Cognitive Insights for Artificial Intelligence

Request for Comments by the Office of Science and Technology Policy (OSTP), White House on developing a National Artificial Intelligence (AI) Strategy that will chart a path for the United States to harness the benefits and mitigate the risks of AI. Document Number 2023-11346

Submitted by Monica Lopez, PhD and Irene Gonzalez, PhD.
Organization, Cognitive Insights for Artificial Intelligence (CIfAI)
July 6, 2023

On behalf of CIfAI, we write in response to the Biden-Harris Administration developing a National AI Strategy that will chart a path for the United States to harness the benefits and mitigate the risks of AI. As this strategy will build on the actions that the Federal Government has already taken to responsibly advance the development and use of AI, we underscore the importance of these efforts given that AI plays not only a fundamental role in the development of the U.S.'s well-known innovation ecosystem, but that AI brings risks to individuals and society at large that necessitate attention. We support OSTP efforts in seeking stakeholder input to help update U.S. national priorities and future actions on AI.

We at CIfAI provide strategic research-based solutions from a human-centered perspective to ensure the safe and ethical design, development, deployment, and management of AI-enabled autonomous systems across various industries. Our values-based approach is founded on accuracy, consistency, and context-dependency, and supports trusted data across every phase of the AI lifecycle to achieve confident and fair decision making.

CIfAI has reviewed all six topics and respective 29 questions provided in the OSTP's request for comment and provides recommendations (total 23) to all eight questions under the topic of **Protecting Rights, Safety, and National Security** below.

Preamble

AI technologies pose multiple risks, and they necessitate adequate guardrails in place. We acknowledge the position of the U.S. regarding the development and use of AI technologies across the nation to forge a

balance between supporting innovation and promoting safety.[1] We further recognize that the private sector is the engine of AI innovation and therefore advocate for the active involvement of industry in the creation of laws, statutes and regulations in partnership with the U.S. government. Collaboration at this critical moment in the advancement of AI is vital to determining the legislation to prepare in the U.S. because industry has technological expertise and government regulatory expertise. Moreover, the U.S. government has the potential to guarantee that AI creates greater opportunities, providing economic and societal benefits for the nation's population.[2] As the U.S. seeks to maintain its competitive edge in the global AI landscape, ensuring the responsible development and deployment of AI-enabled systems becomes an imperative.

Responses to Specific Questions

QUESTION 1. *What specific measures – such as standards, regulations, investments, and improved trust and safety practices – are needed to ensure that AI systems are designed, developed, and deployed in a manner that protects people's rights and safety? Which specific entities should develop and implement these measures*?

Given the complexity of the data collected and selected to train AI algorithms and the processes carried out by AI-enabled systems to generate predictions and make decisions, there is no single standard, regulation, investment, best practice and/or entity that can cover the range of the AI safety problem. In fact, the problem of 'safety' first demands a reexamination of the current environment in which the use of AI-enabled systems has entered everyday life: what kind of relationship do we want between us humans and the AI technology we are developing, and how can we ensure that it is healthy and beneficial for all? To highlight most recent examples, generative AI technologies like ChatGPT,[3] DALL-E,[4] and LLaMA[5] have been released to the public and do not require technological knowledge to use. The coupling of the technologies' user-friendly interfaces and human nature have collided, supporting the engendering of creativity both positive and negative (e.g., thousands of fake photos[6] and videos[7] have been posted to the internet; misleading text generated via human prompts has resulted in fraud and disinformation[8]).

---

[1] Schumer, C. June 21, 2023. Sen. Chuck Schumer Launches SAFE Innovation in the AI Age at CSIS. Transcript, CSIS. https://www.csis.org/analysis/sen-chuck-schumer-launches-safe-innovation-ai-age-csis.

[2] NAIAC Report, Year 1. May 2023. National Artificial Intelligence Advisory Committee (NAIAC). https://www.ai.gov/wp-content/uploads/2023/05/NAIAC-Report-Year1.pdf.

[3] OpenAI. 2023. Introducing ChatGPT. https://openai.com/blog/chatgpt.

[4] NPR Staff. September 20, 2022. Dall-E is now available to all. NPR put it to work. https://www.npr.org/2022/09/30/1125976146/dall-e-art-ai-generator-npr.

[5] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... and Lample, G. (2023). LLaMA: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

[6] Devlin, K. and Cheetham, J. March 24, 2023. Fake Trump arrest photos: How to spot an AI-generated image. BBC News. https://www.bbc.com/news/world-us-canada-65069316.

[7] Denham, H. August 3, 2020. Another fake video of Pelosi goes viral on Facebook. The Washington Post. https://www.washingtonpost.com/technology/2020/08/03/nancy-pelosi-fake-video-facebook/.

[8] Buchanan, B. et al. May 2023. Truth, Lies, and Automation. How language models could change disinformation. Center for Security and Emerging Technology (CSET), Georgetown University. https://cset.georgetown.edu/wp-content/uploads/CSET-Truth-Lies-and-Automation.pdf

Moreover, misguided media coverage[9] and scientifically unproven claims that AI is sentient, conscious and at human-level intelligence[10] raise challenges to policy makers determining how to best regulate AI technologies. The question now becomes: What type of requirement or qualification or set thereof do AI-enabled systems need to have in place such that AI technologies can be built in a safe manner and accountability can be taken for any consequences of misuse and abuse? To address this challenge, government agencies at the state level have already enacted local laws.[11] For example, acts enacted in New York,[12] Illinois,[13] and California[14] impose requirements for businesses utilizing AI systems that directly impact individuals. In response, AI companies have also developed risk management tools[15],[16] like auditing platforms to mitigate the negative effects of high-risk AI systems. Companies have also introduced guardrails to anticipate and prevent the misuse and abuse of their technology.[17] Technical standards as developed by the International Organization for Standardization (ISO)[18] and the Institute of Electrical and Electronics Engineers (IEEE)[19] are already playing an important role in ensuring the development of trustworthy AI.

Recommendation #1: In accordance with the interest to operationalize the NIST AI RMF across public and private sectors here in the U.S. and internationalize it to support global regulatory cooperation on AI,[20] we advise that any governance framework for AI developed dictate the need for the integration of guardrails from the start. Such governance framework will necessarily be developed by industry due to

---

[9] Bender, E. M. April 17, 2022. On NYT Magazine on AI: Resist the urge to be impressed. Medium. https://medium.com/@emilymenonbender/on-nyt-magazine-on-ai-resist-the-urge-to-be-impressed-3d92fd9a0edd.

[10] Metz, C. May 16, 2023. Microsoft Says New AI Shows Signs of Human Reasoning. The New York Times. https://www.nytimes.com/2023/05/16/technology/microsoft-ai-human-reasoning.html.

[11] Ng, A. February 22, 2023. The raucous battle over Americans' online privacy is landing on states. POLITICO. https://www.politico.com/news/2023/02/22/statehouses-privacy-law-cybersecurity-00083775.

[12] The New York City Council. Automated Employment Decision Tools. https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=Advanced&Search.

[13] Riley, T. 2023. Illinois' Biometric privacy law provides blueprint as states seek to curb data collection. Cyberscoop. https://cyberscoop.com/states-bipa-biometric-privacy-legislation-illinois/.

[14] California Consumer Privacy Act. 2023. State of California, Department of Justice. https://oag.ca.gov/privacy/ccpa.

[15] Kazim, E., Koshiyama, A. S., Hilliard, A., & Polle, R. (2021). Systematizing audit in algorithmic recruitment. *Journal of Intelligence*, *9*(3), 46.

[16] Koshiyama, A., Kazim, E., & Treleaven, P. (2022). Algorithm auditing: Managing the legal, ethical, and technological risks of artificial intelligence, machine learning, and associated algorithms. *Computer*, *55*(4), 40-50.

[17] OpenAI. April 5, 2023. Our approach to AI safety. OpenAI Blog. https://openai.com/blog/our-approach-to-ai-safety.

[18] ISO. May 2020. ISO/IEC TR 24028:2020 Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence. https://www.iso.org/standard/77608.html.

[19] IEEE. May 11, 2022. How To Make Autonomous Systems More Transparent and Trustworthy. https://standards.ieee.org/beyond-standards/topic/artificial-intelligence-systems/how-to-make-autonomous-systems-more-transparent-and-trustworthy/.

[20] National Artificial Intelligence Advisory Committee (NAIAC). May 2023. Year 1 Report. https://www.ai.gov/wp-content/uploads/2023/05/NAIAC-Report-Year1.pdf.

technological know-how. At the same time, it is advised that regulations should be created and enacted by the federal and local governments to ensure development of better consumer protections.

Recommendation #2: Attention should be put into convening a group of diverse stakeholders, particularly outside of the group of big technology leaders. This group could include those from small and medium-sized enterprises that have a direct frontline product-to-consumer relationship with their customers and deal with the day-to-day of customer understanding and experience.

Recommendation #3: The above recommendations would benefit from a robust private-public partnership between industry and government to support the development of scientific innovation in tandem with the development of relevant and appropriate laws.

QUESTION 2. *How can the principles and practices for identifying and mitigating risks from AI, as outlined in the Blueprint for an AI Bill of Rights and the AI Risk Management Framework, be leveraged most effectively to tackle harms posed by the development and use of specific types of AI systems, such as large language models?*

The Blueprint for an AI Bill of Rights[21] is essentially a handbook or plan of action created to serve and benefit the American people and protect U.S. democracy and national security from threats generated by unsafe AI tools. The handbook's foundations are based on five principles (i.e., safe and effective systems; algorithmic discrimination protections; data privacy; notice and explanation; human alternatives, consideration, and fallback) that address the design, development, and deployment of AI technologies. While this is a considerable step forward towards the development of regulatory frameworks for AI, the blueprint remains a high-level set of principles without regulation being enacted and implemented. Furthermore, it leaves open the need to determine clear goals and measures of success. In a general sense, AI technology is difficult to regulate. It is a technology advancing at a rapid pace and its level of complexity poses even greater R&D challenges. Of note, there are initiatives and frameworks that have been developed to achieve the integration of AI advances across federal agencies. For example, the National Artificial Intelligence Initiative, that oversees the U.S. national AI strategy, launched the National AI Initiative Act of 2020 that became law on January 1, 2021.[22] Although no regulation is mentioned, this law coordinates all federal agencies to ensure the integration of AI systems for the acceleration of AI research and applications across all sectors of the economy and society. Additionally, the National Institute of Standards (NIST), with the help of many industry experts and individuals, developed an AI Risk Management Framework[23] to support responsible AI. The NIST's framework is voluntary for companies developing AI systems and essentially constitutes 'soft law'. This does not mean, of course, that there are no legal remedies available. Various problematic outcomes can be addressed with existing laws. For example, racially discriminatory loan decisions are covered by the Fair Housing Act,[24]

---

[21] OSTP. Blueprint for an AI Bill of Rights. Making automated systems work for the American people. White House. https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

[22] U.S. Government. 2020. National Artificial Intelligence Initiative. Overseeing the US national AI strategy. https://www.ai.gov/.

[23] U.S. National Institute of Standards and Technology (NIST), Information Technology Laboratory. 2023. AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework.

[24] U.S. Department of Housing and Urban Development. 1968. Housing Discrimination Under the Fair Housing Act. https://www.hud.gov/program_offices/fair_housing_equal_opp/fair_housing_act_overview.

discrimination against any protected class is covered by The Equal Employment Opportunity Act,[25] and a discriminating bias algorithm is covered by the Federal Trade Commission's Algorithmic Accountability Act of 2019.[26]

Recommendation #4: Prior initiatives like the Universal Guidelines for AI (UGAI)[27] and the first human rights framework for AI, stand as a good model for an AI governance framework. The UGAI set out basic rights and obligations for the use of AI that could serve as a helpful guide for the further development of the blueprint. To highlight, the UGAI included transparency, human determination, fairness, accountability, and data quality principles, among others. Moreover, obligations included standards for accuracy, security and public safety, and prohibitions on practices such as secret profiling and social credit scores to prevent government surveillance and discrimination.

Recommendation #5: On the question of large language models (LLMs), a thorough risk assessment of the specific LLM is needed to determine the risks and harms that might be embedded in the trained algorithm. The assessment should consider potential negative consequences at the ethical, individual, social, and legal levels. It is therefore advisable that LLM developers disclose their data sources, and selection thereof, and the training methods and evaluation procedures carried out. This would ensure transparency and explainability of the LLM in terms of its capabilities and limitations, including its potential for bias. Since LLMs use data created and curated by humans, developers should implement measures to protect the privacy and personal data of individuals and develop guidelines that allow for issuance of consent to utilize such data and its subsequent secure handling. This also becomes important to understand how risks can emerge because of the utilization of a single 'foundation' model for a wide range of uses, leading to new questions about allocation of responsibility in the advent of harmful outcomes.

Recommendation #6: Since LLMs contain biases, regular audits of the model should be conducted to determine the source of bias in the training data set as well as the biased outputs themselves. Moreover, many of the impressive capabilities of LLMs are the consequence of Reinforcement Learning from Human Feedback (RLHF) whereby the language model is directly optimized with human feedback.[28] While human feedback for generated text is used both to measure rate of performance and to align a model trained on a general corpus of text data to that of human values, we underscore that what constitutes "good" text and agreed upon human values are open questions. Auditing the LLM would help to ensure that data are diverse and that the criteria utilized for human feedback align with the diverse representation of language itself as it is utilized by different people from different backgrounds, regions and cultures; this would help to promote non-discrimination and fairness of output and values. Audits should be done by independent parties and further evaluated by review boards who have no conflict of interest whatsoever in the system (and company/organization developing and/or using the system) they are auditing. Independent audits and review boards are necessary to ensure accountability as well as to incentivize continuous monitoring and evaluation.

---

[25] U.S. Department of Labor. 1972. Equal Employment Opportunity Act. https://www.dol.gov/general/topic/discrimination.

[26] U.S. Congress. H.R.2231-116th Congress - Algorithmic Accountability Act of 2019. https://www.congress.gov/bill/116th-congress/house-bill/2231/text.

[27] Universal Guidelines for AI. October 2018. The Public Voice. https://thepublicvoice.org/ai-universal-guidelines/memo/.

[28] Lambert, N., Castricato, L., von Werra, L., and Havrilla, A. December 9, 2022. Illustrating Reinforcement Learning from Human Feedback (RLHF). Hugging Face Blog. https://huggingface.co/blog/rlhf.

Recommendation #7: The traditional approach to legislation and law enacting is 'hard law'. Unlike 'soft law', legal instruments create direct enforceable expectations with consequences. While 'soft law' lends itself to rapid adoption due to its adaptability to new situations, like in the case of new AI applications, enforcement is harder to achieve.[29] However, soft law programs do offer an alternative in the absence of hard law. More incentives should be offered to promote the value of soft law initiatives. When guided by empirical evidence and inspired by multiple regulatory perspectives, soft law approaches that, for example, propose standards play a key role in defining technical solutions and therefore can be powerful tools to engender collective support and international cooperation.

Recommendation #8: Despite LLMs being capable of inflicting harm, calls for specific regulation should be considered with caution to avoid the creation and inaction of laws that end up providing substantially less protection. For example, in the Stored Communications Act/Title II of the Electronic Communications Privacy Act of 1986[30] for digital communications (e.g., email), it initially provided less privacy protection due to limitations in email storage. Despite technological advances in communications, the law remains to be updated. Moreover, laws enacted but not enforced can bring unintended consequences. Therefore, it is of interest to consider a broad yet context-specific look into the potential effects that proposed regulations may have.

QUESTION 3. *Are there forms of voluntary or mandatory oversight of AI systems that would help mitigate risk? Can inspiration be drawn from analogous or instructive models of risk management in other sectors, such as laws and policies that promote oversight through registration, incentives, certification, or licensing*?

Various forms of voluntary and mandatory oversight of AI systems that can help mitigate risks associated with their deployment are already in place. They have been crafted from existing risk management models in other sectors. Oversight should include certification and standards by independent regulatory bodies or organizations. AI-enabled systems could undergo a certification process to ensure users and the public at large that they comply with specific safety, security, and ethical standards. The overall goal is to obtain a successful certification process to guarantee standards and at the same time avoid overregulation and enable innovation. In the U.S. there are no comprehensive certification programs tailored to AI systems. However, government agencies such as NIST have been working on developing voluntary standards and guidelines for AI ethics, transparency, and accountability.[31] The IEEE has designed an Ethically Aligned Initiative providing guidelines for the ethical development and deployment of AI systems.[32] At the international level, for example, Germany has proposed a certification of AI systems via a white paper.[33]

---

[29] Gutierrez, C. I. and Marchant, G. May 27, 2021. How soft law is used in AI governance. https://www.brookings.edu/articles/how-soft-law-is-used-in-ai-governance/.

[30] 18 US Code 2703-Required Disclosure of Customer Communications or Records. Cornell Law School. https://www.law.cornell.edu/uscode/text/18/2703.

[31] U.S. National Institute of Standards and Technology (NIST), Information Technology Laboratory. 2023. AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework.

[32] Institute of Electrical and Electronics Engineers. Ethically Aligned Design. First Edition. https://www.businesswire.com/news/home/20190325005314/en/IEEE-Launches-Ethically-Aligned-Design-First-Edition-Delivering-A-Vision-for-Prioritizing-Human-Well-being-with-Autonomous-and-Intelligent-Systems.

[33] Certification of AI Systems. Lernende Systems. Germany's Platform for Artificial Intelligence. https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen_EN/AG1_3_WP_Executive_Summary_Certification_AIsystems.pdf.

Additionally, France has built a Certification of Process for AI.[34] And in the European Union, European standardizers have already prepared a roadmap in response to the upcoming standardization request by the European Commission for the AI Act.[35]

Recommendation #9: Registration and disclosure of information of AI-enabled systems, including their data sources, intended use, and potential risks should be mandatory to provide regulators with the right knowledge of the AI system to assess potential negative risks. In the U.S. there is no dedicated system for AI registration and disclosure. However, some federal agencies have certain regulations and laws that apply to this activity. The U.S. Patent and Trademark Office (USPTO)'s AI Patent Dataset contains AI-related invention patents.[36] The California Consumer Privacy Act (CCPA) and the Health Insurance Portability and Accountability Act (HIPAA) are laws that require organizations to disclose their data collection and uses to ensure the protection of individuals' personal information.[37],[38] The Federal Trade Commission (FTC) regulates and takes action if AI systems engage in unfair or deceptive practices that harm consumers implemented in the Algorithmic Accountability Act.[39]

Recommendation #10: Another way to mitigate AI risks is through licensing and accreditation. In this context, AI companies and their developers demonstrate competence and adherence to guidelines, laws, statutes or regulations. In the U.S. there are no federal comprehensive systems for licensing and accreditation for AI systems. However, some professional organizations such as IEEE offer certifications and accreditations related to AI engineering and ethical practices, and the U.S. Food and Drug Administration (FDA) issued a guidance document for AI/ML Based Software as a Medical Device (SaMD) to meet certain regulatory requirements for approval or clearance.[40] Open and Responsible AI licenses (i.e., OpenRAIL, RAIL) are AI-specific licenses that enable open access, use and distribution of AI artifacts while requiring responsible use of the latter.[41] OpenRAILs and RAILs stand as first steps towards enabling ethics-informed behavioral restrictions with the goal to open communication between stakeholders and provide concrete directives how the licensed artifact can be used.[42]

---

[34] Certification of Processes for AI. Laboratoire National De Métrologie et D'Essais. Republique Française. https://www.lne.fr/en/service/certification/certification-processes-ai.

[35] Soler Garrido, J., Fano Yela, D., Panigutti, C., Junklewitz, H., Hamon, R., Evas, T., André, A. and Scalzo, S., Analysis of the preliminary AI standardisation work plan in support of the AI Act, EUR 31518 EN, Publications Office of the European Union, Luxembourg, 2023, ISBN 978-92-68-03924-3, doi:10.2760/5847, JRC132833.

[36] U.S. Patent and Trademark Office (USPTO). Artificial Intelligence Patent Dataset. https://www.uspto.gov/ip-policy/economic-research/research-datasets/artificial-intelligence-patent-dataset.

[37] California Consumer Privacy Act. 2023. State of California, Department of Justice. https://oag.ca.gov/privacy/ccpa.

[38] U.S. Centers for Disease Control and Prevention (CDC). Health Insurance Portability and Accountability Act of 1996 (HIPAA). https://www.cdc.gov/phlp/publications/topic/hipaa.html.

[39] U.S. Congress. H.R.2231-116th Congress - Algorithmic Accountability Act of 2019. https://www.congress.gov/bill/116th-congress/house-bill/2231/text.

[40] U.S. Food and Drug Administration (FDA). 2021. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. https://www.fda.gov/media/145022/download.

[41] Responsible AI Licenses. https://www.licenses.ai/.

[42] Ferrandis, C. M. August 31, 2022. OpenRAIL: Towards open and responsible AI licensing frameworks. Hugging Face Blog. https://huggingface.co/blog/open_rail.

Recommendation #11: Auditing and impact assessments of AI systems should also be performed to identify biases, risks, and unintended consequences in AI algorithms. These should be conducted by independent bodies to ensure transparency and accountability. In the U.S. there are no systems for auditing at the federal level. However, it would be advisable that companies developing AI systems carry out internal auditing using available risk management systems developed by the private sector.[43],[44],[45] Audits by third-party organizations are also advisable to obtain credible independent evaluations of the AI systems. Companies could also voluntarily adhere to available ethical AI principles[46] and risk management frameworks.[47] Other oversight mechanisms may include public-private collaborations via task forces, guidelines' development and joint initiatives. Another approach would be to provide incentives for organizations to adopt best practices for the development of responsible and ethical AI systems. Incentives could be in the form of grants, tax benefits, or preferential treatment in response to evidence of good faith behavior. Because AI technologies rapidly evolve and change, it would be advisable to also create a mechanism of continuous monitoring of AI systems to update the oversight mechanisms and guidelines.

QUESTION 4. *What are the national security benefits associated with AI? What can be done to maximize those benefits*?

AI offers many security benefits, with the potential to directly affect the balance of power in both military competition and across the global economy. To maximize its benefits, however, we must acknowledge that AI still remains at a fairly early stage and is highly problem-specific and context-dependent. Successful adoption of AI will necessitate investment to educate the technical and management talent and investments for the technical infrastructure (data and its computing and networking capabilities). Another fundamental requirement to highlight again is the need for high-quality data. No matter the use case for AI-enabled intervention, raw data that is accurate, timely and reliable is vital to processing and subsequently transforming the data points into actionable insights.[48]

Recommendation #12: Ways to maximize security benefits related to AI may include enhancing threat detection and intelligence analysis using large datasets to identify in real-time patterns, anomalies, and potential risks. Cybersecurity and defense could be bolstered by AI to detect and respond to cyber threats via the analysis of network traffic patterns. AI can also enhance surveillance capabilities, enabling more efficient and accurate intelligence gathering and actionable intelligence by using automated video analytics, facial recognition, and natural language processing, among others. This could be coupled with decision-making processes by providing data-driven insights and risk assessments. AI could also be used to enable the development of autonomous systems and robotics for national security in applications like

---

[43] Holistic AI. AI Governance, Risk and Compliance. https://www.holisticai.com/.

[44] Complysci. Compliance Program Management Technology. https://www.complysci.com/.

[45] Monetary Authority of Singapore. MAS-led Industry Consortium Releases Toolkit for Responsible Use of AI in the Financial Sector. https://www.mas.gov.sg/news/media-releases/2023/toolkit-for-responsible-use-of-ai-in-the-financial-sector.

[46] Institute of Electrical and Electronics Engineers (IEEE). Standards. Piscataway, NJ, USA. https://www.ieee.org/.

[47] U.S. National Institute of Standards and Technology (NIST), Information Technology Laboratory. 2023. AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework.

[48] UNESCO. 2023. Open data for AI: What now? United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000385841.

uncrewed vehicles, drones, and autonomous robots that could be deployed for surveillance, reconnaissance, and military operations, reducing risks to human personnel.

Recommendation #13: To maximize AI benefits, it would be necessary to coordinate agencies across the federal government and within the Executive Office of the President to address how AI technology intersects with civil rights and equity, the economy, and national security.[49] The Department of State Bureau of Cyberspace and Digital Policy (CDP) comprises the International Cyberspace Security to address responsible behavior in cyberspace. The agency also advances policies that protect the integrity and security of the infrastructure of the Internet, and challenges associated with AI like national security. This will require experts with knowledge in integrating AI across national security programs.[50]

Recommendation #14: Regarding national security, the U.S. Department of Defense (DOD) has invested billions of dollars to develop and integrate AI into defense systems. AI capabilities supporting DOD's war fighting mission are still in development and include facial recognition to enhance weapon systems (e.g., drones, robotic ships) or providing recommendations on the battlefield (e.g., targeted missile strikes).[51]

QUESTION 5. *How can AI, including large language models, be used to generate and maintain more secure software and hardware, including software code incorporating best practices in design, coding and post deployment vulnerabilities*?

AI can contribute to enhancing security in software and hardware in multiple ways that may include secure code generation that provides automated suggestions, code completion, and bug detection. This could help analyze existing code bases, identify vulnerabilities, and reduce human errors and vulnerabilities in software. Vulnerability detection could be identified in software and hardware systems based on code analysis, system behavior, data flow, detection of security weaknesses, among others. This can help improve the security of the system by helping developers to identify and address vulnerabilities early in the development process. AI could also be used in monitoring and analyzing large amounts of data to provide threat intelligence and risk assessment to help proactively patch vulnerabilities and address security risks. Other ways AI could be used are in security and penetration testing, anomaly and intrusion detection, and in continuous monitoring and securing of software and hardware systems throughout its life cycle after deployment.

Recommendation #15: To generate and maintain more secure software and hardware, AI could be used in high-quality training data, including historical security data, vulnerability databases, and code repositories that incorporate best practices and known vulnerabilities. AI could help to adhere to ethical standards and avoid biases and unintended consequences in AI-generated code or security assessments to ensure privacy, fairness, and transparency. It is also crucial to combine domain expertise and AI capabilities effectively. This could be achieved via knowledge sharing and collaboration between AI security experts, developers and researchers, and different stakeholders. This can lead to better security

---

[49] NSCAI: Final Report, Chapter 9; May 2021 Amendment to the U.S. Innovation and Competition Act (USICA) filed by Senators Michael Bennet and Ben Sasse (the amendment was not adopted); June 2022 House Resolution 8027 introduced in the 117th Congress by Representatives Bacon, Franklin, Carbajal, and Lamb (the resolution was not adopted); and November 2022 Platforms Interim Panel Report of the Special Competitive Studies Project.

[50] NAIAC Report, Year 1. May 2023. National Artificial Intelligence Advisory Committee (NAIAC). https://www.ai.gov/wp-content/uploads/2023/05/NAIAC-Report-Year1.pdf.

[51] U.S. Government Accountability Office (GAO). April 2022. How Artificial Intelligence Is Transforming national Security. https://www.gao.gov/blog/how-artificial-intelligence-transforming-national-security#:~:text=.

outcomes and more robust and resilient systems. As an example, NVIDIA has developed secure LLMs to enhance software and hardware security.[52],[53]

QUESTION 6. *How can AI rapidly identify cyber vulnerabilities in existing critical infrastructure systems and accelerate addressing them*?

AI could be used to address cyber vulnerabilities in critical infrastructure by using a variety of methods and approaches that may include leveraging machine learning techniques via algorithm code, configuration and network traffic analysis to help identify vulnerabilities across multiple systems rapidly.[54] AI-enabled systems can continuously monitor critical infrastructure systems for signs of cyber threats and attacks, enabling timely response and remediation of vulnerabilities in real-time. Systems could also analyze historical data and patterns to predict potential vulnerabilities in critical infrastructure systems and prioritize vulnerability mitigation efforts. AI could be used to automate the process of conducting security assessments on critical infrastructure systems, and to identify vulnerabilities rapidly across a large number of systems, reducing the time and effort required for manual assessments. AI can also help integrate with external threat intelligence sources, such as cybersecurity databases, vulnerability feeds, and industry-specific information sharing platforms. By processing and analyzing this data, AI can identify vulnerabilities specific to critical infrastructure systems and provide actionable insights to security teams.

Recommendation #16: The above integration enhances the speed and accuracy of vulnerability remediation and identification. It can also facilitate collaboration and knowledge sharing among organizations operating critical infrastructure systems. The Department of Defense (DoD) is already using many AI applications in cyber defense missions.[55]

Recommendation #17: Collaboration efforts can also accelerate the identification and resolution of vulnerabilities across multiple systems. AI models that link vulnerabilities with cyber-attacks have been developed.[56]

Recommendation #18: Considerations essential to identify and address cyber vulnerabilities' detection in critical infrastructure systems[57] may include access to anonymized comprehensive and diverse data sources, including historical vulnerability data, threat intelligence, and system logs which are crucial for training AI models. Although AI can automate vulnerability detection, human expertise is critical to validate and interpret the findings and thus review and prioritize vulnerabilities, assess their potential

---

[52] NVIDIA. Generative AI For Enterprises. AI and Data Science. https://www.nvidia.com/en-us/ai-data-science/generative-ai/.

[53] Chockalingam, A and Varshney, T. 2023. Secure Large Language Model Conversational Systems. NVIDA Developer. https://developer.nvidia.com/blog/nvidia-enables-trustworthy-safe-and-secure-large-language-model-conversational-systems/.

[54] A comprehensive view of the impact and implications of AI in cyber security can be found here: Montasari, R., & Jahankhani, H. (Eds.). (2021). *Artificial Intelligence in Cyber Security: Impact and Implications: Security Challenges, Technical and Ethical Issues, Forensic Investigative Challenges*. Springer Nature.

[55] Microsoft Corporate Blog. 2022. Applications for artificial intelligence in Department of Defense cyber missions. https://blogs.microsoft.com/on-the-issues/2022/05/03/artificial-intelligence-department-of-defense-cyber-missions/.

[56] Balaji, N. 2023. Exclusive! Scientists Developed an AI Model that Automatically Links Vulnerabilities with Cyber Attacks. Cyber Security News. https://cybersecuritynews.com/ai-model-automatically-links-vulnerabilities-with-cyber-attacks/#:~:text=.

[57] Cybersecurity Infrastructure Security Agency (CISA). America's Cyber Defense Agency. Critical Infrastructure Assessments. https://www.cisa.gov/critical-infrastructure-assessments.

impact, and make informed decisions regarding remediation efforts. This also requires that AI models be continually updated and trained to adapt to evolving cyber threats and new vulnerabilities in critical infrastructure systems. This necessitates that AI also be developed and deployed with robust security measures to prevent it from becoming a potential attack vector.[58]

QUESTION 7. *What are the national security risks associated with AI? What can be done to mitigate these risks*?

While AI offers numerous national security benefits, there are also risks and challenges that need to be addressed. Some of the key national security risks associated with AI and suggestions for mitigating them are outlined as follows:

Recommendation #19: AI systems can be vulnerable to adversarial attacks, where malicious actors manipulate or deceive AI algorithms to produce incorrect or biased results resulting in serious implications for national security applications (e.g., tampering with AI-driven decision-making systems, evading AI-powered security measures).[59] Thus, robust security measures are necessary to detect and defend against adversarial attacks. This includes developing AI models resilient to adversarial manipulation, including implementing rigorous testing, validation procedures, and continuous monitoring. Since AI depends on vast amounts of data, including personal and sensitive information, lack of data privacy and protection measures can lead to unauthorized access, misuse, or breaches of sensitive data posing risks to national security given that classified information and/or personally identifiable information may be compromised.

Recommendation #20: Data protection and privacy laws that govern the collection, storage, and use of data in AI systems should be strictly enforced. Several laws at the federal and state level have already been enacted to address data privacy. Furthermore, robust encryption, access controls, and secure data handling practices should be implemented to safeguard sensitive information while prioritizing transparency and informing users about data collection and usage to build trust by the organizations using the data. This includes addressing the ethical issues in algorithms trained on biased data to prevent ethical and societal concerns. Other issues include efforts to mitigate biases in AI systems via audits and bias testing to ensure the transparency of AI development practices. The explainability and interpretability of AI systems should be addressed to enhance control over AI systems' behaviors. Several government agencies and organizations have already addressed and implemented ethical principles,[60] established risk management frameworks,[61] or developed standards for AI use.[62]

---

[58] Browne, D. and Munger, M. June 27, 2023. Securing the AI Pipeline. Mandiant. https://www.mandiant.com/resources/blog/securing-ai-pipeline.

[59] Townsend, M. 2023. AI Poses National Security Threat, Warns Terror Watchdog. The Guardian. https://www.theguardian.com/technology/2023/jun/04/ai-poses-national-security-threat-warns-terror-watchdog.

[60] Deputy Secretary of Defense. May 2021. Implementing Responsible AI in the Department of Defense. https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF.

[61] U.S. National Institute of Standards and Technology (NIST), Information Technology Laboratory. 2023. AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework.

[62] Institute of Electrical and Electronics Engineers (IEEE). Standards. Piscataway, NJ, USA. https://www.ieee.org/.

Recommendation #21: Because national security is impacted by geopolitical competition, AI has become a focal point of global competition,[63] including concerns about strategic risks associated with AI development and deployment (e.g., military applications, surveillance technologies, potential for AI arms races). Therefore, international cooperation and multilateral agreements with different nations should address geopolitical risks associated with AI like escalating tensions or unintended conflicts. These issues require a multidimensional approach that involves technological advancements, legal and regulatory frameworks, ethical considerations, and international collaboration. The goal is to develop and deploy AI in a manner that upholds security, privacy, fairness, and agreed upon human values.

QUESTION 8. *How does AI affect the United States' commitment to cut greenhouse gases by 50-52% by 2030, and the Administration's objective of net-zero greenhouse gas emissions no later than 2050? How does it affect other aspects of environmental quality*?

AI can optimize energy consumption and resource allocation in various sectors, including manufacturing, transportation, and construction. By identifying patterns and optimizing energy use, AI can help achieve emission reduction targets. AI can facilitate the integration of renewable energy sources into the grid to enhance the efficiency and reliability of renewable energy integration and thus make it easier to transition to cleaner energy sources. AI can optimize grid operations, predict peak demand, and improve grid stability leading to a more efficient integration of renewable energy sources, energy distribution, and reduced use of fossil fuels from the analysis of data from smart meters, weather patterns, and consumer behavior. Other climate approaches include increasing environmental monitoring and conservation efforts by analyzing with AI algorithms satellite imagery, sensor data, and other sources to monitor ecosystems, track biodiversity, detect deforestation, and identify sources of pollution. AI can also improve climate modeling and prediction capabilities. This will help to further understand climate data, atmospheric conditions and historical patterns to provide more accurate climate projections. Land use and sustainable agriculture using AI can help agricultural practices and land use to reduce environmental impacts and reduce carbon emissions from farming activities and thus help to preserve ecosystems.

Recommendation #22: While leveraging AI can support the above benefits of cutting greenhouse gases, the use of AI presents considerable environmental challenges that require careful consideration. For example, AI contributes to higher energy consumption due to computational demand[64] and thus leaves a huge water footprint.[65] Sustainable computer practices need to be addressed by developing energy-efficient AI algorithms and hardware.[66] Due to the massive amounts of data for which AI-enabled systems depend on, safeguarding data privacy, and preventing potential environmental harms associated with data storage and processing should be prioritized, including unintended negative consequences for environmental justice. Finally, to fully harness the potential of AI in achieving environmental objectives, it is desirable to combine AI technologies with public-private partnerships and comprehensive policies

---

[63] Waiker, S. 2021. AI Report: Competition Grows Between China and the U.S. Human Centered Artificial Intelligence. Stanford University. https://hai.stanford.edu/news/ai-report-competition-grows-between-china-and-us.

[64] Cho, R. 2023. AI's growing carbon footprint. Columbia Climate School. https://news.climate.columbia.edu/2023/06/09/ais-growing-carbon-footprint/.

[65] Li, P. et al. 2023. Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. Computer Science-AI Models. arXiv:2304.03271v1. https://arxiv.org/abs/2304.03271.

[66] Martineau, K. 2020. Shrinking deep learning's carbon footprint. MIT News. https://news.mit.edu/2020/shrinking-deep-learning-carbon-footprint-0807.

that are determined via international collaboration with experts from the environmental, technological, and policy domains.

Recommendation #23: The intersection between big data, citizen science, the environment and national security[67] is also worth mentioning. Given that data is the fundamental component of any AI-enabled system, identifying its origin, mode of collection, and type –be it from dedicated agencies, institutes and/ or organizations or from the public itself– can have direct implications for how to monitor events in real-time, what to choose and not choose to use and what subsequent decisions to make. Pursuing the health of the environment and its citizens ultimately becomes a national security imperative.

In summary, regulatory frameworks and laws designed for AI-enabled systems and their creative inventions have large economic and geopolitical implications. As long as regulations do not slow down or interfere with the progress of AI technologies, it is feasible and recommendable to do it in cases where high-risk AI products inflict harm to individuals and society at large. Balancing AI technological development and regulation are clearly important to bring the benefits of AI to the public and to minimize the technology's risks.

[67] de Sherbinin, A., Bowser, A., Chuang, T. R., Cooper, C., Danielsen, F., Edmunds, R., ... and Sivakumar, K. (2021). The critical importance of citizen science data. *Frontiers in Climate*, *3*, 20.