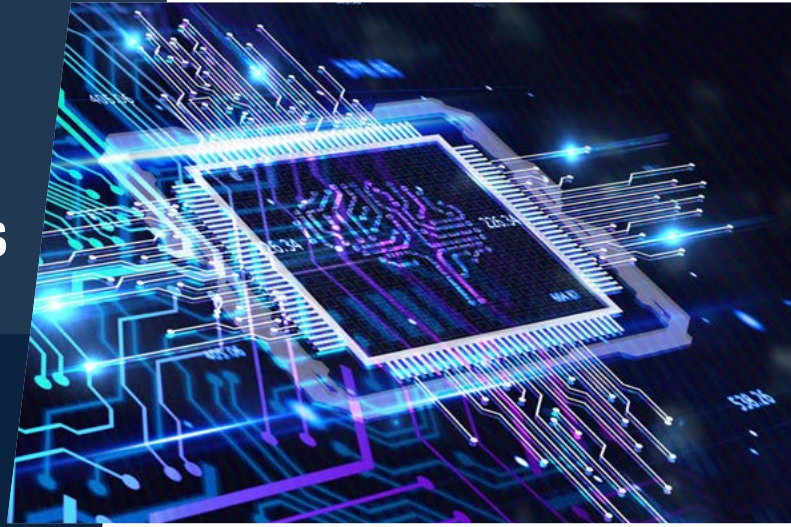


Developing and enhancing artificial intelligence systems users can rely on.



Artificial intelligence (AI) has emerged as a critical factor that will increasingly shape a nation's ability to compete and contribute on the global stage. While new AI systems are promising, most also entail significant ethical and reputational risks. None of the most widely used AI tools are designed to align with the Department of Defense's goal of making data visible, accessible, understandable, linked, trustworthy, secure, and interoperable.

Working alongside technology experts within the U.S. military as well as with researchers at Purdue University and Indiana University, researchers at the University of Notre Dame are working to remedy this problem. They are developing a comprehensive, systematic approach to creating an environment where AI can flourish in a trustworthy manner.

Their project, called "Trusted AI," has identified six widely shared values or dimensions of Trusted AI. The six dimensions are:

Learn More

crc.nd.edu

Contact Us

crcsupport@nd.edu



Explainability - Can we explain how the AI arrives at inferences?

Safety and robustness - Will the AI work as expected—not just in the lab but in real, live contexts?

Fairness - Can we ensure the AI will not reproduce patterns of bias and discrimination?

Privacy - Are we confident that the data the AI uses will be held safely and confidentially?

Environmental wellbeing - Can the AI be trained and developed with minimal negative environmental impact?

Accountability and auditability - Can we identify who is responsible, and can we confirm that the AI is working as expected?

TRUSTED ARTIFICIAL INTELLIGENCE

An Iterative Process

AI technologies—and the data sets they rely on—are constantly evolving. Thus Trusted AI represents an ongoing cycle. Trusted AI researchers work to ensure that each of the dimensions informs every stage in the process, from the initial collection of data to the output, or “inference,” the AI provides. Only when there is an unbroken “chain of trust” can we be sure the end result is trustworthy.

A Data-Centric Approach

One crucial way Trusted AI aligns with the U.S. military’s requirements is through its data-centric approach that places data quality at the heart of the AI development process. Trusted AI needs trusted data, because AI models are only as good as the data they learn from. With this requirement in mind, Trusted AI researchers emphasize the importance of tools and practices that systematically improve the data used to build AI systems. By prioritizing data quality, encompassing aspects such as cleanliness, relevance, diversity, and contextual richness, AI systems can achieve better performance with potentially simpler architectures.

Workforce Development

Trusted AI is part of the Scalable Asymmetric Lifecycle Engagement (SCALE) workforce development program, and a key part of its effort is to train the future military officers, scholars, and tech industry leaders who will make trusted AI a reality. Students directly involved in the research process will gain critical skills that will help them guide the ethical development of future AI tools. The Trusted AI team also educates younger students in the principles of Trusted AI through presentations and by welcoming dozens of high school students to Notre Dame’s campus for the Center for Research Computing’s Summer Scholars program.

Fostering National and International Collaboration

Through the Notre Dame-IBM Technology Ethics Lab, the University convenes research and practitioner-focused events related to ethical AI. These have included the Rome Call for AI Ethics: A Global University Summit, which was co-organized by the Pontifical Academy for Life, IBM, and Notre Dame and hosted by the Lab. Leaders from Notre Dame and seven other universities signed the Rome Call during the summit.

The Center for Research Computing at Notre Dame

The Notre Dame researchers guiding the Trusted AI project are part of the Center for Research Computing (CRC). The CRC is an innovative and multidisciplinary research environment that supports collaboration to facilitate multidisciplinary discoveries through advanced computation, software engineering, artificial intelligence, and other digital research tools. The Center enhances the University’s innovative applications of cyberinfrastructure, provides support for interdisciplinary research and education, and conducts computational research. Learn more at crc.nd.edu.

