NCITE Project Summary

Malevolent Creativity: Agentic AI and Robotic Systems as Emerging Threats

September 2025

Sam Hunter (PI), University of Nebraska at Omaha
Joel Elson (Co-PI), University of Nebraska at Omaha

This project examines how next-generation artificial intelligence—agentic (autonomous decision-making) and physical/embodied systems (robotics)—may enable new forms of malevolent creativity linked to terrorism. The team will map documented and potential malicious use cases and test how these technologies shape harmful idea generation. The project will have two experimental studies that vary group identity salience and perceived justification (defense vs. offense), and compare AI use for information-gathering, planning, or ideation outsourcing. Findings will clarify which AI capabilities most increase malevolent ideation and under what conditions, and distinguish risks posed by agentic vs. physical systems. The work will also identify counter-use opportunities and produce practitioner-focused guidance, exercises, and briefs to help DHS apply these technologies responsibly while mitigating emerging threats.

Impact Statement

This research gives DHS operators concrete threat maps, evidence-based risk cues, and practical counter-use playbooks for agentic and robotic AI, enabling earlier detection, safer procurement, and more targeted prevention and response across the homeland security and defense field.

Policy Impact

- Executive Order 14179, “Removing Barriers to American Leadership in Artificial Intelligence” (Jan. 23, 2025): Evidence on misuse pathways and contextual risk factors supports agency implementation that accelerates AI while guarding against rights, safety, and security harms in operational settings.
- Executive Order 14320, “Promoting the Export of the American AI Technology Stack” (Jul. 23, 2025): The project highlights potential foreign and domestic security externalities of exporting full-stack AI/robotics packages, informing interagency risk reviews and monitoring priorities relevant to DHS.
- Office of Management and Budget Memorandum M-25-21, “Accelerating Federal Use of AI through Innovation, Governance, and Public Trust” (Apr. 3, 2025): Experimental findings on when and how AI elevates malevolent ideation inform risk assessment, impact testing, and model governance requirements for federal AI use.
- Office of Management and Budget Memorandum M-25-22, “Driving Efficient Acquisition of Artificial Intelligence in Government” (Apr. 3, 2025): Use-case analyses and red-flag criteria help acquisition officials specify safeguards for agentic/robotic AI in requirements, evaluations, and vendor oversight.
- DHS Directive 139-08, “Artificial Intelligence Use and Acquisition” (Jan. 16, 2025): The project’s risk mapping and translation products directly inform safe, human-centered DHS AI use and procurement, including distinguishing high-risk agentic/embodied capabilities and providing operator-ready mitigations.

ACKNOWLEDGMENT AND DISCLAIMER

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 20STTPC00001-05. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

End User Offices with Direct Operational Impact

DHS Office of Intelligence & Analysis – Emerging Technology Branch	DHS Science and Technology Directorate
Office of the Director of National Intelligence – National Counterterrorism Center	Customs and Border Protection – Chief Information Officer (ICO)

Expected Findings and Outputs

- A structured catalog of documented and plausible malicious use cases for agentic AI (e.g., autonomous planning, target selection, amplification of harmful decisions) and physical AI (e.g., robotic surveillance, coercion, or harm), with distinct threat vectors for each.
- Under conditions of strong group identity cues and perceived justification (defense or offense), access to agentic/physical AI is expected to increase malevolent ideation and novelty of harmful ideas compared to controls.
- Outsourcing ideation to AI versus using it for information-gathering/planning is expected to yield different risk profiles; prompt characteristics and user engagement patterns will be linked to higher-risk outputs.
- Practical counter-use opportunities (e.g., scenario planning, real-time detection cues) will be identified for ethical, secure application by DHS practitioners.

NCITE Strategic Priority

Tactics – The project directly addresses how threat actors could operationalize agentic and embodied AI to generate novel methods of harm, providing DHS with tactics-focused indicators and mitigations.

NCITE Operational Area of Excellence

Research and Development Translation and Transition — Deliverables (case briefs, tabletop exercises, co-design workshops) convert findings into actionable guidance for DHS end-users and acquisition officials.

Methodology

The project will commence in three phases: (1) systematic reviews synthesize documented/potential malicious use cases of agentic and physical AI using threat reports and primary sources (e.g., chat logs, case reports, court documents); (2) controlled experimental studies to manipulate group identity salience, perceived role (defense vs. offense), and AI use (information retrieval vs. ideation outsourcing) to measure effects on malevolent ideation; (3) co-development with DHS of tabletop exercises and workshops to translate findings into ethical, secure counter-use guidance and products.

Please visit the NCITE website for more information on the project at ncite.unomaha.edu

ACKNOWLEDGMENT AND DISCLAIMER

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 20STTPC00001-05. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.