelerator chips, other server hardware, networking hardware, and energy separately, this approach can provide more accurate training costs. We find that the most expensive publicly-announced training runs to date are OpenAI's GPT-4 at $40M and Google's Gemini Ultra at $30M. Among frontier models, defined as models within the top 10 most compute-intensive models when they are released, we find that training has become 2.4× more expensive per year since 2016 (90% CI: 2.0× to 2.9×).

We then compare this approach to the cloud-price approach that was first presented in the AI Index [6]. Instead of estimating hourly compute costs in detail, the cloud-price approach simply uses historical rental rates from cloud platforms. The cloud-price approach shows a similar growth rate (2.5× per year with a 90% CI of 2.1× to 3.1×), but

[1]Epoch AI. [2]Stanford University.

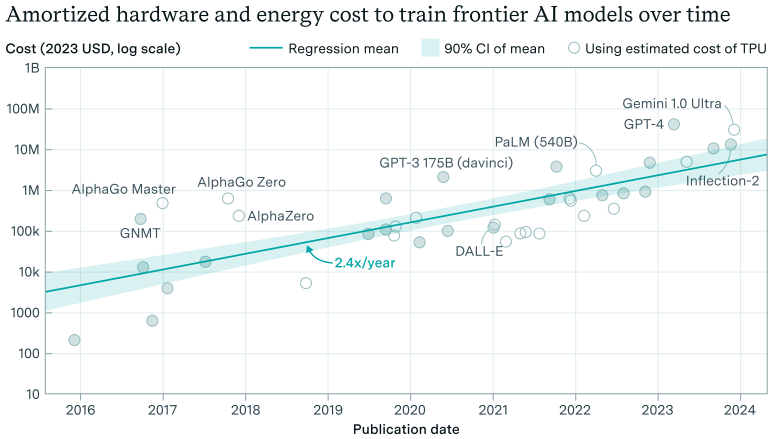


Figure 1: Amortized hardware cost plus energy cost for the final training run of frontier models. The selected models are among the top 10 most compute-intensive for their time. Amortized hardware costs are the product of training chip-hours and a depreciated hardware cost, with 23% overhead added for cluster-level networking. Open circles indicate costs which used an estimated production cost of Google TPU hardware. These costs are generally more uncertain than the others, which used actual price data rather than estimates.

yields costs that are about twice as large on average. We expect the cloud-price approach to overestimate frontier model costs, since model developers usually either own or have private rental agreements for their training hardware. Using both approaches helps validate our estimate of cost growth, while also highlighting the uncertainty of individual costs.

Our third and most in-depth approach breaks down hardware, energy, and R&D staff costs over the entire development of the model (i.e. both experiments and training). We select four especially notable models for this approach—GPT-3, OPT-175B, GPT-4, and Gemini Ultra. For these models, we find that R&D staff costs including equity are between 29% and 49% of the total amortized cost. Computing hardware makes up 47–64%, while energy comprises only 2–6%. However, if we exclude equity the fraction for R&D staff drops to 19–33%, and the fractions of computing hardware costs and energy rise to 61–76% and 2–7% respectively.

By taking into account hardware purchase costs, energy costs, and the more opaque costs of R&D labor, our analysis provides a clearer picture of the true costs of AI development. This sheds light on not only current costs but also the economic hurdles that lie ahead as AI continues to scale.

All of our results can be reproduced using the code and data available at `https://github.com/epoch-research/training-cost-trends`.

2 Methodology

2.1 Datasets and frontier model selection

Our investigation draws upon the Notable AI Models database, which documents 796 notable models across the history of machine learning [7]. Key details captured for each model include training compute, dataset size, and parameter count. To focus on the largest-scale models, we initially filtered the database to models that had training compute estimates and that were published on or after 1 October 2015 (the start of the large-scale ML era according to [8]) and up to 31 December 2023. This resulted in 276 selected models. For these models, we recorded the training time, hardware type and quantity, and utilization rate sourced from each model's original publication, where possible.

For our main results, we examined 41 models that were historically at the frontier of compute. Specifically, we filtered for models that were in the top 10 of training compute as of their release.[1] Appendix B.1 provides further details on this selection procedure and a comparison to three alternative methods.

[1]We excluded models that are fine-tuned versions of a separately listed model, to avoid double-counting costs.

Power Hungry Processing: ⚡ Watts ⚡ Driving the Cost of AI Deployment?

Alexandra Sasha Luccioni
Yacine Jernite
sasha.luccioni@huggingface.co
Hugging Face
Canada/USA

Emma Strubell
Carnegie Mellon University, Allen Institute for AI
USA

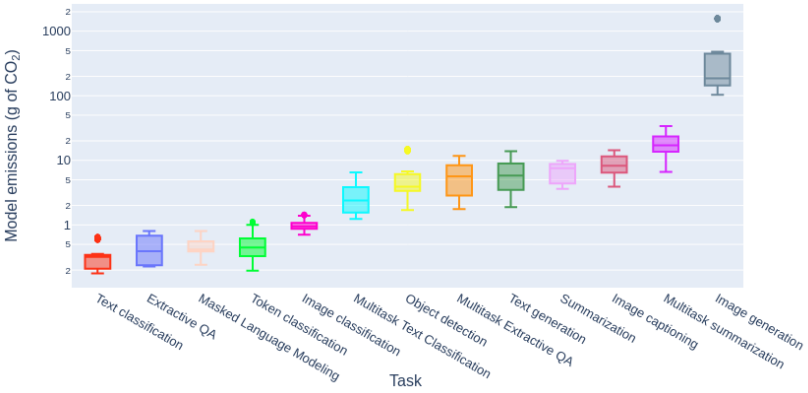


Figure 1: The tasks examined in our study and the average quantity of carbon emissions they produced (in g of CO_2eq) for 1,000 queries. N.B. The y axis is in logarithmic scale.

ABSTRACT

Recent years have seen a surge in the popularity of commercial AI products based on generative, multi-purpose AI systems promising a unified approach to building machine learning (ML) models into technology. However, this ambition of "generality" comes at a steep cost to the environment, given the amount of energy these systems require and the amount of carbon that they emit. In this work, we propose the first systematic comparison of the ongoing inference cost of various categories of ML systems, covering both task-specific (i.e. finetuned models that carry out a single task) and 'general-purpose' models, (i.e. those trained for multiple tasks). We measure deployment cost as the amount of energy and carbon required to perform 1,000 inferences on representative benchmark dataset using these models. We find that multi-purpose, generative architectures are orders of magnitude more expensive than task-specific systems for a variety of tasks, even when controlling for the number of model parameters. We conclude with a discussion around the current trend of deploying multi-purpose generative ML systems, and caution that their utility should be more intentionally weighed against increased costs in terms of energy and emissions. All the data from our study can be accessed via an interactive demo to carry out further exploration and analysis.

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-0450-5/24/06
https://doi.org/10.1145/3630106.3658542

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; **Neural networks**; • **Hardware** → **Impact on the environment**; **Power estimation and optimization**.

ACM Reference Format:
Alexandra Sasha Luccioni, Yacine Jernite, and Emma Strubell. 2024. Power Hungry Processing: ⚡ Watts ⚡ Driving the Cost of AI Deployment?. In *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT '24), June 3–6, 2024, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3630106.3658542

1 INTRODUCTION

Understanding the environmental impacts of different industries is an important first step towards developing effective strategies to mitigate those impacts. For newer industries such as information and communication technologies (ICT) of which Artificial Intelligence (AI) and Machine Learning (ML) are considered to be a part

of, more work is needed to understand the extent of their environmental impacts and the factors that influence it. Between 2017 and 2021, the electricity used by Meta, Amazon, Microsoft, and Google, the main providers of commercially-available cloud compute, more than doubled [22]. According to the most recent figures available, global data centre electricity consumption has grown by 20-40% annually in recent years, reaching 1-1.3% of global electricity demand and contributing 1% of energy-related greenhouse gas emissions in 2022 [21]. However the contribution of the AI sector specifically towards these figures is unclear.

Recent work documenting the environmental impacts of ML has focused largely on quantifying the operational energy and carbon required to perform the training phase of the ML model life cycle [12, 30, 41, 49] due to the relative ease of measuring per-model energy use for that phase and the impressive quantity of energy required to perform a single training run [41, 49]. Yet, other phases of the ML model life cycle, such as inference, stand to impact the environment just as much, or more, than training due to the computational resources required to deploy modern models at scale. While inference on a single example requires much less computation than that required to train the same model, inference happens far more frequently than model training — as many as billions of times a day for a model powering a popular user-facing product such as Google Translate.[1] Yet, in-depth work quantifying the costs of model inference and deployment is limited and their environmental impacts, in terms of energy and carbon as well as water and mining of rare earth minerals, have yet to be estimated. According to AWS, the largest global cloud provider, inference is estimated to make up 80 to 90% of total ML cloud computing demand [2, 28], whereas a 2021 publication by Meta attributed approximately one-third of their internal end-to-end ML carbon footprint to model inference, with the remainder produced by data management, storage, and training [57]; similarly, a 2022 study from Google attributed 60% of its ML energy use to inference, compared to 40% for training [40]. Given the increasing ubiquity of AI model deployment, it is crucial to go beyond these high-level statistics to get a better idea of the energy requirements and carbon emissions of model inference for different models and tasks. In particular, looking at inference rather than training leads to drastically different conclusions when considering the multi-purpose (or "general-purpose") aspect specifically. Training a single model for multiple tasks can indeed be more energy-efficient when considering training costs only, but these gains can easily be lost and even reversed over the course of the model's lifetime, given how much inference is carried out when these models are deployed in user-facing applications like chat and web search.

To help shed light on this issue, we perform an extensive study measuring the amount of energy required to deploy various ML models and architectures, including large language models (LLMs)- as such, our study is, to our knowledge, the first to focus solely on the *inference* phase of the ML model life cycle. We study 88 models across 10 tasks and 30 datasets, spanning applications in natural language and computer vision, analyzing the impact of end task, modality, model size, architecture, and learning paradigm (i.e. task-specific or multi-task/multi-purpose) on energy efficiency. We identify orders-of-magnitude differences in the amount of energy required per inference across models, modalities and tasks and shine light on an important trade-off between the benefit of multi-purpose systems, their energy cost, and ensuing carbon emissions. By painting a more detailed picture of widely varying energy requirements for ML model inference, we hope this study can be useful for practitioners to better understand accuracy-efficiency trade-offs across tasks and models, as well as enabling better estimates, and projections and policy decisions at the sector level.

[1]Google reported translating more than 100 billion words per day in 2016, assuming an average query length of 100 words yields an estimate of 1 billion queries to the model per day. Source: https://blog.google/products/translate/ten-years-of-google-translate/

2 PREVIOUS WORK

Estimating the energy and emissions of ML models has remains a relatively under-explored topic, albeit one that has been gathering traction since Strubell et al's seminal article quantifying the energy and carbon emissions of a variety of then-large NLP models [2019]. Since then, most studies have focused on estimating the energy consumed and carbon emitted during the training phase of neural networks – this includes studies by Patterson et al. [2022, 2021], who compared different models and analyzed factors influencing their emissions. There have also been studies of specific model architectures, e.g. BLOOM [31] and Nour [27], which carried out in-depth analyses of the different steps in the models' life cycle and their relative contribution towards the final quantity of carbon emissions. Given the increasing deployment of ML models in the cloud, several studies have therefore looked at cloud-specific ways to reduce the emissions of ML models such as delayed scheduling, workload elasticity and choosing the least carbon-intensive electricity available Chien et al. [6], Dodge et al. [12], Hanafy et al. [19].

Despite these empirical studies, there is currently a lack of standardized methodology for quantifying and comparing the energy consumption and carbon emissions of ML models. There are several tools that exist, such as Code Carbon [47], MLCO2 [26] and LLM-Carbon [13], all of which adopt different approaches and output different results (see [1] for a detailed comparison). It is therefore difficult to systematically compare the carbon footprints of different models. Existing tools and studies have also largely focused on the dynamic power consumption (i.e. the electricity necessary for powering hardware) and its resulting emissions. However, there have been several proposals to also take into account the embodied emissions of ML models (i.e. the emissions that can be attributed to the manufacturing of computing equipment) into carbon emissions estimates. This has been impeded by a lack of transparency from the designers of common computing hardware such as GPUs, although recent estimates have revealed that the embodied carbon footprint of an LLM trained and deployed on Meta's compute cluster constitutes up to 50% of its carbon footprint [57]. While the majority of existing work has been focused on ML model training given that it is a more tractable part of the model life cycle (i.e. it is most often carried out over a set period of time on a specific compute instance), model inference has started to also become the subject of scholarship [6, 11]. Luccioni et al.'s study of BLOOM was the first of its kind to look at the specific energy costs related to deploying an LLM [31] and found that, over time, this can represent a significant portion of a model's overall carbon footprint.