HUMAN FACTORS and ERGONOMICS SOCIETY

The AI Danger In the Making

Problem

The high-level push for untested, immature AI in defense operations poses a significant danger to our troops and innocent bystanders as well as weakens our national defense. By introducing AI, which can act in unexpected and unreliable ways and that is vulnerable to manipulation by foreign actors, we undermine the very national security we wish to support.

Background

AI has the potential to be beneficial,[1] but it is still not sufficiently intelligent or capable of performing reliably or competently in complex, real-world environments where human life and safety are on the line.[2; 3] Despite the hype and the potential of AI, the technology still faces many practical limitations.

Recent studies show that AI systems (including large language models; LLMs) have problems with "unauthorized compliance with non-owners, disclosure of sensitive information, execution of destructive system-level actions, denial-of-service conditions, uncontrolled resource consumption, identity spoofing vulnerabilities, cross-agent propagation of unsafe practices, and partial system takeover." [4] They often have difficulties correctly perceiving the world,[5] make up facts and information that do not exist,[6; 7] misreport on their status and actions,[4] poorly summarize gathered information,[8] and are ineffective and unsafe for fully autonomous operation.[10]

Not only is AI technology not ready for critical applications, but its very presence can simultaneously undermine the decision-making and performance of the people who oversee it. A National Academies of Science study cited decades of research showing that automation and AI lead people to suffer "from a poor understanding of what the systems are doing, high workload when trying to interact with AI systems, poor situation awareness and performance deficits when intervention is needed, biases in decision making based on system inputs, and degradation of manual skills."[9] Therefore, dangerous AI behavior may go undetected or not be addressed promptly when needed. Furthermore, sustained reliance on AI can erode workforce skills, further reducing people's capacity for effective intervention.

Recent events in which the government pushed Anthropic to allow the Dept of War to use its AI without constraints are potentially disastrous. Anthropic's statement that AI technology is not suitable for autonomous operations or mass surveillance is consistent with many AI experts and the scientific record cited above. For taking this position, Anthropic was punished with the loss of its government contracts and being labeled "a supply chain risk." This punitive response suggests a reckless disregard for the safety, security, and effectiveness of AI technology by decision makers, and encourages inappropriate risks.

HUMAN FACTORS AND ERGONOMICS SOCIETY
2001 K STREET NW, THIRD FLOOR NORTH
WASHINGTON, DC 20006
TEL. + 1 (202) 367-1114 | EMAIL: INFO@HFES.ORG

Automated technologies will continue to develop to be able to enhance mission-critical functions. However, the current push to implement AI with insufficient safeguards and testing is ripe for unintended consequences.[11] Every major technology being put into weapons systems requires careful and rigorous testing prior to operational deployment to protect the people using it and those potentially affected by it. A new, complex technology like AI should be no exception.

The high-level push for untested, immature AI in defense operations not only poses a significant danger to our troops and innocent bystanders but also undermines our national defense. By introducing AI that is vulnerable to manipulation by foreign actors, and which can act in unexpected and unreliable ways, we undermine the very national security we wish to support.

Action

We urge Congress to prohibit the use of AI in critical applications such as defense, aviation, healthcare, power systems, and transportation, unless:

1) It has appropriate safeguards in place, to:
    a. Support human oversight of AI actions by providing transparency to humans about its actions and intentions, its current state or mode, its goals and currently assigned functions, its ability to perform tasks in current and upcoming situations, how well it is performing tasks, and its projected actions;
    b. Be ultimately accountable to human decision makers while adhering to safety constraints; and
    c. Protect from cyber-attacks, manipulation, and interference by outside actors; and
2) It has been carefully tested and validated with respect to its ability to:
    a. Perform appropriately and accurately in the context of use;
    b. Detect and correct for emergent and/or unexpected behaviors or actions; and
    c. Support the people working with and monitoring the AI to enable them to effectively understand AI decisions and actions, as well as detect and respond to AI errors.

About the Human Factors and Ergonomics Society

With over 3,000 members, the Human Factors and Ergonomics Society (HFES) is the world's largest nonprofit association for human factors and ergonomics professionals. HFES members include psychologists, engineers, and other professionals who share a common interest in developing safe, effective, and practical human use of technology, particularly in challenging settings. HFES professionals have performed extensive research on how people perform when operating with automation and AI over the past 50 years.

References

1. Reimer, B., & Lindkvist, M. (2025). How to make AI useful. London, UK: LID Publishing.
2. Endsley, M. R. (2023). Ironies of artificial intelligence. Ergonomics, 66(11), 1656–1668.
3. Cummings, M. (2021). Rethinking the maturity of artificial intelligence in safety-critical settings. AI Magazine, 42(1), 6–15.

HUMAN FACTORS AND ERGONOMICS SOCIETY
2001 K STREET NW, THIRD FLOOR NORTH
WASHINGTON, DC 20006
TEL. + 1 (202) 367-1114 | EMAIL: INFO@HFES.ORG

4. Shapira, N., Wendler, C., Yen, A., Sarti, G., Pal, K., Floody, O., . . . Prakash, N. (2026). Agents of Chaos. arXiv preprint arXiv:2602.20021.
5. Ollikka, N., Abbas, A., Perin, A., Kilpeläinen, M., & Deny, S. (2024). A comparison between humans and AI at recognizing objects in unusual poses. arXiv preprint arXiv:2402.03973.
6. Chelli, M., Descamps, J., Lavoué, V., Trojani, C., Azar, M., Deckert, M., . . . Ruetsch-Chelli, C. (2024). Hallucination rates and reference accuracy of ChatGPT and bard for systematic reviews: comparative analysis. Journal of medical Internet research, 26(1), e53164.
7. Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., & Ho, D. E. (2025). Hallucination-free? Assessing the reliability of leading AI legal research tools. Journal of empirical legal studies, 22(2), 216–242.
8. Adel, A., & Alani, N. (2025). Can generative AI reliably synthesise literature? exploring hallucination issues in ChatGPT. AI & SOCIETY, 1–14.
9. National Academies of Sciences Engineering and Medicine. (2021). Human-AI teaming: State-of-the-art and research needs. Washington, DC: National Academies Press.
10. Rabanser, S., Kapoor, S., Kirgis, P., Liu, K., Utpala, S., & Narayanan, A. (2026). Towards a Science of AI Agent Reliability. arXiv. https://doi.org/10.48550/arXiv.2602.16666.
11. Department of War. (2026). Artificial Intelligence Strategy for the Department of War. Washington, DC. (January 9, 2026)