

Ethical Principles for Artificial Intelligence

These principles will apply to both combat and non-combat functions and assist the U.S. military in upholding legal, ethical and policy commitments in the field of AI. The department's AI ethical principles encompass five major areas:

1. Responsible. DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities.
2. Equitable. The Department will take deliberate steps to minimize unintended bias in AI capabilities.
3. Traceable. The Department's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedure and documentation.
4. Reliable. The Department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire life-cycles.
5. Governable. The Department will design and engineer AI capabilities to fulfill their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior.

References

DOD Adopts Ethical Principles for Artificial Intelligence, U.S. Department of Defense, 24 February 2020,

<https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/>

A Closer Look: The Department of Defense AI Ethical Principles, The Joint Artificial Intelligence Center, 24 February 2020, https://www.ai.mil/blog_02_24_20-dod-ai_principles.html