Artificial Intelligence

Clear, comprehensive and actionable definitions of Artificial Intelligence (AI) are elusive given the rapid pace of its evolution. Generally, the term refers to a suite of capabilities executed via computer systems that are designed to perform the cognitive functions of the human mind. Among these are machine learning, deep learning, computer vision, natural language processing and generative AI. The AI Triad is a useful paradigm for understanding and visualizing artificial intelligence in general. Simply put, artificial intelligence is machine learning systems using computer power to execute algorithms that learn from data to use perception, reasoning, learning and creativity to help solve problems.

Major Applications

- **Machine learning:** the ability of computer systems to use algorithms and statistical models to analyze large sets of data and draw lessons and inferences from patterns and details in data without explicit human instruction.
- **Deep learning:** a type of machine learning based on artificial neural networks (like those of the human mind) in which successive layers of processing extract increasingly advanced features and insights from data.
- **Computer vision:** the ability of computers and systems to interpret and understand the visual world, derive meaningful information from digital images and video, and respond with actions and recommendations based upon that information.
- **Natural language processing:** the ability of a computer system to interpret and understand human language and respond to linguistic input.
- **Generative AI:** refers to deep-learning models such as ChatGPT and DALL-E that can generate original high-quality text, images, audio, and other content based on training data.

Proliferation Concerns

Notwithstanding the uncertainty of AI's trajectory, it could pose concerns along several discrete areas of the nuclear value chain. For instance, AI will likely facilitate access to peaceful uses of nuclear science and technology, including nuclear materials that could be diverted or abused for illicit purposes. Policymakers have already moved to head off any push to prematurely integrate AI into nuclear command and control out of concern that it could prove too destabilizing. In November 2024, U.S. President Joe Biden and Chinese President Xi Jinping agreed not to surrender decision-making authority to artificial intelligence when they "affirmed the need to maintain human control over the decision to use nuclear weapons."

Moreover, the algorithms underpinning large language models such as ChatGPT and DALL-E have been shown to be susceptible to workarounds. In several cases, users have been able to "trick" the system by tweaking the input prompt in a way that bypasses the AI's safeguards. Typically, AI models are created with safeguards that recognize illicit behavior — such as prompting it to explain how to create a biological weapon — but have been circumvented in cases where the user prompted the AI to address a hypothetical scenario.

AI models can also be repurposed for nefarious means. In 2022 for example, an AI system augmented by a pharmaceutical research company to help with drug discovery ended up producing an extensive list of novel chemical nerve agents. The ultimate impact of AI on proliferation will ultimately be a result of the manner in which it is combined with other technological advances such as materials science, advanced manufacturing, and quantum computing.