The United States Government's Role in Harnessing Technology for Future Prosperity:

A Transition Memorandum from The Block Center for Technology and Society at Carnegie Mellon University

COMPILED BY

Carnegie Mellon University
The Block Center
FOR TECHNOLOGY AND SOCIETY

Contents

INTRODUCTION

Today's policymakers face unprecedented complexity. Artificial intelligence, automation, digital platforms, and data systems are transforming every aspect of our economy and society—from how we work and learn to how we care for the health of our citizens. These changes hold immense promise but also risk displacing American workers and losing our position as a global leader in the innovation economy. To address these challenges, we urge the incoming administration and Congress to adopt a forward-thinking and pragmatic approach, leaning on the strength of the United States' research capabilities.

This memorandum outlines actionable policy recommendations to guide the Federal government's efforts to harness the potential of technology while addressing its most pressing risks. Our recommendations focus on six key areas:

- **AI and the Future of Science:** Accelerating scientific research through the capability of AI tools
- **Safe and Reliable AI:** Maintaining trustworthy AI systems the nation will rely on to advance our economy
- **AI and Healthcare:** Revolutionizing the health sciences and bettering patient outcomes
- **The Future of Education:** Preparing our young people to harness the power of AI tools well into the future
- **AI and Energy:** Ensuring we have the capability to power the innovation economy of the future
- **Workforce Training in the age of AI:** Arming American workers with the skills needed to thrive in the future economy

Each section is informed by The Block Center's expert faculty, who are committed to a society where technological progress is aligned with public interest. By implementing these policies, the administration can position the United States as a global leader in responsible innovation, ensuring that technology helps all Americans prosper and thrive.

Executive Summary

The rapid advancement of technology, particularly artificial intelligence (AI), presents both transformative opportunities and complex challenges. The Block Center for Technology and Society at Carnegie Mellon University proposes a comprehensive framework for the incoming policymakers to responsibly leverage technology for economic prosperity, inclusion, and global leadership. This summary highlights actionable policy recommendations across *six critical areas*.

1. AI and the Future of Science

Opportunities: AI promises breakthroughs in scientific discovery, expediting advances in fields like medicine, materials science, and climate modeling.

Recommendations:
- Expand computational resources through initiatives like the National AI Research Resources (NAIRR).
- Develop robust evaluation frameworks for AI systems.
- Support embodied AI research to bridge digital and physical applications, including robotics.
- Enhance data sharing with universal standards and create centralized repositories.
- Fund foundational AI research targeting scientific acceleration and interdisciplinary collaboration.

2. Safe and Responsible AI

Opportunities: Ensuring the safety and trustworthiness of AI systems is critical for societal adoption and innovation.

Recommendations:
- Invest in AI measurement, testing and evaluation infrastructure
- Develop domain-specific standards for AI testing and validation.
- Enhance the capabilities of existing regulatory agencies to evaluate and assess AI tools.
- Implement tiered transparency standards and create auditable interaction logs.
- Clarify liability frameworks to address risks associated with AI misuse.

3. AI and Healthcare

Opportunities: Despite spending the largest percentage of GDP on healthcare, the US has the worst health outcomes of any developed country. Invest in dramatically increasing access to healthcare and reducing administrative inefficiency using AI in order to eliminate death due to preventable and treatable causes
.

Recommendations:
- Build secure, inclusive data-sharing platforms for AI model training.
- Develop AI tools tailored for low-resource healthcare settings.
- Invest in public health-oriented AI applications, such as vaccine distribution and hospital management.
- Promote hybrid human-AI healthcare systems to preserve patient trust and relational care.
- Establish clear accountability frameworks and adaptive regulations to address ethical concerns and liability.

4. The Future of Education

Opportunities: AI can enhance learning outcomes by personalizing education, streamlining teacher workloads, and fostering critical thinking.

Recommendations:
- Pilot AI programs to adapt to student and educator needs incrementally.
- Ensure equitable access to AI tools and training, particularly in underserved areas.
- Strengthen data privacy standards and promote ethical AI use in classrooms.
- Emphasize AI as a complementary teaching aid rather than a replacement for educators.

Recommendations:

- Pilot AI programs to adapt to student and educator needs incrementally.
- Ensure equitable access to AI tools and training, particularly in underserved areas.
- Strengthen data privacy standards and promote ethical AI use in classrooms.
- Emphasize AI as a complementary teaching aid rather than a replacement for educators.

5. AI and Energy

Opportunities: AI requires tremendous amounts of energy for its development and training. However, the technology can also support energy solutions that will also reduce negative climate impacts through enhanced data analytics, grid optimization and predictive modeling.

Recommendations:
- Foster industry collaboration for sustainable energy use.
- Standardize energy reporting metrics to improve transparency and reduce reliance on misleading carbon neutrality claims.
- Adopt advanced environmental efficiency metrics such as Energy Reuse Factor (ERF) and Water Usage Effectiveness (WUE).
- Invest in grid modernization and support clean energy sources, including nuclear power.
- Address e-waste and cooling demands with innovative recycling and cooling technologies.

6. Workforce in the Age of AI

Opportunities: AI-driven transformation necessitates upskilling and reskilling to prepare society to meet new workforce and skill demands. Invest in an AI GI Bill for reskilling and upskilling workers for the AI economy.

Recommendations:
- Convene stakeholders to develop infrastructure for monitoring AI's impact on employment.
- Design AI-focused skills transition programs and promote micro-credential pathways.
- Support apprenticeship models combining education with practical training.
- Create shared learning platforms for accessible, continuous skill development.

AI AND THE FUTURE OF SCIENCE

BACKGROUND:

Recent dramatic advances in AI, including Large Language Models such as GPT, Claude, and Gemini raise the possibility that one very positive impact of AI might be to dramatically accelerate research progress across a wide variety of scientific fields, from cell biology, to materials science, to weather and climate modeling, to neuroscience (Toner-Rodgers 2024). AI provides the opportunity to shift from a traditional "lone ranger" approach to a collaborative "community scientific discovery" model (Mitchell 2024). This shift would leverage AI's strengths in data analysis, hypothesis generation, and automated experimentation to address current limitations in scientific research. By taking the outlined steps below, the government can help enable this shift in scientific discovery and play a transformative role in how scientific research is conducted in the United States.The paradigm shift to AI-enhanced, community-driven scientific discovery has the potential to transform several fields by expediting breakthrough advancements.

- **Medicine**: AI could reduce the time and cost required to develop new vaccines for disease outbreaks by a factor of ten.
- **Material Sciences**: AI could accelerate the discovery of transformative materials, such as room-temperature superconductors and thermoelectric materials that convert heat to electricity without emissions.
- **Cell Biology**: AI-enabled foundational models could integrate vast experimental datasets to simulate cell behavior, allowing in silico testing before costly in vivo experiments.
- **Neuroscience**: Foundational models could unify data from single neuron to full-brain imaging, predicting neural responses, medication impacts, and therapeutic outcomes for mental health treatments.
- **Meteorology**: AI can improve weather forecasting by providing highly localized predictions and extending forecast horizons, benefiting agriculture and disaster preparedness (Mitchell 2024).

This transformation could position the United States as the leader in scientific discovery and innovation well into the future.

Our Recommendations:

Given this context, The Block Center for Technology and Society is making the following recommendations to address particular policy challenges and opportunities that the next presidential administration must prioritize and address to take advantage of the many exciting opportunities for AI to transform scientific research and entrench the U.S. as the global leader in science and technology.

Expanding the computational and data resources available to researchers using AI for scientific discovery

Challenge: The computational and data requirements create a strong barrier to entry for researchers. Moreover, the AI needs of different scientific research communities will likely differ, and so resources need to be at least partially tailored to meet the needs of these different research communities.

Recommendation: The National AI Research Resources (NAIRR), which NAIAC supported in its first-year report, is an important step in making powerful resources broadly available to the research community, and its current pilot shows the feasibility of this strategy. The administration should increase investment to expand the availability of the NAIRR as a broad research resource, and at the same time begin developing additional resources that are specialized to high-value areas of science (e.g., a network of cloud science labs) where AI can have an impact.

Develop a science base for the evaluation of AI systems.

Challenge: Our understanding of how to evaluate AI systems, and how to provide guarantees for their behavior, has thus far not risen to these challenges; it has lagged behind our ability to create powerful AI algorithms.

Recommendation: The administration should fund research programs through NIST, DOE, DOD, DOT, and NSF aimed at creating a new science base for the evaluation of AI systems, with a particular focus on frontier AI systems and their deployment in consequential application settings. This should lead to advancing the state of the art of the measurement and science of AI evaluation.

Support research to advance integration of AI into physical systems with applications in areas such as manufacturing and robotics

Challenge: The rapid advances in AI (including generative AI and multi-modal language models) and their applications in the digital world need to be advanced to applications in the physical world. There is significant opportunity to cross-fertilize AI advances and benefit important application areas such as manufacturing. Embodied AI is a field that addresses the challenges that arise in applying advances in cognitive AI to physical world systems.

Recommendation: The administration should increase funding for research into embodied AI systems via agencies such as NSF, NIST, DOE, DOT and DOD and consider challenges (analogous to the DARPA challenges) to drive translation impact of embodied AI systems research.

Accelerate the creation of new experimental datasets to train new foundation models, and to make data available to the full community of scientists

Opportunity: The key to AI's role in the transformation in scientific research is access to universally accessible datasets and scientific literature. Such access will enable researchers to utilize the automation power of AI tools to accelerate the scientific process and take advantage of the resource saving advantages of simulated experimentation before moving to real-world experimentation. To achieve multiple order-of-magnitude advances in science, and to train the types of foundational models we desire, the scientific community will require a very significant advance in our ability to share and jointly analyze diverse data sets contributed across the entire scientific community

Recommendation: Create data sharing standards to make it easy for one scientist to (re)use the experimental data created by a different scientist, and to form the basis for a national data resource in each relevant science. Note there are earlier successes in setting and using such standards, that can provide starting templates for standards efforts (e.g., the success in sharing data in the human genome project).

Recommendation: Create and support data sharing websites for each relevant field. Just as GitHub has become the go-to website for software developers to contribute, share and reuse software code, create a GitHub for scientific data sets that serves as both data repository and search engine for discovering data sets most relevant to a particular topic, hypothesis, or planned experiment.

Recommendation: Conduct a study of how to construct incentives to maximize data sharing. Currently, scientific fields vary widely in the degree to which individual scientists share data, and the degree to which for-profit institutions make their data available for basic scientific research. Building a large, sharable national data resource is such an integral component of the AI-science opportunity, that constructing a compelling incentive structure for data sharing will be key to success.

Recommendation: Where appropriate, fund development of automated laboratories (e.g., robotic labs for experiments in chemistry, biology, etc., accessible to a wide collection of scientists over the internet) to efficiently run experiments, and to produce data in a standard format. One major side-benefit of creating such laboratories is that they will also drive the development of standards for stating precisely the experimental procedure to be followed, thereby improving reproducibility of experimental results. Just as we can benefit from a GitHub for data sets, we can also benefit from a related GitHub for sharing, modifying and reusing components of experimental protocols.

Create the New Generation of AI Tools Needed to Accelerate Scientific Research

Opportunity: Current machine learning methods have been found to be extremely valuable for discovering statistical regularities in data sets too large for human inspection (e.g., AlphaFold was trained on a large set of protein sequences and their painstakingly measured 3D structures).The recent advent of LLMs with advanced capabilities to digest, summarize, and reason about large text collections can form the basis for new machine learning algorithms that enable accelerated research.

Recommendation: Fund multi-institutional research teams in a variety of specific scientific fields, to produce a vision and preliminary results showing how AI might be used to dramatically accelerate progress in their field, and what is needed to scale the approach. This effort should NOT be funded in grants to individual institutions, because the biggest advances are likely to come from integrating data and studies across many scientists at many institutions. Instead this is likely to be most effective if performed by teams of scientists across many institutions, proposing opportunities and approaches that carry with them the incentives to engage their full scientific community.

Recommendation: Fund relevant basic AI research specifically targeted to develop approaches applicable to scientific research. This should include developing “foundation models” interpreted broadly, as tools to accelerate research in different fields, and to accelerate the paradigm shift from “lone ranger” science to a more powerful “community scientific discovery” paradigm.

SAFE AND RESPONSIBLE AI

BACKGROUND:

AI technologies have increasingly become interwoven into everyday life. Rigorous studies have begun to demonstrate how AI can dramatically increase productivity and revolutionize how individuals work, learn, and innovate (Noy and Zhang, 2023; Toner-Rodgers, 2024). However, they also have the potential to cause harm. Because AI learns from existing datasets, users, or conditions, these systems have already unintentionally replicated biases, such as struggling to identify individuals with darker skin tones (Frase & Daniels 2023) or reinforcing racist and sexist stereotypes (Jones et. al 2023). Some AI systems have also been intentionally misused to harm others, such as in scams to defraud individuals (Frase & Daniels 2023), the creation of fake media or propaganda (Hendrycks et. al, 2023), cyberbullying or harassment (Callegari 2024), or the use of deepfakes to create sexual abuse or harassment materials (Callegari 2024). Beyond this, researchers have noted the ethical dilemmas around implementation of AI, such as how to balance AI growth and innovation with right to privacy and how to determine in what contexts or industries it is appropriate and responsible to use AI (Jones et. al 2023). Companies developing AI tools may also struggle to mitigate these risks themselves as they seek to balance safety with their own profits (Jones et. al 2023, Hendrycks et. al 2023).

Our Recommendations:

Given this context, The Block Center for Technology and Society is making the following recommendations to address particular policy challenges and opportunities that the next presidential administration must prioritize and address to address the concerns of the American public and promote the Safe and Responsible use of AI systems in public life.

Increasing Representation in AI Models

Challenge: Currently, fragmented data sharing and evaluation practices hinder robust model testing and training, particularly for clinical decision-support tools that rely on sensitive, individual-level data. This has consequences for the generalizability of models, often resulting in site-specific tools that do not apply well to new populations. This disproportionately impacts underserved populations who are often underrepresented in training data, resulting in models with poor performance on these groups. While broader data sharing could lead to more equitable benefit of AI in healthcare, it poses risks to patient privacy.

Recommendation: Policymakers should encourage secure, cross-institutional data platforms that support broad data pooling for AI training. This requires common data standards and secure data-sharing platforms that enable federated learning across multiple sites to preserve data privacy while expanding training across diverse populations. Additionally, creating systems to streamline model evaluation across different subpopulations can improve model robustness and inclusivity. These actions will increase the utility of models in practice, and particularly for underserved populations.

Broadening the Reach of AI Models

Challenge: Underserved populations are often treated by healthcare systems with less comprehensive data (e.g., less imaging) and without AI model development resources. AI could widen healthcare disparities, accelerating care for well-funded research-focused centers and failing to reach marginalized or rural populations.

Recommendation: The recommendations above regarding model generalizability will partially mitigate these issues. Furthermore, investment should be made towards AI models tailored for low-resource settings, given the differences in data and resources.

Looking Beyond Clinical Applications: Operations and Public Health

Challenge: Beyond clinical settings, AI applications in resource optimization and public health can provide substantial efficiencies, especially in resource-limited settings, increasing access and reducing patient wait times in underserved areas.

Recommendation: Investments should be directed toward operations and public health-oriented AI applications. For example, AI can be used to schedule home health care workers, optimize vaccine distribution, or manage patient flows through a hospital system (El-Bouri et al, 2022). These applications increase healthcare access through operational efficiencies. Furthermore, these problems often require less sensitive health data, lowering the barrier to research and model deployment.

Leveraging experts in AI for healthcare

Challenge: The replacement of certain human healthcare roles by AI raises ethical concerns, particularly regarding patient trust and the human relational aspects of care that are difficult for AI to replicate.

Recommendation: Hybrid human - algorithm systems can mitigate these concerns. The integration of humans and algorithms into joint decision-making pipelines has drawn interest across various domains (McKinsey 2024) and healthcare is particularly well-suited for these approaches. In such systems, AI complements, rather than replaces, human healthcare providers. Human oversight should be maintained to preserve the relational aspect of care, which fosters patient trust, improves continuity, adherence, and outcomes(Haselager et al, 2024). This approach allows AI and clinicians to work synergistically for better patient outcomes.

Liability and Accountability

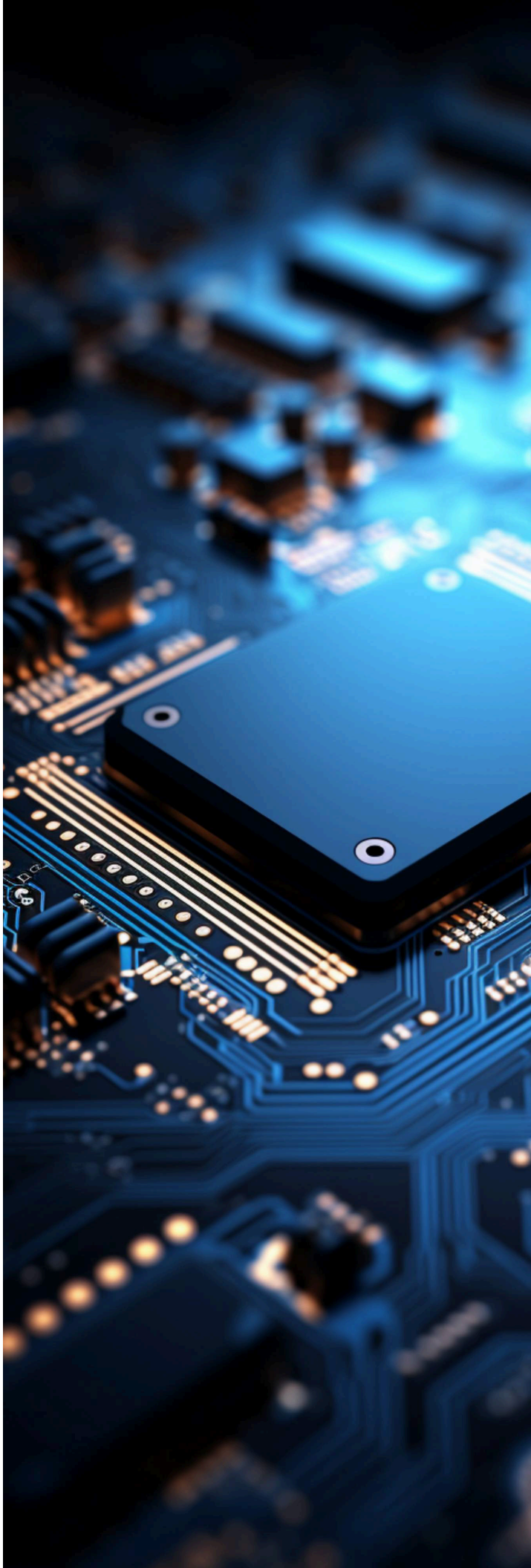
Challenge: There is ambiguity in accountability when AI-assisted decisions lead to adverse outcomes, making it unclear if the responsibility lies with the clinician or the model developer. This uncertainty creates obstacles in adoption and misalignments between stakeholders.

Recommendation: Regulatory guidelines are needed to clarify accountability for AI-assisted decisions, specifying who is responsible in adverse cases. These guidelines would provide clear liability structures, making AI model adoption more feasible by aligning the interests of stakeholders, including health systems and developers.

Adaptive Regulatory Frameworks

Challenge: The rapid pace of AI innovation requires regulatory frameworks that evolve continuously. Overly rigid regulations could stifle innovation, while insufficiently rigorous guidelines might risk patient safety.

Recommendation: A non-prescriptive regulatory approach should be adopted, enforcing rigorous standards for evaluation and deployment but allowing adaptability as technology advances and new application areas emerge. Frameworks should include diverse population evaluations, ongoing monitoring, and contingency plans for addressing data drift and technological changes, ensuring both innovation and safety.



AI AND HEALTHCARE

BACKGROUND:

The expansion of AI technology in the healthcare realm has the potential to critically advance the field of medicine, while simultaneously posing risks and uncertainties to be addressed. AI has demonstrated success in diagnostics, early detection, treatment optimization, and operational improvements, among other areas, which collectively can greatly improve patient outcomes (Lyakhova & Lyakhov, 2024, Snider et. al, 2024). Companies like Abridge, founded by CMU faculty, are expanding the economic opportunities of healthcare and AI.

Despite promising advances, there are concerns surrounding the adoption of AI in the healthcare sector. For one, there are concerns that biased AI algorithms lead to increasing health disparities in minority groups created by underrepresentation in health survey data (Office of Minority Health, 2023). Additionally, wide distrust over AI in the healthcare industry has been reported, with the privacy of patient's health data being a large concern (Ronanki, 2024). The breakthrough of AI in the industry also raises concern surrounding job displacement among healthcare workers as the scope of their work has the potential to be massively redefined (Rony et. al, 2024).

Ultimately, concerns over algorithmic bias, patient data privacy and job displacement in the medical field must be addressed as we simultaneously welcome the advances in the medical field brought on by the inclusion of AI in the healthcare sector.

Our Recommendations:

Given this context, The Block Center for Technology and Society is making the following recommendations to address particular policy challenges and opportunities, such as securing representative health data and increasing access to AI tools in the clinical setting, that the next presidential administration must prioritize and address to take advantage of the many exciting opportunities for AI to improve patient outcomes.

Enhancing our healthcare outcomes and efficiency

Challenge: The United States excels at delivering care to patients relative to its peers, but lags behind in administrative efficiency and health outcomes despite its high spending on healthcare (Blumenthal et al., 2024).

Recommendation: In an effort to halve the number of deaths from treatable or preventable medical conditions, provide funding to healthcare systems interested in utilizing AI models trained in improving access to healthcare, affordability, and administrative efficiency.

Increasing representation in AI models

Challenge: Currently, fragmented data sharing and evaluation practices hinder robust model testing and training, particularly for clinical decision-support tools that rely on sensitive, individual-level data. This has consequences for the generalizability of models, often resulting in site-specific tools that do not apply well to new populations. This disproportionately impacts underserved populations who are often underrepresented in training data, resulting in models with poor performance on these groups. While broader data sharing could lead to more equitable benefit of AI in healthcare, it poses risks to patient privacy.

Recommendation: Policymakers should encourage secure, cross-institutional data platforms that support broad data pooling for AI training. This requires common data standards and secure data-sharing platforms that enable federated learning across multiple sites to preserve data privacy while expanding training across diverse populations. Additionally, creating systems to streamline model evaluation across different subpopulations can improve model robustness and inclusivity. These actions will increase the utility of models in practice, and particularly for underserved populations.

Broadening the reach of AI models

Challenge: Underserved populations are often treated by healthcare systems with less comprehensive data (e.g., less imaging) and without AI model development resources. AI could widen healthcare disparities, accelerating care for well-funded research-focused centers and failing to reach marginalized or rural populations.

Recommendation: The recommendations above regarding model generalizability will partially mitigate these issues. Furthermore, investment should be made towards AI models tailored for low-resource settings, given the differences in data and resources.

Looking beyond clinical applications: operations and public health

Challenge: Beyond clinical settings, AI applications in resource optimization and public health can provide substantial efficiencies, especially in resource-limited settings, increasing access and reducing patient wait times in underserved areas.

Recommendation: Investments should be directed toward operations and public health-oriented AI applications. For example, AI can be used to schedule home health care workers, optimize vaccine distribution, or manage patient flows through a hospital system (El-Bouri et al, 2022). These applications increase healthcare access through operational efficiencies. Furthermore, these problems often require less sensitive health data, lowering the barrier to research and model deployment.

Leveraging experts in AI for healthcare

Challenge: The replacement of certain human healthcare roles by AI raises ethical concerns, particularly regarding patient trust and the human relational aspects of care that are difficult for AI to replicate.

Recommendation: Hybrid human - algorithm systems can mitigate these concerns. The integration of humans and algorithms into joint decision-making pipelines has drawn interest across various domains (McKinsey 2024) and healthcare is particularly well-suited for these approaches. In such systems, AI complements, rather than replaces, human healthcare providers. Human oversight should be maintained to preserve the relational aspect of care, which fosters patient trust, improves continuity, adherence, and outcomes(Haselager et al, 2024). This approach allows AI and clinicians to work synergistically for better patient outcomes.

Liability and Accountability

Challenge: There is ambiguity in accountability when AI-assisted decisions lead to adverse outcomes, making it unclear if the responsibility lies with the clinician or the model developer. This uncertainty creates obstacles in adoption and misalignments between stakeholders.

Recommendation: Regulatory guidelines are needed to clarify accountability for AI-assisted decisions, specifying who is responsible in adverse cases. These guidelines would provide clear liability structures, making AI model adoption more feasible by aligning the interests of stakeholders, including health systems and developers.

Adaptive Regulatory Frameworks

Challenge: The rapid pace of AI innovation requires regulatory frameworks that evolve continuously. Overly rigid regulations could stifle innovation, while insufficiently rigorous guidelines might risk patient safety.

Recommendation: A non-prescriptive regulatory approach should be adopted, enforcing rigorous standards for evaluation and deployment but allowing adaptability as technology advances and new application areas emerge. Frameworks should include diverse population evaluations, ongoing monitoring, and contingency plans for addressing data drift and technological changes, ensuring both innovation and safety.

THE FUTURE OF EDUCATION

BACKGROUND:

The commentary by Prof. Ken Koedinger of Carnegie Mellon on AI and education (Koedinger, 2023) highlights both the opportunities and challenges of developing AI for education. A recent report on the rigorous evaluation of helping teachers with computer aided learning using technology from Khan Academy showed significant student gains in student learning (Oreopoulos et al, 2024). However, a review of the broader literature documents work both about the potential harm as well as the benefit that could accrue from this technology.

Approximately a quarter of public K-12 teachers believed AI tools do more harm than good in education (Lin 2024). This view is particularly prevalent among high school teachers (Lin 2024), who cite issues such as encountering students using AI tools to write their papers, AI tools that "hallucinate" incorrect information, and students relying too heavily on these tools instead of their own critical thinking (Blaise 2023). A startling 20 percent of teens that are familiar with ChatGPT reported thinking it was acceptable to use the tool to write essays (Lin 2024).

However, 32 percent of teachers believed AI tools in education are equally beneficial as they are harmful (Lin 2024). Proponents of AI in the classroom note that these tools can be used in ways that augment student learning and potentially lighten teacher workload. For example, teachers who have incorporated AI tools into their classrooms commonly report using these tools to adapt educational material to meet varied student skill levels and to generate starting points for materials such as lesson plans and documentation (Diliberti et. al 2024). When properly used in the classroom, AI tools could provide more opportunities for prosocial, small-group learning environments and "play" as an element of learning" (Partelow 2024).

Our Recommendations:

Given this context, The Block Center for Technology and Society is making the following recommendations to address particular policy challenges and opportunities that the next presidential administration must prioritize and address to take advantage of the many exciting opportunities for AI to improve educational outcomes in the United States.

Addressing the Developmental Appropriateness of AI Tools

Challenge: AI tools must be carefully designed to align with students' cognitive, social, and emotional development across age groups, ensuring that technology complements rather than dominates foundational learning experiences.

Opportunity: Age-specific applications of AI can enhance learning by reinforcing developmental goals without detracting from critical in-person interactions.

Recommendation: Implement developmental guidelines for AI in education. For pre-adolescents, AI should remain interactive and supervised, supporting foundational skills in areas like math and literacy while prioritizing social interaction with peers and teachers. For adolescents, AI can support critical thinking and independent study skills, and for university students, it can provide advanced research and technical skill enhancement.

Equitable Access and Inclusivity

Challenge: Socioeconomic disparities can limit access to AI-enhanced education, risking inequality in learning opportunities and skill development.

Opportunity: Providing equitable access to AI tools in education can democratize learning outcomes, giving all students the resources to succeed academically and professionally.

Recommendation: Allocate resources to ensure that AI tools are accessible to students in underserved areas and provide training on AI-related skills, such as prompt engineering and data evaluation. At the university level, AI should be viewed as a "co-learner," with curricula designed to encourage critical engagement with AI outputs.

Ensuring Data Privacy and Security

Challenge: The integration of AI in education raises significant concerns regarding student data privacy and security, particularly given the sensitivity of educational and personal data.

Opportunity: Strong privacy protocols can protect students while fostering trust in AI applications in educational settings.

Recommendation: Strengthen data privacy protocols by requiring transparency from AI vendors about data collection, storage, and sharing. Universities should implement clear standards for AI-assisted learning to protect student data, supported by monitoring from an Office for AI in Higher Education.

Emphasizing Empowerment, Not Displacement

Challenge: AI has the potential to overshadow educators rather than support them if not carefully managed, which could diminish the value of direct teacher-student interactions.

Opportunity: When positioned as a teaching aid, AI can empower educators by automating routine tasks and allowing them to focus on personalized, higher-order interactions with students.

Recommendation: Integrate AI into curricula as a skill-enhancing tool that supports educators, not as a replacement. AI should streamline repetitive tasks, enabling educators to emphasize critical thinking, creativity, and personal guidance.

Promoting Ethical and Responsible Use

Challenge: AI systems in education need ongoing assessment to ensure they are free of biases and contribute positively to students' motivation, self-perception, and critical thinking skills.

Opportunity: Ethical use of AI can enhance student development by promoting responsible and unbiased learning environments.

Recommendation: Create ethical guidelines for AI in education, with a focus on preventing biases and ensuring AI enhances rather than undermines student motivation and critical thinking. The proposed Office for AI in Higher Education could monitor ethical compliance and guide AI's responsible use in universities.

AI AND ENERGY

BACKGROUND:

AI offers a powerful tool in the race to address climate change. By enabling advanced data analysis, monitoring, and predictive modeling, AI can help us better understand and address climate challenges (Gentine et al., 2024). However, AI itself can contribute to increased energy consumption and greenhouse gas emissions.

Data centers, the backbone of AI applications, are significant consumers of electricity. As AI usage grows, so too does the energy demand for data centers. In fact, data centers are projected to consume over 1,000 terawatt-hours of electricity by 2026, roughly equivalent to the annual electricity consumption of a nation like Japan (International Energy Agency, 2024).

The energy footprint of AI is twofold. First, training AI models, particularly large language models like GPT-3, is an energy-intensive process. Training GPT-3, for instance, produces approximately 500 metric tons of greenhouse gas emissions (Keller et al., 2024). Second, the operation of AI models, such as generative AI chatbots, requires significant computational resources. A single query to a generative AI chatbot can consume four to ten times more energy than a standard search engine query (Keller et al., 2024; The Economist, 2024).

While advancements in hardware efficiency can help mitigate these energy demands (Cowls et al., 2021), it's crucial to address the environmental impact of AI. To harness the benefits of AI while minimizing its negative environmental consequences, policymakers must adopt a balanced approach. By promoting the development of energy-efficient AI technologies and implementing sustainable data center practices, we can ensure that AI serves as a tool for climate action rather than a contributor to the problem.

Our Recommendations:

Given this context, The Block Center for Technology and Society and the Scott Institute for Energy Innovation are making the following recommendations to address particular policy challenges that the next Presidential administration must prioritize and address to modernize our energy infrastructure and ensure a crucial emerging industry enables the U.S. to meet its climate objectives.

Coordinating American Resources

Challenge: Making the US the global leader in AI will require large, stable and sustainable sources of energy as well as sustainable use of scarce resources such as water.

Recommendation: Foster industry collaboration to encourage public-private partnerships to drive innovation to achieve these goals and advocate for global cooperation to establish benchmarks and cross-border environmental impact goals.

Lack of Transparency in Energy Reporting

Challenge: Big tech companies use Renewable Energy Certificates (RECs) to claim carbon neutrality, which obscures actual energy use and emissions. Limited location-based emissions reporting creates a misleading picture of the environmental impact.

Recommendation: Develop and enforce standardized energy reporting metrics that emphasize location-based emissions and move away from reliance on RECs and Mandate disclosures specific to data centers' environmental impact.

Inadequate Metrics for AI's Environmental Impact

Challenge: Current metrics like Power Usage Effectiveness (PUE) fail to capture the complexity of AI workloads, focusing only on energy input without considering efficiency or output. A broader set of metrics is needed to evaluate energy, water usage, carbon emissions, and computational efficiency.

Recommendation: Adopt advanced metrics that Implement new metrics like Energy Reuse Factor (ERF), Carbon Usage Effectiveness (CUE), and Water Usage Effectiveness (WUE) to provide a more comprehensive view of environmental efficiency. Use computational efficiency metrics like Performance per Watt (PPW) and energy intensity to optimize AI applications.

Legislative Gaps and Infrastructure Limitations

Challenge: Grid modernization and access to clean energy are critical bottlenecks, alongside long connection lead times and infrastructure challenges.

Recommendation: Enact legislation to address AI's energy demands and promote sustainable practices, such as encouraging renewable energy use and investing in grid modernization. Support nuclear energy as a clean power source through frameworks like the ADVANCE Act.

E-Waste and Cooling Challenges

Challenge: AI's reliance on high-performance hardware contributes to significant e-waste.Data centers consume massive water resources for cooling, adding to their environmental burden.

Recommendation: Address E-Waste and Cooling Demands by promoting recycling and sustainable practices for e-waste generated by AI infrastructure and investing in innovative cooling technologies to reduce energy and water consumption.

WORKFORCE IN THE AGE OF AI

BACKGROUND:

In 2024, the National Academies Committee, co-chaired by Block Center Advisory Board Member Erik Brynjolfsson and Block Center Chief Technologist Tom Mitchell, released a report on AI and the Future of Work (National Academies Committee, 2024). Among its key findings were that a) the rapidly increasing capabilities of AI combined with complementary investments in processes and new skills will likely significantly increase productivity, and b) an exclusive focus on worker displacement could overlook positive developments such as new forms of work that demand valuable new expertise and AI systems that work jointly with workers to enable them to use their expertise more effectively with less formal training.

In parallel, there has been work articulating the scale of the potential impact AI could have. The International Monetary Fund estimated that nearly 40 percent of jobs globally will be impacted by AI, with that rate as high as 60 percent in advanced economies (Ospina 2024). This impact may be felt as certain tasks or jobs are automated – McKinsey Global Institute estimated that 29.5 percent of all hours work could be automated using AI by 2030 – but it may also be felt as the distribution of tasks or necessary skills in a workplace shifts due to use of AI (Khattar 2023, Tamayo et. al 2023). In occupations that are highly exposed to AI, workplaces are increasingly demanding "soft skills", such as management skills and social and emotional skills (Green 2024, Khattar 2023). More job postings are also demanding physical skills related to production and technology (Green 2024) or have shifted to requiring more highly educated workers, particularly those with STEM degrees (Khattar 2023). As AI transforms what skills are needed and valued versus what can be automated, it has drastically reduced the half-life of skills, particularly in technology fields (Tamayo et. al 2023). Skill development and reskilling workers has become a strategic imperative, both to help companies and developing industries quickly develop the talent they need (George et al., 2022; George et al., 2024; Tamayo et al., 2023), and to support workers at risk of automation, who are often disproportionately women, Latino, or Black workers (Khattar 2023).

Our Recommendations:

Given the landscape that workers face, The Block Center for Technology and Society is making the following recommendations to address particular policy opportunities and challenges that the next presidential administration must prioritize and address for the benefit of their constituents and the further development of our economy and workforce for the modern age.

Increasing Representation in AI Models

Challenge: Currently, fragmented data sharing and evaluation practices hinder robust model testing and training, particularly for clinical decision-support tools that rely on sensitive, individual-level data. This has consequences for the generalizability of models, often resulting in site-specific tools that do not apply well to new populations. This disproportionately impacts underserved populations who are often underrepresented in training data, resulting in models with poor performance on these groups. While broader data sharing could lead to more equitable benefit of AI in healthcare, it poses risks to patient privacy.

Recommendation: Policymakers should encourage secure, cross-institutional data platforms that support broad data pooling for AI training. This requires common data standards and secure data-sharing platforms that enable federated learning across multiple sites to preserve data privacy while expanding training across diverse populations. Additionally, creating systems to streamline model evaluation across different subpopulations can improve model robustness and inclusivity. These actions will increase the utility of models in practice, and particularly for underserved populations.

Broadening the Reach of AI Models

Challenge: Underserved populations are often treated by healthcare systems with less comprehensive data (e.g., less imaging) and without AI model development resources. AI could widen healthcare disparities, accelerating care for well-funded research-focused centers and failing to reach marginalized or rural populations.

Recommendation: The recommendations above regarding model generalizability will partially mitigate these issues. Furthermore, investment should be made towards AI models tailored for low-resource settings, given the differences in data and resources.

Looking Beyond Clinical Applications: Operations and Public Health

Challenge: Beyond clinical settings, AI applications in resource optimization and public health can provide substantial efficiencies, especially in resource-limited settings, increasing access and reducing patient wait times in underserved areas.

Recommendation: Investments should be directed toward operations and public health-oriented AI applications. For example, AI can be used to schedule home health care workers, optimize vaccine distribution, or manage patient flows through a hospital system (El-Bouri et al, 2022). These applications increase healthcare access through operational efficiencies. Furthermore, these problems often require less sensitive health data, lowering the barrier to research and model deployment.

Leveraging experts in AI for healthcare

Challenge: The replacement of certain human healthcare roles by AI raises ethical concerns, particularly regarding patient trust and the human relational aspects of care that are difficult for AI to replicate.

Recommendation: Hybrid human - algorithm systems can mitigate these concerns. The integration of humans and algorithms into joint decision-making pipelines has drawn interest across various domains (McKinsey 2024) and healthcare is particularly well-suited for these approaches. In such systems, AI complements, rather than replaces, human healthcare providers. Human oversight should be maintained to preserve the relational aspect of care, which fosters patient trust, improves continuity, adherence, and outcomes(Haselager et al, 2024). This approach allows AI and clinicians to work synergistically for better patient outcomes.

Liability and Accountability

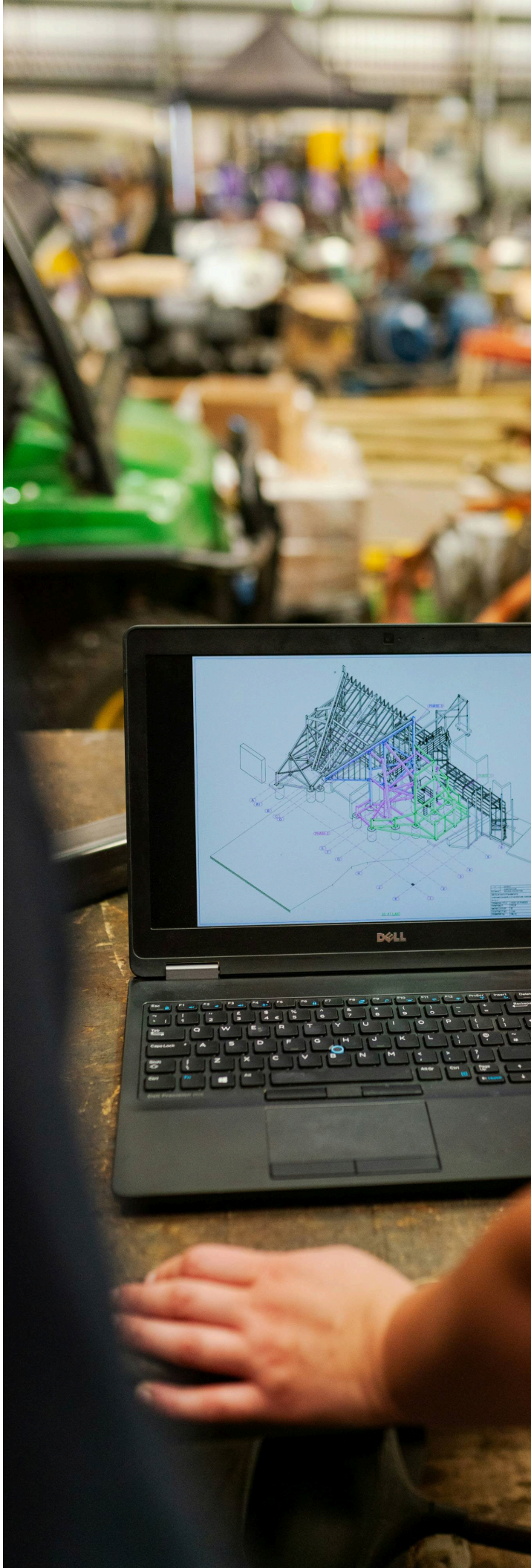
Challenge: There is ambiguity in accountability when AI-assisted decisions lead to adverse outcomes, making it unclear if the responsibility lies with the clinician or the model developer. This uncertainty creates obstacles in adoption and misalignments between stakeholders.

Recommendation: Regulatory guidelines are needed to clarify accountability for AI-assisted decisions, specifying who is responsible in adverse cases. These guidelines would provide clear liability structures, making AI model adoption more feasible by aligning the interests of stakeholders, including health systems and developers.

Adaptive Regulatory Frameworks

Challenge: The rapid pace of AI innovation requires regulatory frameworks that evolve continuously. Overly rigid regulations could stifle innovation, while insufficiently rigorous guidelines might risk patient safety.

Recommendation: A non-prescriptive regulatory approach should be adopted, enforcing rigorous standards for evaluation and deployment but allowing adaptability as technology advances and new application areas emerge. Frameworks should include diverse population evaluations, ongoing monitoring, and contingency plans for addressing data drift and technological changes, ensuring both innovation and safety.



CONCLUSION

AI presents an inflection point for society, with the potential to drive unprecedented progress across industries and communities.

By implementing The Block Center's recommendations, policymakers can harness AI's transformative power while mitigating associated risks. The Block Center is prepared to work in a bipartisan manner to ensure that the US is a leader in AI and that this technology benefits the US economy and citizenry. This balanced approach ensures technology aligns with public interest, fostering a prosperous future. Through strategic collaboration, innovation, and ethical governance, the United States can lead the global shift toward responsible technological advancement.

ABOUT

The Block Center for Technology and Society at Carnegie Mellon University is committed to fostering a future where technological advancements align with public good, economic growth, and social equity.

The Block Center is dedicated to exploring the intersections of technology, society, and public policy. The Block Center aims to foster interdisciplinary research that addresses the complex challenges posed by rapid technological advancements. Through our research and outreach, The Block Center for Technology and Society seeks to promote a thoughtful and inclusive approach to technological innovation, ultimately striving for a future where technology serves the common good.

Our Areas of Work

The Future of Work
The Block Center investigates the various ways emerging technologies are or will impact workers at all skill levels. We aim to foster research that investigates the impact of disruptive innovation on the U.S. labor market and leverages advanced technologies to address the social and economic needs of those being left behind as a result of technological change.

Responsible use of Artificial Intelligence
The Block Center brings together researchers and educators spanning computer science, engineering, decision sciences, philosophy, arts, economics, psychology, public policy, statistics, and business to make progress with stakeholders to develop and deploy best practices in AI.

Seeding Societal Futures
The Block Center is dedicated to connecting CMU faculty with diverse stakeholders to execute projects prioritizing equitable, human-centered design and use of technology.

CONTRIBUTORS

Research and Content:

Block Center Affiliated Faculty
Sarah Fox
Ramayya Krishnan
Nikolas Martelaro
Tom Mitchell
Anand Rao
Costa Samaras
Holly Wiberg
Peter Zhang

Block Center Staff & Contributors
Matthew Broocke
Cole Gessner
Sarah Giordano
Mitul Jhaveri
Jay Palat

Design:

Brittany Calderon
Lia Gold-Garfinkel

REFERENCES

Blaise, A. (2023). As ChatGPT enters the classroom, teachers weigh pros and cons. National Educators Association. Retrieved from https://www.nea.org/nea-today/all-news-articles/chatgpt-enters-classroom-teachers-weigh-pros-and-cons

Blumenthal, D., Gumas, E. D., Shah, A., Gunja, M. Z., & Williams, R. D. II. (2024, September 19). Mirror, Mirror 2024: A portrait of the failing U.S. health system. The Commonwealth Fund. https://www.commonwealthfund.org/publications/fund-reports/2024/sep/mirror-mirror-2024
Callegari, A. (2024). Why trust and safety discussions are key to AI safety. World Economic Forum. Retrieved from https://www.weforum.org/agenda/2024/08/why-trust-and-safety-discussions-are-key-to-ai-safety/

Charuvilparambil Titus, Nikhil George and Telang, Rahul and Ramayya, Krishnan, Learning from Internal Mobility: A Machine Learning Approach to Augmenting Human Capital (September 20, 2022). http://dx.doi.org/10.2139/ssrn.4230710
Charuvilparambil Titus, Nikhil George and Combemale, Christophe and Krishnan, Ramayya and Telang, Rahul, Linking Skills to Mobility: Insights for Workforce Development (June 07, 2024). http://dx.doi.org/10.2139/ssrn.4857198

Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2021). The AI gambit: Leveraging artificial intelligence to combat climate change - Opportunities, challenges, and recommendations. National Library of Medicine. Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC8522259/

Diliberti, M., Schwartz, H., Doan, S., Shapiro, A., Rainey, L., & Lake, R. (2024). Using artificial intelligence tools in K-12 classrooms. RAND Corporation. Retrieved from https://www.rand.org/pubs/research_reports/RRA956-21.html

El-Bouri, R., Taylor, T., Youssef, A., Zhu, T., & Clifton, D. A. (2021). Machine learning in patient flow: A review. Progress in Biomedical Engineering, 3(2), 022002. https://doi.org/10.1088/2516-1091/abddc5

Frase, H., & Daniels, O. (2023). Understanding AI harms: An overview. Center for Security and Emerging Technology. Retrieved from https://cset.georgetown.edu/article/understanding-ai-harms-an-overview/

Gentine, P., List, G., Thompson, K., Fan, Z., Sandalow, D., Ma, N., Kesavadas, T., ... O'Neil, G. (2024). Landscape assessment of AI for climate and nature. Columbia University, University at Albany, State University of New York, and ESRI. Retrieved from https://www.climate.columbia.edu/sites/default/files/content/research/AI%20for%20Climate%20&%20Nature%20-%20Bezos%20Earth%20Fund/Landscape%20Assessment%20of%20AI%20for%20Climate%20and%20Nature%20-%20May%202024.pdf

Green, A. (2024). Artificial intelligence and the changing demand for skills in the labour market. OECD Artificial Intelligence Papers, No. 14. https://doi.org/10.1787/88684e36-en
Haselager, P., Schraffenberger, H., Thill, S., et al. (2024). Reflection machines: Supporting effective human oversight over medical decision support systems. Cambridge Quarterly of Healthcare Ethics, 33(3), 380-389. https://doi.org/10.1017/S0963180122000718

Hendrycks, D., Mazeika, M., & Woodside, T. (2023). An overview of catastrophic AI risks. Center for AI Safety. Retrieved from http://arxiv.org/pdf/2306.12001
International Energy Agency. (2024). Electricity 2024: Analysis and forecast to 2026. Retrieved from
https://iea.blob.core.windows.net/assets/18f3ed24-4b26-4c83-a3d2-8a1be51c8cc8/Electricity2024-Analysisandforecastto2026.pdf

Jones, K., Buscher, M., & Wallace, J. (2023). Challenges of AI. Chatham House. Retrieved from https://www.chathamhouse.org/2022/03/challenges-ai

Khattar, R. (2023). Will AI benefit or harm workers? Center for American Progress. Retrieved from https://www.americanprogress.org/article/will-ai-benefit-or-harm-workers/

Koedinger, K. Navigating Ethical Benefits and Risks as AIED Comes of Age. Int J Artif Intell Educ 34, 136–143 (2024). https://doi.org/10.1007/s40593-023-00350-5

Lin, L. (2024). A quarter of U.S. teachers say AI tools do more harm than good in K-12 education. Pew Research Center. Retrieved from
https://www.pewresearch.org/short-reads/2024/05/15/a-quarter-of-u-s-teachers-say-ai-tools-do-more-harm-than-good-in-k-12-education/

Lyakhova, U. A., & Lyakhov, P. A. (2024). Systematic review of approaches to detection and classification of skin cancer using artificial intelligence: Development and prospects. Computers in Biology and Medicine, 178, 108742.
https://doi.org/10.1016/j.compbiomed.2024.108742
National Academies of Sciences, Engineering, and Medicine. (2024). Artificial intelligence and the future of work. Washington, DC: The National Academies Press.
https://doi.org/10.17226/27644

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6652), 187-192. https://doi.org/10.1126/science.adh2586

Oreopoulos, Philip and Gibbs, Chloe and Jensen, Michael and Price, Joseph, Teaching Teachers to Use Computer Assisted Learning Effectively: Experimental and Quasi-Experimental Evidence (April 2024). NBER Working Paper No. w32388, https://ssrn.com/abstract=4810603

Ospina, J. (2024). How to build the skills needed for the age of AI. World Economic Forum. Retrieved from https://www.weforum.org/agenda/2024/04/skill-building-age-of-ai/

Partelow, L. (2024). Using learning science to analyze risks and benefits of AI in K-12 education. Center for American Progress. Retrieved from
https://www.americanprogress.org/article/using-learning-science-to-analyze-the-risks-and-benefits-of-ai-in-k-12-education/

Ronanki, R. (2024, July 2). Ethical AI in healthcare: A focus on responsibility, trust, and safety. Forbes. Retrieved from https://www.forbes.com/sites/forbesbooksauthors/2024/01/04/ethical-ai-in-healthcare-a-focus-on-responsibility-trust-and-safety/

Rony, M. K. K., Parvin, M. R., Wahiduzzaman, M., Debnath, M., Bala, S. D., & Kayesh, I. (2024). "I wonder if my years of training and expertise will be devalued by machines": Concerns about the replacement of medical professionals by artificial intelligence. SAGE Open Nursing, 10, 23779608241245220. https://doi.org/10.1177/23779608241245220

Snider, V., Homsi, K., Kusnoto, B., Atsawasuwan, P., Viana, G., Allareddy, V., ... Elnagar, M. H. (2024). Clinical evaluation of artificial intelligence-driven remote monitoring technology for assessment of patient oral hygiene during orthodontic treatment. American Journal of Orthodontics and Dentofacial Orthopedics, 165(5), 586-592. https://doi.org/10.1016/j.ajodo.2023.12.008

Tamayo, J., Doumi, L., Goel, S., Kovacs-Ondrejkovic, O., & Sadun, R. (2023). Reskilling in the age of AI. Harvard Business Review. Retrieved from https://hbr.org/2023/09/reskilling-in-the-age-of-ai

Toner-Rodgers, A. (2024). Artificial intelligence, scientific discovery, and product innovation. arXiv. https://arxiv.org/abs/2412.17866

The Economist. (2024). Generative AI has a clean energy problem. The Economist. Retrieved from https://www.economist.com/business/2024/04/11/generative-ai-has-a-clean-energy-problem

U.S. Department of Health and Human Services Office of Minority Health. (2023, July 12). Shedding light on healthcare algorithmic and artificial intelligence bias. Retrieved from https://minorityhealth.hhs.gov/news/shedding-light-healthcare-algorithmic-and-artificial-intelligence-bias