AI and American National Security

by Julie George

U.S. Vice President Kamala Harris looks on as U.S. President Joe Biden signs an executive order after delivering remarks in the White House on advancing the safe, secure, and trustworthy development and use of artificial intelligence on October 30, 2023. Biden's executive order aimed for the U.S. to "lead the way" in global efforts at managing the new technology's risks. The order directed federal agencies to set new safety standards for AI systems and required developers to "share their safety test results and other critical information with the US government," according to a White House statement. BRENDAN SMIALOWSKI/AFP VIA GETTY IMAGES

Artificial intelligence (AI), especially generative AI, is often claimed as an emerging technology that will disrupt all facets of society. Generative AI is a field of AI that can produce novel content such as text, images, videos, audio, and music. Artificial intelligence is rapidly developing in applications and use cases, from large language models like ChatGPT to deepfakes. Deepfakes are photos, videos, or audio recordings that have been manipulated using AI to make them appear real. As a dual-use technology, artificial intelligence leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind. This technology can enhance visual perception, speech recognition, decision-making, and translation at breakneck speeds. We have witnessed substantial progress in generative AI and how AI can be misused through disinformation and algorithmic discrimination. Experts assert that international cooperation is needed to foster the opportunities that these new technologies hold while protecting societies from their risks. This chapter outlines how AI impacts U.S. national security, discusses pathways of global governance, and highlights policy debates on AI. Overall, attention to artificial intelligence will be increasingly important as this innovation continues to take shape and impact the nature of national and international security.

DR. JULIE GEORGE *received her PhD in the government department at Cornell University, specializing in international security. Broadly, her research examines the proliferation of emerging technologies. This focus has led her to engage a broad selection of scholarship across science and technology studies, history, international organizations, and law. Most recently, she was a postdoctoral fellow in the International Security Program within the Belfer Center for Science and International Affairs at the Harvard Kennedy School. She is an affiliate of the Center for International Security and Cooperation (CISAC) at Stanford University and a Fellow at the Cornell Brooks School Tech Policy Institute.*

Definition and consequences of AI

In one survey, only 33% of respondents believed they used technology that features artificial intelligence. However, the actual number is, shockingly, much higher. In reality, 77% of the population uses an AI-powered service. AI is all around us—from self-driving cars to virtual personal assistants, AI can learn from experience, understand natural language, recognize patterns, solve problems, and make decisions. In simple terms, AI enables machines to perform tasks that typically require human assistance. Perhaps more significantly, artificial intelligence is a dual-use technology that has and will continue to impact and interact with various components of society—governments and institutions, social media, conflict, and much more.

Regarding governments and institutions, artificial intelligence impacts political power, regime dynamics such as in democracies and authoritarian countries, and perception of trust. AI techniques can shape information flows and perceptions, affect democratic processes, and strengthen authoritarian regimes. For example, artificial intelligence has impacted electoral processes, as seen with voter profiling through targeted political advertisements based on social media profiles in the 2016 U.S. presidential election and the United Kingdom Brexit referendum. In addition, authoritarian regimes have largely used social media bots to enforce propaganda efforts both inside and outside a home country. Artificial intelligence can also enable the spread of disinformation and influence government operations, leading to societal disruption.

Regarding social media, artificial intelligence is used for data collection, refining algorithms for content, and amplifying information. Recent efforts have focused on "responsible AI" as seen across social media platforms. In 2021, X, the microblogging site, focused on "responsible machine learning (ML)," an initiative that was designed not only to increase the transparency of the AI systems used by X but also to improve the fairness of the algorithms and to provide users with "algorithmic choice" when it comes to the technologies that might affect them. Looking at Facebook/META, the company has claimed to help advance this emerging field and spread the impact of such work by creating an interdisciplinary responsible AI (RAI) team several years ago. Within RAI, the fairness team works with product teams across the company to foster informed, context-specific decisions about measuring and defining fairness in AI-powered products. National and international security concerns are increasingly raised with social media platforms, such as the popular platform TikTok, which is banned in India as well as on U.S. federal government devices.

! Before you read, download the companion **Glossary** that includes definitions, a guide to acronyms and abbreviations used in the article, and other material. Go to **www.fpa.org/great_decisions** and select a topic in the Topic Resources section.

The Enigma machine was patented in 1918 by the German engineer Arthur Scherbius and produced commercially from 1923. The German government, impressed by its security, acquired the rights to the machine and adapted it for military use. Throughout World War II, Germany and its allies encrypted military messages using Enigma machines, and by 1945 over 40,000 were in use. The Germans considered the Enigma code to be unbreakable. However, thanks to combined Allied efforts, the codebreakers at Bletchley Park in Buckinghamshire (the British Army's intelligence center) managed to intercept and decipher the code with the help of Colossus, the world's first electronic programmable computer. This feat is said to have dramatically shortened the war. SCIENCE & SOCIETY PICTURE LIBRARY/GETTY IMAGES

Regarding warfare and future conflict, we cannot ignore the potential use of artificial intelligence in this domain. For example, in the Russia-Ukraine War, there are current applications of semiautonomous drones, including the Switchblade 600 and Lancet, which emphasize the human-machine relationship. A semiautonomous unmanned system is a "mode of operation wherein the human operator and/or the unmanned system plan and conduct a mission and requires various levels of human-robot interaction." Guided by human operators, these drones have been used for surveillance and target identification purposes. As it stands, the Switchblade 600 requires a human operator to appoint targets over a live video feed. Thus, we witness the increasing ways in which artificial intelligence intersects with governments, civil society, and the military domain.

In this essay, I begin with a history of artificial intelligence from the 1940s to the present day while also clarifying the various definitions of AI over the decades. After defining artificial intelligence, I then outline the dual-use nature of advanced artificial intelligence, investigating its applications in both society and the military. Afterward, I underscore the implications of artificial intelligence, discuss key policy debates on AI, and offer policy recommendations for AI and U.S. national security. I also highlight the pathways of global governance of AI, including various stakeholders such as governments, the private technological sector, and civil society. Last, I outline my key takeaways on the opportunities and risks posed by this emerging technology.

History of AI: past, present, and future

As a dual-use technology, artificial intelligence leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind. This technology can enhance visual perception, speech recognition, decision-making, and translation at breakneck speeds. We have witnessed substantial progress in generative artificial intelligence, such as in large language models (LLMs) within foundational models, especially in recent years. Trained on large amounts of data, LLMs can analyze and understand natural language, as well as craft responses. However, it is also essential to understand the history of artificial intelligence and the landmark goals states and non-state actors have achieved. In this section, I discuss the history of artificial intelligence since the mid-20th century, which can also be seen in the chronological timeline. This history shows that AI expertise, infrastructure, and institutions have shaped the technology.

In the 1940s and 1950s, many scientists from various fields, such as mathematics, psychology, engineering, economics, and political science, met to discuss the possibility of artificial intelligence. Considered the father of modern computer science, Alan Turing was famous for his research developing the first modern computers, decoding the encryption of German Enigma machines during World War II, and creating the Turing test, a basis for artificial intelligence. Turing discussed the logical framework behind intelligence machines and how to both build and test their intelligence in his 1950 paper, "Computing Machinery and Intelligence." The 1950s included computers that were learning checker strategies, speaking English, and solving word problems in algebra. In 1956, the first artificial intelligence program, the Logic Theorist, was presented at the Dartmouth Summer Research Project on Artificial Intelligence hosted by John McCarthy and Marvin Minsky.

In the 1960s, the U.S. Department of Defense began to fund substantial artificial intelligence research and establish laboratories. Computer scientist John McCarthy coined the term "artificial intelligence" in 1965. During this first wave of AI research, scientists studied generality in the models of cognition. In other words, they fed machines a large amount of training data from which the machines could extract patterns or other meaningful information, which they could use to learn and make accurate predictions when fed with new data.

In the second wave of AI interest

THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

1940s

After successfully breaking the German Enigma code in World War II, scientists from various fields gathered to discuss the possibilities of early computing. One such pioneer, Alan Turing, proposed the **Turing test**, n which participants in a blind study were asked to distinguish between human- and machine-generated written communication.

1950s

Early programming allowed computers to learn checker strategies, speak in English, and solve word problems in algebra. In 1956, the first artificial intelligence program, the **Logic Theorist**, was presented at the Dartmouth Summer Research Project on Artificial Intelligence hosted by John McCarthy and Marvin Minsky.

1960s and 1970s

The **U.S. Department of Defense** begins funding labs to study artificial intelligence, a term first coined by John McCarthy in 1965. Computers fed large amounts of data could automatically extract patterns and other information, then use it to make accurate predictions with subsequent data.

1980s

Experiments with neural networks and deep learning advances helped develop more sophisticated language learning models and computer vision. Edward Feigenbaum established expert systems to imitate a human expert's decision-making process, which then went on to be used mainly in industries.

1990s and 2000s

Among significant AI achievements, **IBM's Deep Blue** defeated a reigning world chess champion, Garry Kasparov. Similarly, Google's **AlphaGo** defeated world Go champion Ke Jie in 2017.

Recent developments

The latest advancements in AI include visual deep learning models like **DALL-E** and **DALL-E 2**, which are text-to-image models, and generative pretrained transformers such as **ChatGPT** and **GPT-4**, and most recently, **OpenAI's** release of **o1**. The rise of generative AI across saw natural language processing, image and video generation, and data processing. Other large language models include **Llama**, **Claude**, **AlphaGo**, **Siri**, and **Alexa**, which utilize deep learning and natural language processing. Machine learning, however, has raised concerns about establishing guidelines regarding the safety and ethics of artificial intelligence.

in the 1980s, scholars focused on experimenting with neural networks and deep learning advances to develop more sophisticated language learning models and computer vision. Notably, Edward Feigenbaum established expert systems that would imitate a human expert's decision-making process, which then went on to be used mainly in industries. The 1990s and 2000s saw significant AI achievements, including IBM's Deep Blue defeating the reigning chess champion Garry Kasparov. Similarly, Google's AlphaGo defeated world Go champion Ke Jie in 2017.

Recent advancements in AI include visual deep learning models like DALL-E and DALL-E 2, which are text-to-image models, and generative pretrained transformers such as ChatGPT and GPT-4. These have sparked great interest from the general public, industry competitors, and governments across the globe. In September 2024, OpenAI released o1, which is a model designed to spend more time thinking before it responds to complex tasks and difficult problems in science, coding, and math. In the latest decade, the world has witnessed the rise of generative AI across natural language processing, image and video generation, and data processing.

Most recently, AI has made it possible for machines to learn from experience based on annotated data, recognize patterns in the data, modify algorithms, create text and images, and execute human-like tasks. Artificial intelligence automates repetitive learning through data, adapts progressive learning algorithms, and works to increase accuracy through deep neural networks. Some recent examples of artificial intelligence include large language models (e.g., ChatGPT, Llama, Claude), AlphaGo, Siri, and Alexa, which utilize deep learning and natural language processing. These technologies have the potential to function and respond without direct human involvement due to breakthroughs in machine learning. However, artificial intelligence has raised concern with many individuals as algorithms pick up or amplify societal biases around factors such as race or gender. Governments, civil society groups, and the technology industry are currently exploring rules and guidelines regarding the safety and ethics of artificial intelligence. The following section will delve into military and civilian uses of AI.

Dual-use technology: the nexus of AI and national security

Dual-use technologies are innovations that can be used for both civilian and military purposes. Artificial intelligence is one such dual-use technology, which has significant commercial and military value. Regarding military applications, AI can be used for defense, intelligence, surveillance, economic and financial tools of statecraft, cybersecurity, and the nuclear domain. For example, the U.S. Department of Defense's Artificial Intelligence Strategy (2018) outlined how AI, the ability of machines to perform tasks that typically require human intelligence, will strengthen the military, increase the effectiveness and efficiency of operations, and enhance the security of the homeland.

Artificial intelligence can impact many aspects of national security. Former Deputy Assistant Secretary of Defense for Force Development and Emerging Capabilities Michael Horowitz argues that the effect of artificial intelligence is an "enabler"—the effect of artificial intelligence on military power and international conflict will depend on particular AI applications for militaries and policymakers. There are three potential application areas of AI, which illustrate why militaries have interest: 1) narrow AI applications to process information offer the potential to speed up the data interpretation process, freeing human labor for higher-level tasks, 2) from hypersonics to cyberattacks, senior military and civilian leaders believe the speed of warfare is increasing, and 3) AI could enable a variety of new military concepts of operation on the battlefield, such as the "loyal wingman" idea, which posits a human airplane pilot or tank driver who could coordinate a number of uninhabited assets as well. In this section, I will highlight the dual uses of AI and prominent scholarship regarding defense, intelligence, operations, cybersecurity, and the nuclear domain.

Defense & AI

Artificial intelligence can be leveraged across combat operations. Specifically, AI can bolster situational awareness through deep neural networks and small robotic sensors. Through these applications, AI can collect large amounts of data, process that information, and help with classification. For example, the U.S. Department of Defense (DoD) had an artificial intelligence initiative known as Project Maven, which sought to leverage machine learning and deep learning that autonomously extracts objects of interest from moving or still imagery from drones. The objective behind Project Maven was to improve drone targeting. The DoD says Project Maven, "enhances the performance of the human-machine team by fusing intelligence and operations through AI/ML and augmented reality technology. Project Maven seeks to reduce the time required for decision-making to a fraction of the time needed without AI/ML."

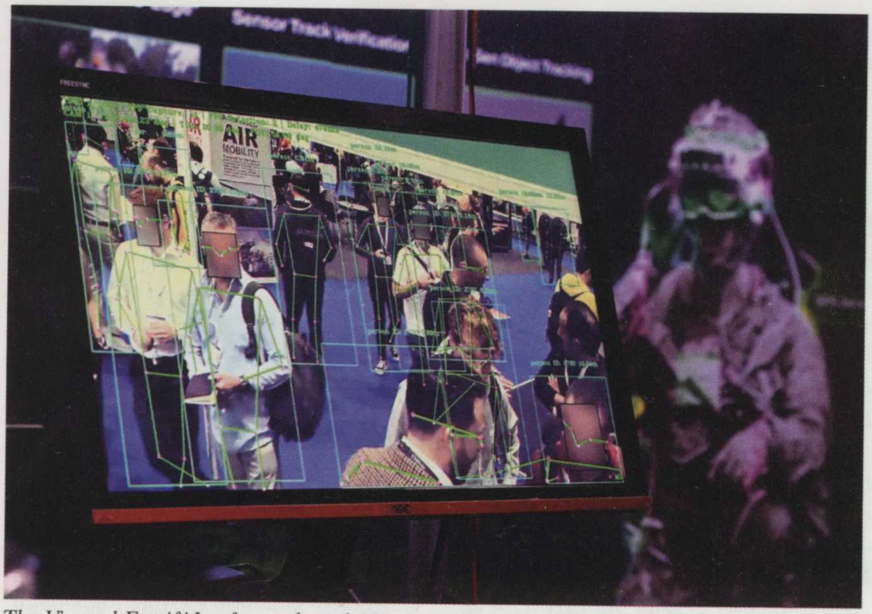
Examples of dual-use applications of AI also include decoys and camouflage through generative adversarial networks (GANs). GANs are a type of machine learning model that creates new data that is similar to training data; GANs can be leveraged to produce puzzling and convincing deepfakes or spoofing attacks. AI-powered spoofing attacks use AI to produce false content that appears realistic and can be used to deceive people.

Furthermore, AI can be used within settings such as command and control systems and the electromagnetic spectrum domain. With the former, autonomous systems that have been delegated authority for specific efforts may potentially result in decision-making with the assistance of AI. For example, AI could provide processed information to commanders at faster speeds. With this assistance, commanders could better understand the conflict and overall environment. Regarding the latter, AI can enhance the electromagnetic spectrum domain in terms of jamming and communication. AI can send various signals that could complicate the environment, making it difficult for the adversary to assess the signals. AI may also reduce decision-making time, making it harder for adversaries to anticipate, respond, or communicate effectively.

Intelligence, operations, & AI

AI is also used for intelligence purposes, such as data collection and robust analysis. For example, AI helps sort through big data to make connections from the information, identify suspicious activity, assess patterns, connect networks, and predict future efforts.

AI is far more commonly employed in various military and national security tasks, such as classifying targets in satellite imagery, streamlining the analysis of video gathered by drones, and operating physical systems like autonomous planes. First, AI could pose challenges to operational coordination by complicating burden-sharing and the interoperability of multinational forces. Second, AI could hamper alliance and coalition decision-making by straining the processes and relationships that undergird decisions on the use of force. Third, by increasing the speed of warfare, AI could decrease the time leaders, from the tactical to strategic levels, have to debate policies and make decisions. MIT Professor Erik Lin-Greenberg argues that in the military domain, a rival could poison imagery data to throw off AI target recognition systems, leading the system to miss military targets, classify them as nonmilitary ones, or identify civilian infrastructure as military facilities.



The Vizgard FortifAI software-based AI engine for surveillance monitoring is displayed on September 26, 2023, in London, England. The system operates face redaction to avoid non-compliance with regulations by anonymizing faces and license plates with a real-time redactor. The AI engine integrates with camera-based security systems and unmanned platforms to reduce operator burden and increase the probability of detecting potential threats by recognition of basic human physical behaviors such as walking, running, climbing, crouching, or crawling. JOHN KEEBLE/GETTY IMAGES

Cybersecurity & AI

The cybersecurity domain is ripe for the integration of AI. Notably, in 2016 the National Security Agency (NSA) Director Michael Rogers stated that the agency sees AI as "foundational to the future of cybersecurity." Over the years, government agencies have invested in understanding the nexus of cybersecurity and AI. For example, the Defense Advanced Research Program Agency (DARPA) launched the Cyber Grand Challenge, which called for a "head-to-head fight between autonomous machines in cyberspace." In this challenge, each system was able to automatically identify and exploit cyber vulnerabilities in its adversaries while patching its own weaknesses and defending itself from outside cyberattacks. The Department of Defense also established Project Voltron, which was used to establish and deploy autonomous cybersecurity systems to identify and patch vulnerabilities in the U.S. military.

Nuclear & AI

AI also has interesting implications for the nuclear domain. Stanton Nuclear Fellow Heather Williams and NATO Director of Nuclear Policy Jessica Cox highlight the ubiquitous nature of artificial intelligence in the military and civilian environments and across the globe. Using Michael Horowitz's framework, they define artificial intelligence as "systems that select and engage targets on their own" (for example, computer-guided precision weapons) or "intelligent machines capable of cognitive judgments on par with humans." They argue that artificial intelligence can offer opportunities and risks, depending on its application. They also assert that artificial intelligence can assist in deterrence and arms control, which has previously been overlooked or ignored in scholarship. Finally, the authors write how incorporating artificial intelligence into early warning and decision-making could provide time for de-escalation. The nuclear policy community continues to recruit and bolster talent with technical expertise and knowledge of artificial intelligence, cyber, and nuclear weapons.

One 2022 report published by the European Leadership Network called

REAIM
Responsible AI in the Military domain Summit
인공지능의 책임있는 군사적 이용에 관한 고위급회의
REAIM

South Korean Foreign Minister Cho Tae-yul delivers a speech at the closing session of the Responsible AI in the Military Domain summit in Seoul on September 10, 2024. The global summit agreed that humans, not AI, should make the key decisions when it comes to using nuclear weapons. JUNG YEON-JE/AFP VIA GETTY IMAGES

“Nuclear Decision-Making, Complexity and Emerging and Disruptive Technologies” outlines how emerging and disruptive technologies may impact nuclear decision-making and potentially increase escalation in a regional conventional conflict. The report highlights several technologies impacting nuclear decision-making: artificial intelligence and big data analytics, AI-enabled cyber operations, cheaper and smarter space assets and space weapons, autonomous systems, hypersonic weapons, and quantum technology. However, they also highlight how such technologies may encourage restraint in some scenarios, primarily based on various combinations of such technologies. Overall, they encourage decision-makers to understand the prospects and limitations of these technologies, especially artificial intelligence, and fully partake in risk assessments and efforts for cooperation.

In sum, AI has concerning and interesting implications for both the military and civilian domains. At the same time, many scholars highlight the problem of overconfidence in the possibilities of AI—specifically, how overconfidence may lead militaries to apply machine learning to situations that are too complex for brittle algorithms. There are three levels at which a lack of integration and socialization of AI can cause breakdowns: 1) between new AI systems and legacy systems, 2) between the human operators and decision-makers and AI systems, and 3) between organizations that utilize AI systems. However, the contributors, Jacek Durkalec, Anna Péczeli, and Brian Radzinsky, note areas where AI raises issues for military applications and command and control. For example, “data poisoning” attacks would allow adversaries to manipulate algorithms by injecting bad data into training datasets. In this piece, the authors also point out that many military applications of AI will be inconsequential. Still, others could be concerning, such as using AI for swarming, nuclear operations, or lethal decision-making. In the next section, I will outline key policy debates regarding AI, including both opportunities and limits on global governance of the technology.

Key policy debates on AI: opportunities and limits on global AI rules of the road

Efforts to govern artificial intelligence are a salient issue in the U.S. and across the globe. As one can imagine, there are competing interests in the governance of AI, and this task will require the committed efforts of governments, the private sector, nongovernmental organizations, and civil society. Debates of AI governance include concerns over data privacy, export controls of hardware, potential regulations for industry, and assessment of AI risks. Below are a few examples of leading initiatives focused on minimizing AI risks while benefiting from the technology’s capabilities.

The U.S. (Blueprint for an AI Bill of Rights, CHIPS Act, and Executive Order 14110)

The White House Office of Science and Technology Policy’s (OSTP) Blueprint for an AI Bill of Rights serves as a guide for a society that protects people from these threats while also using technologies in ways that reinforce the country’s highest values. In October 2022, OSTP established a road map for the responsible use of AI with five core principles to guide and govern the effective development and implementation of AI systems. These five principles include safe and effective systems, algorithmic discrimination protections, data privacy, notice and explanation, and human alternatives to AI systems.

In addition, in 2022 the U.S. passed the CHIPS and Science Act, which focused on hardware components for artificial intelligence. Through the act, the Biden administration aimed to strengthen American manufacturing, supply chains, and national security; and invest in research development, science and technology, and lead-in industries, such as artificial intelligence, quantum computing, nanotechnology, and clean energy. The CHIPS and Science Act committed $280 billion focused on increasing scientific research and advanced semiconductor manufacturing capacity to boost U.S. competitiveness against China. In turn, China has announced that it seeks to reduce its "external [foreign] dependence for key technologies and advanced equipment." While China is also a leading country in AI, the country is struggling to obtain core AI components. As a result, it must pursue other forms of development, such as domestic efforts, due to export controls. Specifically, U.S. export controls on AI components such as semiconductors have stopped AI computer chip designers like NVIDIA and AMD from selling their high-end chips for AI and supercomputing to China.

More recently, the Biden administration published Executive Order 14100, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, on October 30, 2023. Executive Order 14100 focuses on "responsible AI development and deployment through federal agency leadership, regulation of industry, and engagement with international partners." Moreover, the order directs federal entities on issues such as safety and security, innovation and competition, worker support, consideration of AI bias and civil rights, consumer protection, privacy, federal use of AI, and international leadership. This order works alongside the Blueprint for an AI Bill of Rights and the National Institute of Standards and Technology's AI Risk Management Framework.

The United Nations Educational, Scientific, and Cultural Organization (UNESCO)

In November 2021, the 193 member states at UNESCO's General Conference adopted the Recommendation on the Ethics of Artificial Intelligence, the very first global agreement on the subject. The agreement aims to protect and promote human rights and human dignity, and to be an ethical guiding compass and global normative bedrock with respect for the law in the digital world. Specifically, the standard looks at protecting data, banning the use of AI systems for social scoring and mass surveillance, monitoring and evaluating AI systems for countries and companies, and protecting the environment. According to UNESCO, the agency stated that AI is eliciting unprecedented challenges: "We see increased gender and ethnic bias, significant threats to privacy, dignity, and agency, dangers of mass surveillance, and increased use of unreliable Artificial Intelligence technologies in law enforcement, to name a few. Until now, there were no universal standards to provide an answer to these issues." Notably, UNESCO's global agreement highlights the role of states and the private sector in using artificial intelligence for various domains. UNESCO also underscored that "industry self-regulation is clearly not sufficient to avoid these ethical harms, which is why the recommendation provides the tools to ensure that AI developments abide by the rule of law, avoiding harm, and ensuring that when the harm is done, accountability and redressal mechanisms are at hand for those affected." All UNESCO member states endorsed the UNESCO global framework, seeking to increase the benefits of AI and decrease technology risks.

European Union AI Act

In April 2021, the European Union (EU) put forth the AI Act, the first proposed law on artificial intelligence by a significant regulator. The act assesses AI applications based on three risk categories: 1) "applications and systems that create an unacceptable risk, such as government-run social scoring of the type used in China, are banned; 2) high-risk applications, such as a CV-scanning tool that ranks job applications, are subject to specific legal requirements; and 3) applications not explicitly banned or listed as high-risk are largely left unregulated." In December 2022, the EU member states, also known as the Council of the EU, approved an updated version of the AI Act after amendments and discussions. The AI Act as a draft law was passed in June 2023, a large milestone. The European Parliament then passed the AI Act on March 13, 2024, and it came into force on August 1, 2024. The AI Act is the first legislation in the EU

Director-General of UNESCO Audrey Azoulay speaks during the signing of an agreement with eight technology companies to build more ethical AI at the 2nd Global Forum on the Ethics of Artificial Intelligence in Kranj, Slovenia, on February 5, 2024. LUKA DAKSKOBLER/SOPA IMAGES/LIGHTROCKET VIA GETTY IMAGES to regulate AI systems, providing rules for the safe and trustworthy placing of products on the EU market with AI components.

The AI Act aims to ensure that AI systems placed and used in the European market are safe and respect existing legislation on fundamental rights and values of the EU, among which is the General Data Protection Regulation (GDPR). Article 22 of the GDPR is a general restriction on automated decision-making and profiling. It applies only when a decision is based solely on automated processing—including profiling—which produces legal effects or similarly significantly affects the data subject. Article 15 of the GDPR is linked explicitly to automated, individual decision-making and profiling that fall within the narrow scope of Article 22. These include the "existence" of automated decision-making, including profiling, "meaningful information about the logic involved," and "the significance and the envisaged consequences of such processing" for the individual. Given existing frameworks and regulations, Europe seeks to regulate the use of AI to protect the fundamental rights and safety of EU citizens. The EU also wants to establish itself as a global leader in AI and governance.

Organization for Economic Co-Operation and Development (OECD) Council Recommendation

The OECD Principles on Artificial Intelligence "promotes AI that is innovative and trustworthy and that respects human rights and democratic values." They were adopted in May 2019 by OECD member countries when they approved the OECD Council Recommendation on Artificial Intelligence. The OECD AI Principles are the first such principles signed up by governments. They include concrete recommendations for public policy and strategy, and their general scope ensures they can be applied to AI developments around the world. The OECD AI Principles are standards for technology that are both practical and flexible in terms of AI developments over time. In addition to the OECD's 36 member countries, Argentina, Brazil, Colombia, Costa Rica, Peru, and Romania have adopted the OECD principles, bringing the total to 42 countries. This is a substantial achievement in developing standards, working towards responsible stewardship, creating policies on deploying trustworthy AI systems, and fostering secure AI ecosystems.

The leading countries in advanced artificial intelligence, as highlighted in the Stanford University 2023 AI Index Report the governance of emerging technologies, especially dual-use ones like AI, is complex and tainted with challenges from countries and the private sector. However, there have been large strides in governing this dual-use technology. Efforts like the AI Bill of Rights in the U.S., UNESCO's global agreement on AI, the EU's AI Act, and the OECD's AI Principles are substantial efforts in developing and regulating the technology. The following section will outline the current approach in the U.S. to AI and national security.

The current approach to AI and security in the U.S.

The U.S. government invested over $4 billion in AI research and development in fiscal year 2023. Still, the U.S. is not the sole country committed to increasing federal spending on artificial intelligence. Regarding military funding of AI, an October 2021 report by the Center for Security and Emerging Technology estimated that yearly Chinese military spending on AI was "in the low billions of US dollars." According to the National Defense Industrial Association's magazine *National Defense*, this level of funding for AI is on par with the Pentagon's investments. Other countries leading in AI investment efforts include Israel, the United Kingdom, Canada, India, Japan, Germany, Singapore, and France. From both a national and global vantage point, it is clear that interest in artificial intelligence is growing quickly.

Current export control policies and AI

Export control laws and regulations restrict the transfer of goods and technology for entities outside the country of origin that could contribute to the military potential of international adversaries or advance a country's foreign policy goals. They are also enacted for economic and trade purposes. These restrictive efforts to possess technologies are often discussed in terms of national security considerations, dual-use activity, and protection of sensitive technology. Certain technologies, like artificial intelligence, may have applications for potential conventional weapons, intelligence collection, weapons of mass destruction, or terrorist applications, and provide other countries with a qualitative military or intelligence advantage (for example, Nvidia A100 chips used by an affiliated Chinese military body for training deep learning models). Notably, microprocessors and graphic processing units (GPUs) are needed to assist with machine learning, neural networks, and computer vision.

The U.S. is concerned with where and how certain technologies spread across borders, especially to strategic competitors. Therefore, the country has made several concerted efforts that touch on AI, including the Export Control Reform Act, the CHIPS and Science Act, and Bureau of Industry and Security restrictions. In 2018, the U.S. established the Export Control Reform Act. Section 1758 of the act authorizes the Bureau of Industry and Security (BIS) to develop appropriate controls on exporting, reexporting, or transferring emerging and foundational technologies essential to the country's national security. In 2022, the U.S. passed the CHIPS and Science Act. This act aims to bolster American manufacturing in semiconductor chips, supply chains, and national security. In 2024, the BIS, which controls the export of dual-use and less sensitive military items through the Export Administration Regulations, released an interim final rule on China's ability to obtain advanced computing chips.

Strengths

The main bottleneck issues concerning AI frontier models center on the costs and accessibility of GPUs. By enforcing export controls on certain GPUs and related resources, the U.S. Department of Commerce is rightly focused on the specialized hardware needed to build frontier models and who has access to these materials. Specifically, U.S. export controls have prevented AI computer chip designers like Nvidia and AMD from selling specific high-end chips and supercomputers to China. The U.S. also requires companies to obtain export licenses for other countries, including Saudi Arabia, the United Arab Emirates, and Vietnam, regarding advanced processors.

Weaknesses

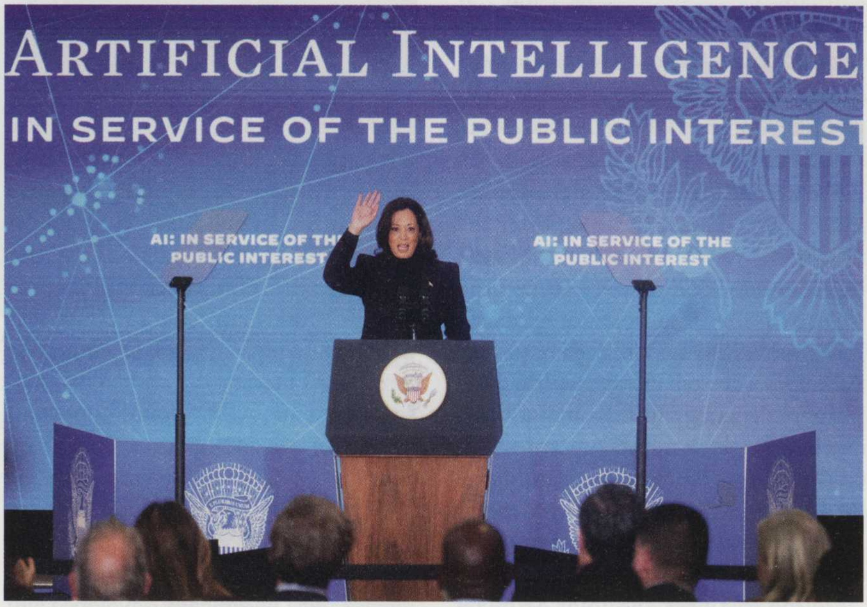
It is important also to note the weaknesses of the current U.S. approach to AI core components and export controls. In September 2022, Nvidia was banned from exporting the A100 and H100 to China without a license. As a result, the company then released the slower A800 and H800 to comply with the regulations and meet market needs, which could arguably be seen as a loophole (these chips were later banned in October 2023). Moreover, high-risk actors have obtained these advanced chips from companies in places other than the U.S., for example third-party sellers in India, Taiwan, and Singapore. The following section will discuss various potential AI and security policy solutions for the U.S.

How should the U.S. approach AI and security?: policy options

This section focuses on how the U.S. can approach the nexus of AI and security through improved export controls, increased AI literacy, and strengthened AI ethics across the country and its stakeholders.

Policy solution #1: export controls on AI

An alternative approach to export controls concerning AI and certain GPUs centers on robust risk assessment and security measures. An additional layer of security could be added for companies or actors seeking to rent or purchase cloud computing services on U.S. soil, so they undergo an extensive review. It has been reported that some actors may have circumvented the cloud computing ban by using servers containing restricted chips and cloud computing on U.S. soil (for example, ByteDance with Oracle). The U.S. Department of Commerce would do well to consider this loophole and require cloud providers to authenticate foreign customers' profiles and objectives. This effort could also prevent high-risk actors from using advanced chips to train frontier models from U.S.-based data centers. While this policy recommendation would add another step to an intensive process, the benefit of bolstering national security outweighs the burdens of bureaucracy.


ARTIFICIAL INTELLIGENCE
IN SERVICE OF THE PUBLIC INTEREST
AI: IN SERVICE OF THE PUBLIC INTEREST
AI: IN SERVICE OF THE PUBLIC INTEREST

U.S. Vice President Kamala Harris gives a speech on AI at the U.S. Embassy in London, on November 1, 2023, before attending the AI Safety Summit, the first global summit on the safe use of AI. PRESS ASSOCIATION VIA AP IMAGES

Policy solution #2: AI literacy

The widespread influence of AI can have significant effects across civil society and the military—from election interference to misinformation and disinformation.

Indeed, the impact of AI is a salient topic in the 2024 general election in the U.S. Moving forward, the U.S. should commit more national attention, financial resources, and training to support AI education across federal entities and civil society. Perhaps more importantly, the U.S. federal government should formalize an AI education strategy with timeline-specific goals, underscoring

Garry Kasparov in action on February 10, 1996, in Philadelphia during a match against the IBM supercomputer Deep Blue. Kasparov, a world champion, played six games against the AI chess program, winning three, losing one, and playing two to a draw. AL TIELEMANS/SPORTS ILLUSTRATED VIA GETTY IMAGES

both short-term and long-term goals. Specifically, U.S. policymakers must prioritize an AI-informed society, safeguarding transparency, and determining how to best equip the military.

Some progress has been made along these lines. For example, the Pentagon's 2020 AI Education Strategy emphasizes priority areas and skills required to accelerate AI adoption, from software and coding to data management and infrastructure. The strategy discusses how to build up AI capabilities, increase AI awareness for senior leaders, and bolster training on the responsible use of AI. While this is a good initial step, the strategy lacks the specifics of a timeline.

In the past year, the Joint Artificial Intelligence Center rolled out AI education pilot courses for thousands of Defense Department employees, ranging from education for general officers to coding boot camps. It would be advantageous to extend these initiatives beyond the Defense Department and to classify them into annual, five-year, and ten-year plans. The U.S. would greatly benefit from robust educational initiatives and AI investments across its departments—especially in Defense, Education, Homeland Security, and State—to reinforce the country's national security.

In March 2021, former Google CEO Eric Schmidt and former U.S. Deputy Secretary of Defense Bob Work, who led the National Security Commission on AI, wrote in the commission's final report: "America is not prepared to defend or compete in the AI era." However, this does not have to be the nations' future regarding AI. Decoding AI through AI literacy is a critical national security issue. AI infiltrates almost all aspects of our daily lives in the U.S.

Policy solution #3: AI ethics

The U.S. recognizes that artificial intelligence must be ethically developed and operated in a secure manner. On October 30, 2023, the U.S. made a productive step when President Biden signed the Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI. The Biden administration created this executive order to promote transparency, protect civil rights, and foster innovation while mitigating AI risks. The administration believed that only by doing this "only then can Americans trust AI to advance civil rights, civil liberties, equity, and justice for all." Notably, President-elect Donald Trump stated that he will repeal the executive order.

The Department of Defense published its recommendations for AI ethical principles, which would apply to combat and non-combat efforts, in 2020, as seen in the bullets below. At the time, former Secretary of Defense Mark Esper said, "The United States, together with our allies and partners, must accelerate the adoption of AI and lead in its national security applications to maintain our strategic position, prevail on future battlefields, and safeguard the rules-based international order." The principles focus on five critical areas: responsibility, equitability, traceability, reliability, and governance. At the core of these principles and behind the deployment and use of AI is the critical role of humans, who will exercise appropriate levels of judgment and work to minimize unintended bias in AI capabilities.

- **Responsible**. DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities.
- **Equitable**. The department will take deliberate steps to minimize unintended bias in AI capabilities.
- **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.
- **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 lifecycles.
- **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.

The prominence of AI and rising investments in the U.S. indicate the need for continued dialogue, research, and discussion of responsible deployment of the technology both within the country and globally.

Indeed, these policy recommendations are not limited to the U.S.; a multilateral approach among countries on AI framing and assessment could be more beneficial, especially as the technology advances. While AI has significant opportunities, there are concrete risks if AI goes unregulated at the national and global levels. Coupled together, these discussions at the national level and beyond can foster cooperation and robust dialogue on security.

Global perspectives on AI: competition, cooperation, and coordination

Open AI chief executive officer Sam Altman speaks at the Advancing Sustainable Development through Safe, Secure, and Trustworthy AI event at Grand Central Terminal on September 23, 2024, in New York. BRYAN R. SMITH/POOL/AFP VIA GETTY IMAGES

The National Defense Strategy notes that rapid technological advancements and the changing character of war will impact U.S. national security. It points out several disruptive technologies, including "advanced computing, big data analytics, AI, autonomy, robotics," which ensure the U.S. can fight and win the wars of the future. It's quite clear that the U.S. will not be the only country facing these new developments and threats; other key strategic competitors, such as China and Russia, are leveraging AI in multifaceted ways. Along with competing for AI leadership, countries must reckon with multiple global AI governance initiatives on AI norms and standards.

We would be remiss to ignore the pivotal role of non-state actors in the critical developments of technological spread, especially regarding AI. It's not that technology companies should have a seat at the table regarding emerging technologies—they are already vocal, active players in this arena along with states. Technology companies have varied opinions on AI issues, including some advocating for more regulations (for example, Sam Altman, chief executive officer of OpenAI, in a recent Congressional hearing on the need for AI regulations), while others note the repercussions of emerging technologies. The private sector has put continued effort into enacting national-level policies and initiatives, which shape state-private sector dynamics on AI. Governments, the private sector, and the general public all have a vested interest in AI and its societal implications.

The big picture

Artificial intelligence, especially generative AI, is often claimed as an emerging technology that will disrupt all facets of society. With the significant strides in AI, both opportunities and risks come to the forefront for states. The development, deployment, and governance of AI are widely discussed in international and regional forums, often in discussion with one another to foster progress.

AI can be considered a "moving target" and that could not ring truer now. Within international institutions, actors can collaborate to participate in the following efforts: defining and building consensus concerning emerging technologies, developing legislation or legal frameworks for their uses (and improper uses), and collaborating on monitoring and transparency. The dearth of international institutions and legal frameworks regarding emerging technologies increases risk, vulnerabilities, and the ambiguity of the technology's proper use. As AI advances, it will be important for governments to craft common standards, foster cooperation, and mitigate conflict.

Discussion questions

1. What are the interests of the U.S. in national and global AI governance?

2. Some experts argue that AI is the leading emerging technology to transform the conduct of warfare. Do you agree or disagree with this statement? Why?

3. There are competing interests in the global governance of artificial intelligence. How should countries cooperate on potential regulations of this rapidly developing technology?

4. What risks does AI pose to international security? What opportunities does AI offer to society?

5. Technology companies largely drive artificial intelligence with their advanced frontier models. What is the role of stakeholders such as the private sector in AI policies?

6. What risks does AI pose to personal security and privacy?

Suggested readings

The Oxford Handbook of AI Governance, edited by Justin B. Bullock, Yu-Che Chen, Johannes Himmelreich, Valerie M. Hudson, Anton Korinek, Matthew M. Young, and Baobao Zhang.

In "How Artificial Intelligence Can Strengthen Nuclear Stability," Jessica Cox and Heather Stanton highlight the ubiquitous nature of artificial intelligence both in military and civilian domains and globally. Using Michael Horowitz's framework, the authors define artificial intelligence as "systems that select and engage targets on their own" (e.g., computer-guided precision weapons) or "intelligent machines capable of cognitive judgments on par with humans." They argue that artificial intelligence can offer opportunities and risks, depending on its application.

In "Power to the People: How Open Technological Innovation Is Arming Tomorrow's Terrorists," Audrey Kurth Cronin focuses on how modern technologies have made their way to the public. The diffusion of modern technology (robotics, cyber weapons, 3D printing, autonomous systems, and artificial intelligence) to ordinary people has given them access to weapons of mass violence previously monopolized by the state. She argues that governments must develop countermeasures to preempt militants from co-opting innovations to catastrophic effect.

In "AI, the International Balance of Power, and National Security Strategy," Michael Horowitz, Gregory Allen, Edoardo Saravalle, Anthony Cho, Kara Frederick, and Paul Scharre discuss how artificial intelligence could transform nearly every aspect of national security. They consider AI applications for defense, intelligence, diplomacy, surveillance, cybersecurity, and economic tools of statecraft. Moreover, the authors underscore how the U.S. should anticipate and prepare accordingly for AI uses by competitors.

In *AI Superpowers,* Kai-Fu Lee, Lee explores how the development of AI may deliver huge changes sooner than expected, especially with the U.S.-China rivalry pushing each country to strive to become the technological superpower.

In *AI 2041,* Kai-Fu Lee and Chen Qiufan use short stories and technical analysis to explore how AI will impact the world by 2041.

In "The Potential Impact of Emerging Technologies on Democratic Representation: Evidence from a Field Experiment," Sarah Kreps and Douglas L. Kriner assess individuals' willingness to use AI-enabled technologies and their levels of trust in these capabilities. The authors look at AI-enabled technologies in different domains, such as armed drones, general surgery, police surveillance, self-driving cars, and social media content moderation. The article assesses public attitudes toward support and discusses the implications of integrating AI-enabled technologies in multiple settings.

In "Keeping Humans in the Loop Is Not Enough to Make AI Safe for Nuclear Weapons," Peter Rautenbach highlights the challenges of artificial intelligence and the need to keep the "human in the loop" along with technical solutions. The author underscores how the technology can be brittle and can break when used in unfamiliar territory on which it was not trained. While AI can perform certain functions at fast speeds, human operators can work to increase safety through critical thinking, support, and oversight of these systems.

Don't forget to vote!

Download a copy of the ballot questions from the Resources page at www.fpa.org/great_decisions

To access web links to these readings, as well as links to additional, shorter readings and suggested websites,

GO TO www.fpa.org/great_decisions

and click on the topic under Topic Resources, on the top of the page.