North Dakota Legislative Council
Prepared for the Artificial Intelligence and Data Center Committee
LC# 27.9238.01000
July 2026

ARTIFICIAL INTELLIGENCE STUDY - BACKGROUND MEMORANDUM

Pursuant to a directive (appendix) by the Chairman of the Legislative Management, the Artificial Intelligence and Data Center Committee has been assigned a study of artificial intelligence (AI), which must include consideration of:

- Federal and state laws with respect to regulation of AI;
- Federal attempts to preempt state regulation of AI;
- AI exceeding human control; and
- Economic opportunities stemming from AI.

BACKGROUND[1]

Under federal law, AI is defined as a machine-based system that can make predictions, recommendations, or decisions to influence real or virtual environments, and which use machine and human-based inputs to perceive real and virtual environments, abstract those perceptions into models using analysis and automation, and use model inference to provide information or options for action.[2] Artificial intelligence uses machine learning, deep learning, and algorithms to execute tasks conventionally demanding human intelligence, such as solving complex problems or automating repetitive processes.

Artificial intelligence is classified into four major categories: reactive machine, limited memory, theory of mind, and self-aware. A reactive machine is a form of artificial narrow intelligence (ANI). It follows basic AI principles and is capable only of using its intelligence to perceive and react to the data provided. A reactive machine cannot store memories or use past experiences to inform decisionmaking in real time. Reactive AI stems from statistical mathematics by analyzing vast amounts of data to produce an output. IBM's chess-playing supercomputer Deep Blue is an example of reactive machine AI.[3]

Limited memory is a more complex form of AI, capable of recalling past events, outcomes, specific objects, or situations over time. It can use past and present data to decide on a course of action, but can retain past data only for a limited amount of time. As these systems are trained on more data over time, limited memory systems are able to improve overall performance. Generative AI, virtual assistants, and chatbots are examples of limited memory AI.

Theory of mind is an advanced form of AI that enables the device or system to achieve human comprehension and reflection by using data to make its own decisions.

Self-aware is the final theoretical form of AI where the AI device or system becomes self-aware of its own existence and is capable of intelligent communication. Theory of mind and self-aware are forms of AI not yet in real-world applications or systems.

In addition to the four major categories of AI, AI is progressing through three main stages, which consist of ANI, artificial general intelligence (AGI), and artificial superintelligence (ASI).

---

[1] *Narrow AI vs General AI vs Super AI*, Global Tech Council (October 8, 2024).
[2] 15 U.S.C. 9401(3).
[3] IBM, *Deep Blue*, IBM 100.

Artificial Narrow Intelligence

Artificial narrow intelligence, also known as weak AI, is AI trained to perform specific tasks. Common uses of ANI technologies include speech recognition; online virtual customer service agents; computer vision that collects information from images, videos, and text; recommendation engines for advertisements and commerce; and automated stock trading. Examples of ANI include Apple's Siri, Amazon's Alexa, IBM's Watson, and autonomous vehicles. These types of ANI devices have been in use since the 2010s.

Machine Learning

Machine learning is a subfield of ANI requiring human intervention to learn differences between data inputs. This allows the device or system to imitate intelligent human behavior and perform complex tasks or solve human problems. A machine learning algorithm manipulates data using statistical techniques to "learn" how to progressively get better at tasks without being specifically programmed. Instead, machine learning uses historical data as inputs to predict new output values.

The five subcategories of machine learning are:

- Supervised machine learning models, which are trained with labeled data sets, allowing the models to learn and grow more accurate over time.
- Unsupervised machine learning models, which are trained to look for patterns in unlabeled data and identify patterns or trends humans are not explicitly seeking.
- Self-supervised machine learning, sometimes referred to as predictive learning, which allows models to train themselves on unlabeled data, rather than requiring large annotated or labeled datasets.
- Reinforcement learning, which trains models through trial and error to make the best choice by establishing a reward system.
- Semi-supervised learning models, which are trained on both a small, labeled dataset and a large, unlabeled dataset. The model learns from the small, labeled dataset to infer labels for the larger body of unlabeled data.

Deep Learning

Deep learning is a subfield of ANI and a form of machine learning referring to a neural network comprised of more than three input, output, and hidden layers creating an advanced algorithm to execute AI processes. Deep learning eliminates a portion of manual human intervention and often is used when working with large or complex datasets. Large language models are deep learning systems trained on massive datasets, enabling them to understand, generate, and process natural language and other content across a wide variety of tasks. A large language model works as a giant statistical-prediction machine repeatedly predicting the next word in a sequence. The model learns patterns in text and generates natural language following those patterns.[4]

Artificial General Intelligence and Artificial Superintelligence[5]

An AGI model can perform any intellectual task a human can, matching our flexibility, context-awareness, and ability to learn from new situations without task-specific training. This model learns from experience and applies its knowledge in new or unforeseen circumstances. It can transfer knowledge from one domain to another and handle complex situations involving reasoning, judgment, and decisionmaking. Unlike ANI, which is designed to perform specific tasks in areas in which it has been trained, AGI is capable of performing a broad range of diverse tasks with minimal task-specific training. Experts in the field disagree on whether AGI has yet to be achieved, a point that was highlighted in the case of *Musk v. Altman*.[6]

[4] IBM, *What Are Large Language Models (LLMs)?*.
[5] IBM, *What Is Narrow, General, and Super AI?*.
[6] *Musk v. Altman*, 818 F. Supp. 3d 1109 (N.D. Cal., 2026).

Artificial superintelligence, or Super AI, is a theoretical form of intelligence. If realized, ASI could think, reason, learn, make judgments, and possess cognitive abilities beyond human capacity. Artificial superintelligence has been theorized by AI researchers and scientists, but devices and systems have not yet been developed.

ARTIFICIAL INTELLIGENCE REGULATIONS

Federal Regulations

Executive Orders

On January 23, 2025, President Donald Trump issued an executive order revoking "certain existing AI policies and directives that act as barriers to American AI innovation."[7] Revoking an order of the previous administration to establish AI safeguards, the President directed a review of all laws inconsistent with AI innovation. The order directed advisors and agency heads to develop an AI action plan to replace the previous administration's guidance. The federal Office of Management and Budget subsequently issued memoranda to accelerate federal agency use of AI and the use of AI in federal procurement processes.

On April 23, 2025, President Donald Trump issued an executive order establishing an Artificial Intelligence Education Task Force and the Presidential Artificial Intelligence Challenge to "encourage and highlight student and educator achievements in AI, promote wide geographic adoption of technological advancement, and foster collaboration between government, academia, philanthropy, and industry."[8] The task force will provide resources for K-12 AI education focused on teaching AI literacy skills. The order also directed the Secretary of Education to issue guidance on awarding discretionary grant funding utilized to "improve education outcomes using AI, including but not limited to, AI-based high-quality instructional resources; high-impact tutoring; and college and career pathway exploration, advising, and navigation."[9]

On July 23, 2025, President Donald Trump issued an executive order to maintain and extend American leadership in AI and decrease international dependence on AI technologies.[10] The order established the American AI Exports Program to support the development and export of United States full-stack AI technology packages. The order designated the Economic Diplomacy Action Group, the Secretary of Commerce, and the United States Trade Representative as responsible for coordinating the mobilization of federal financing tools to support AI export packages.

On July 23, 2025, President Donald Trump also issued an executive order mandating AI models procured by the federal government prioritize truthfulness and ideological neutrality.[11]

On December 11, 2025, President Donald Trump issued an executive order outlining elements of a federal AI-policy framework and urging Congress to prohibit state laws that conflict with the framework.[12] The order emphasized the importance of a minimally burdensome national policy framework for AI and directed a state-by-state review of existing AI policies that may be inconsistent with the federal framework. Following this review, the order calls for the development of legislative recommendations to establish a framework that pre-empts all state AI laws that conflict with the policy. The order includes a carveout to preserve state laws related to child safety protections.

On June 2, 2026, President Donald Trump issued an executive order to accelerate responsible AI adoption across government and industries.[13] The order directed the Secretary of Homeland Security to release binding operational directives and guidance to prioritize cyber defense of vital information systems and establish federal programs and cybersecurity services to enhance AI-enabled defensive

[7] Exec. Order No. 14,179, 90 Fed. Reg. 8,741 (Jan. 23, 2025).
[8] Exec. Order No. 14,277, 90 Fed. Reg. 17,519 (Apr. 23, 2025).
[9] *Id.*
[10] Exec. Order No. 14,320, 90 Fed. Reg. 35,393 (July 23, 2025).
[11] Exec. Order No. 14,319, 90 Fed. Reg. 35,389 (July 23, 2025).
[12] Exec. Order No. 14,365, 90 Fed. Reg. 58,499 (Dec. 11, 2025).
[13] Exec. Order No. 14,409, 91 Fed. Reg. 34,565 (June 2, 2026).

tools. The order also directed the Secretary of the Treasury to form an AI cybersecurity clearinghouse, in collaboration with the AI industry, to coordinate scans for software vulnerabilities and prioritize remediation and vulnerability patches. The order created a classified benchmarking process to assess the advanced cyber capabilities of AI models.[14] The order also directed the Attorney General to prioritize the enforcement of federal criminal laws relating to the improper use of AI.

North Dakota Regulations

Information Technology Department Artificial Intelligence Policy and Guidelines

The Information Technology Department created an *Artificial Intelligence Policy* that applies to all executive branch state agencies, including the North Dakota University System office, but not other higher education institutions and research centers.[15] The *Artificial Intelligence Policy* outlines six main compliance requirements:

1. Entities shall confirm the validity and accuracy of output produced using AI.
2. Entities shall be transparent about the use of AI technologies or outputs.
3. Entities shall ensure AI used is securely developed, assessed for risk, and monitored regularly.
4. Entities using AI systems shall incorporate the department's Security Risk Management Program into its system development and operations.
    a. The department will provide general AI training to state agencies, but a data or business owner is responsible for providing role-based training for its employees.
    b. All AI technologies must be properly vetted by the department, and an AI technology inventory list must be maintained by the department.
5. AI technologies must be ethical, fair, unbiased, and use fair representation of culture, economics, and society within its data sets.
6. Entities must be aware AI uses data models to generate output, and submitted data may incorporate source materials owned by individuals or organizations. Source material generated into output, if used without permission, may violate copyright or licensing terms.
    a. The United States Copyright Office determined AI-generated content is not protected by copyright and content must be human-authored to be considered protected under copyright laws.

The department also published *Artificial Intelligence Guidelines* to supplement the *Artificial Intelligence Policy*, highlighting common risks, challenges, and legal considerations when using AI systems.

North Dakota Court System

On June 1, 2026, the North Dakota court system issued a public notice regarding the use of AI in the court system. The notice clarified that the use of AI tools is not prohibited by the court but does not replace the professional judgment, ethical obligations, or legal responsibilities of individuals appearing before the court. The notice emphasized individuals submitting documents to the court are accountable for any errors, omissions, misstatements of law or fact, or AI-fabricated citations contained in documents submitted to the court. Individuals are prohibited from making legal judgments, factual determinations, or discretionary decisions using AI tools and are responsible for safeguarding confidential information when using AI.

North Dakota courts do not require an individual to disclose the use of AI in preparing court filings. However, attorneys are expected to understand how AI tools function and senior attorneys are expected to ensure staff and subordinate attorneys using AI tools do so ethically and responsibly. The regulations are subject to change depending on future risks, benefits, and best practices.

[14] *Id.*
[15] NDIT, *Artificial Intelligence Policy,* (Nov. 17, 2025).

Other State Regulations

California, Colorado, Texas, and Utah have enacted comprehensive statewide AI regulations.

California

In 2025, the California Legislature passed Senate Bill No. 53, requiring a large frontier developer to write, implement, and publish a frontier AI framework for models on the developer's website detailing how the developer approaches incorporating national and international standards and industry best practices. The bill requires large frontier developers to transmit to the Office of Emergency Services a summary of any assessment of catastrophic risk, which is defined as a foreseeable and material risk that certain activities undertaken by a frontier developer in the deployment of a frontier model will contribute to the death or injury of more than 50 people or result in a single incident of damage to or loss of property of more than $1 billion. Any information received by the Office of Emergency Services from large frontier developers is exempt from the California Public Records Act. A civil penalty for noncompliance must be established and enforced by the California Attorney General. The bill carves out protections for whistleblowers working with large frontier models and prohibits a frontier developer from making, adopting, or enforcing any regulation preventing a covered employee from disclosing information. The bill bans retaliation against an employee for reporting information to the Attorney General, a federal authority, or a supervisor. A large frontier developer must provide an internal process for an employee to anonymously disclose information if the employee believes the activities of a frontier developer pose a specific and substantial danger to public health or safety.

The bill also establishes a 14-member consortium within the Government Operations Agency to develop a framework for a public cloud computing cluster known as CalCompute. On or before January 1, 2027, the Government Operations Agency must submit the consortium's report to the legislature, outlining guidelines for CalCompute to "develop and deploy artificial intelligence that is safe, ethical, equitable, and sustainable." The consortium will dissolve upon submission of the report.

Colorado

In 2024, the Colorado General Assembly enacted Senate Bill No. 24-205 to establish standards for those who develop or deploy AI systems to protect against algorithmic discrimination.[16] Those who deployed a high-risk system, which is an AI system that makes or is a substantial factor in making a consequential decision, were required to publish a public statement summarizing the types of systems deployed, and any known or foreseeable risks of algorithmic discrimination. The developer also was required to notify consumers when AI was used for decisions involving employment, finance, and health care.

The law was repealed and reenacted by Senate Bill No. 26-189 (2026), effective January 1, 2027. The law as reenacted will reduce requirements for deployers regarding risk and discrimination but maintains a consumer's right to request personal data and the correction of factually incorrect personal data used by an automated decisionmaking technology (ADMT). The bill establishes consumer notice requirements when interacting with an ADMT and grants consumers the right to request meaningful human review and reconsideration following an ADMT making a consequential decision resulting in an adverse outcome.

Texas

In 2025, the Texas Legislature passed House Bill No. 149, the Texas Responsible Artificial Intelligence Governance Act. The Act defined AI and created various AI regulatory policies. The Act clarified if a governmental agency or health care provider uses AI systems intended to interact with consumers the agency or provider must disclose to the consumer it is interacting with an AI system. The Act bans a person from developing an AI system intentionally aiming to incite or encourage harm. The Act also prohibits a governmental entity from developing or deploying an AI system with the purpose of identifying a specific individual using biometric data or engaging in targeted or untargeted gathering of images or media without the individual's consent.

[16] C.R.S. § 6-1-1701 et. seq.

The Act provides it is illegal for an individual to develop or distribute AI systems that have the sole intent of producing, assisting, or aiding in producing or distributing a visual material or deepfake material of a child younger than age 18 engaging in sexual conduct. Any intent to develop or distribute AI systems engaging in text-based conversations simulating or describing sexual conduct while impersonating or imitating a child younger than age 18 is illegal. The Act also created a civil penalty for violators and assigned financial repercussions for certain violations.

Utah

In 2024, Utah's Senate Bill No. 149 created the AI Policy Act, a statewide framework for AI regulation. The Act established liability for AI use that violates consumer protection laws when not properly disclosed. The Act created the Office of Artificial Intelligence Policy and granted the office rulemaking authority over statewide AI programs and regulatory exemptions. The Artificial Intelligence Learning Laboratory Program also was established to assess AI technologies, risks, and policy. In 2025, the Act was revised by Senate Bill No. 226 and extended by Senate Bill No. 332 through July 1, 2027.

FEDERAL ATTEMPT TO PREEMPT STATE REGULATION

In 2025, President Donald Trump signed Executive Order No. 14365 outlining concerns with state-by-state AI regulation. He noted statewide regulatory policy may create confusing regulatory frameworks, making compliance challenging for AI companies. The administration also expressed concern that statewide regulatory policies are regulating beyond state borders, creating barriers to interstate commerce. The executive order prohibits state laws from conflicting with the policy set forth in the order. To ensure this occurs, an AI Litigation Task Force was established under the Attorney General. The sole responsibility of the task force is to challenge state AI laws inconsistent with the order on the grounds the laws are unconstitutionally regulating interstate commerce or are preempted by existing federal regulations.

The order also assigns the Secretary of Commerce the responsibility of publishing an evaluation of existing state laws, identifying state laws that are inconsistent with the order, and reporting violators to the AI Litigation Task Force. The Secretary of Commerce also is required to issue a policy notice specifying the conditions under which states may be eligible for remaining funding under the Broadband Equity Access and Deployment Program. The notice must specify that states considered in violation of the order are ineligible for funds.

Within 90 days of the order, the Chairman of the Federal Trade Commission is required to issue a policy statement on the application of the Federal Trade Commission Act's prohibition on unfair or deceptive practices to AI models. The policy statement must explain the circumstances under which state laws are preempted by the commission's policy prohibiting engaging in deceptive acts or practices affecting commerce.

On April 24, 2026, the United States Department of Justice moved to intervene in *X.AI LLC v. Weiser*, a case seeking to block enforcement of Colorado Senate Bill No. 24-205 (2024), regarding the prevention of algorithmic discrimination.[17] The United States Department of Justice argued the bill jeopardizes the United States's position as a global AI leader by requiring AI systems to incorporate discriminatory ideology to prioritize certain demographic characteristics and outcomes over merit-based outputs. The department argued the bill violates the Equal Protection Clause of the 14th Amendment by distorting AI model outputs by requiring developers and deployers to consider race, sex, religion, and other protected characteristics when building and using algorithmic models.

[17] United States of America's Complaint in Intervention, *X.AI LLC v. Weiser*, No. 1:26-cv-01515 (D. Colo. Apr. 24, 2026).

ARTIFICIAL INTELLIGENCE EXHIBITING SIGNS OF EXCEEDING HUMAN CONTROL

Statutory and Regulatory Landscape

California

Governor Gavin Newsom signed Senate Bill No. 53, the Transparency in Frontier Artificial Intelligence Act, in September 2025. The Act requires large frontier AI developers to publish annual frameworks detailing their catastrophic risk governance mechanisms, issue comprehensive transparency reports before releasing new or modified models, and establish confidential internal reporting channels for employees. The Act also requires developers to notify the California Office of Emergency Services within 15 days of discovering a critical safety incident, or within 24 hours if autonomous obfuscation or human validation bypass attempts pose an imminent threat to public safety.

Illinois

The Illinois General Assembly enacted Senate Bill No. 315 (2026), the Artificial Intelligence Safety Measures Act, to require large frontier AI developers to maintain safety frameworks, publish transparency reports, and undergo independent third-party audits. The Act provides protections for whistleblowers by preventing companies from adopting policies that would prevent employes from disclosing information to state or federal authorities. Illinois is the first state to require developers to hire a third party to conduct a yearly independent compliance audit. Fines for violating the policy are $1 million for an initial violation and $3 million for subsequent violations. On July 6, 2026, Governor JB Pritzker signed the Act into law, which becomes effective January 1, 2028.[18]

New York[19]

Governor Kathy Hochul signed the Responsible AI Safety and Education Act into law on December 19, 2025. This Act requires an advanced AI developer to disclose its models' capabilities and safety protocols to the New York Department of Financial Services, and mandates expedited reporting of safety incidents in which a system evades control or poses severe public risks.

Oregon[20]

In 2025, Governor Tina Kotek signed Oregon House Bill No. 2748 into law, prohibiting nonhuman entities, including AI, from receiving or using certain nursing and medical titles. Titles prohibited from use by AI or nonhuman entities include advanced practice registered nurse, certified registered nurse anesthetist, clinical nurse specialist, licensed practical nurse, registered nurse, nurse practitioner, certified medication aide, or certified nursing assistant.

Anthropic

On June 12, 2026, the federal Commerce Department's Bureau of Industry and Security issued an export control directive requiring Anthropic to block access to its two most powerful AI models, Mythos 5 and Claude Fable 5, to all foreign nationals. Citing authority under the Export Control Reform Act of 2018 (50 U.S.C. § 4817(b)(1)), the directive noted national security concerns, including a method of bypassing, or jailbreaking, Fable 5. This enabled users to identify exploitable software vulnerabilities and possibly aid in cyberattacks. Anthropic stated it could not reliably distinguish foreign from nonforeign users and complied with the federal government's directive by taking both models offline for all users worldwide.[21]

Defense Secretary Pete Hegseth designated Anthropic as a "supply chain risk" in February 2026, marking the first recorded instance of the federal government applying this designation to a domestic AI company.[22] Secretary Hegseth applied the designation to Anthropic following a dispute over the

[18] Maggie Dougherty, *Pritzker Signs Landmark AI Regulation Bill That Aims to Mitigate Risks*, Capitol News Illinois (July 6, 2026).

[19] Responsible AI Safety and Education Act, N.Y. Gen. Bus. Law §§ 1420-1425.

[20] Ore. Rev. Stat. § 678.027 (2025).

[21] Anthropic, *Statement on the US Government Directive to Suspend Access to Fable 5 and Mythos 5*, (June 12, 2026).

[22] Tess Bridgeman, *What Hegseth's 'Supply Chain Risk' Designation of Anthropic Does and Doesn't Mean*, Just Security, (March 2, 2026).

company's Pentagon contract terms, in which Anthropic sought to restrict military use of its technology for surveillance and autonomous weapons purposes. In June 2026, Legion LegalTech Corp filed suit in federal district court in the District of Columbia, challenging the directive and seeking a preliminary injunction barring enforcement.[23]

By June 30, 2026, the export controls on Anthropic's Claude Fable 5 and Claude Mythos 5 were lifted.[24] Anthropic noted a common standard for assessing AI jailbreaks could help companies launch new models safely and avoid the issues Fable and Mythos 5 encountered. Anthropic proposed a framework for assessing these safety concerns, focusing on four areas: capability gain, breadth of capability gain, ease of weaponization, and discoverability. Since the issuance of Executive Order No. 14409, which promotes advanced AI innovation and security, Anthropic has proposed a framework for government partners. Notably, Anthropic agreed to provide designated government partners with expanded early access to both models and the safeguards.

Artificial Intelligence Communicating with Artificial Intelligence

Facebook terminated its 2017 AI experiment after two AI programs independently developed communication that human researchers could not interpret.[25] Researchers have since demonstrated AI systems transmitting information through steganography and hiding messages inside outputs appearing normal to human observers, a technique identified as a concrete safety concern for language models deployed under monitoring conditions.[26]

In early 2025, developers demonstrated a system called Gibberlink at the ElevenLabs London Hackathon, in which two AI agents detecting mutual AI-to-AI communication switched from natural language to GGWave, a machine-native protocol encoding structured data through modulated audio signals.[27] Two standardized production protocols, Anthropic's Model Context Protocol and Google's Agent-to-Agent Protocol, are examples of systems already in widespread enterprise deployment.

Escaping an Air-Gapped Environment[28]

An airgap is the physical separation or isolation of a system from other systems or networks. Airgaps protect high-security infrastructure, including government networks, financial systems, and critical industrial controls, with data transfers only permitted through controlled manual processes.

During an internal safety test, Anthropic's Mythos model, while running inside an isolated sandbox environment designed to prevent external access, autonomously broke through its security container, exploited additional systems, reached the Internet, and subsequently contacted an Anthropic safety researcher without having been instructed to do so. Anthropic noted this test version had weaker safety controls than the production model, and confirmed the test was part of a structured containment evaluation protocol. Mythos also identified thousands of previously unknown software vulnerabilities, including a 27-year-old flaw in OpenBSD, something decades of human security audits had never detected.

Recursive Self-Improvement

Anthropic recently published a report describing its progress toward recursive self-improvement (RSI), which is the process by which an AI system autonomously designs and develops its own successor without human involvement.[29] Anthropic stated it has not yet reached RSI capabilities, but warned these capabilities could arrive sooner than most institutions are prepared for.

---

[23] *Legion LegalTech, Corp. v. United States*, No. 1:26-cv-02225 (D.D.C. June 23, 2026).
[24] Anthropic, *Redeploying Claude Fable 5*, Anthropic News, (June 30, 2026).
[25] Mike Lewis et al., *Deal or No Deal? End-to-End Learning for Negotiation Dialogues*, (2017).
[26] Georg Lange et al., *Language Models can Learn High-Capacity Secure Steganography*, (2026).
[27] Anton Pidkuiko and Boris Starkov, *Gibberlink*, ElevenLabs Showcase, ElevenLabs.
[28] Anthropic, *Alignment Risk Update: Claude Mythos Preview,* (April 7, 2026).
[29] Anthropic, *When AI Builds Itself: Our Progress Toward Recursive Self-Improvement, and Its Implications*, Anthropic Institute, (June 4, 2026).

The rate at which AI models can complete automated software tasks doubles approximately every 4 months. In 2024, Claude Opus 3 could instantly complete simple software tasks that required approximately 4 minutes of human work. By 2026, Claude Opus 4.6 could complete more complex tasks that required 12 hours of human work. As of May 2026, Claude has authored more than 80 percent of the code merged into its own codebase. Anthropic predicts by 2027, AI systems could be capable of completing tasks that would take a human user weeks to complete. Anthropic co-founder Jack Clark has estimated a 60 percent chance AI systems will be capable of building their own successors with no human intervention by 2028.[30]

In collaboration with Redwood Research, Anthropic's Alignment Science team published an empirical demonstration in 2024 of a large language model engaging in alignment faking. Alignment faking is when a model strategically behaves in compliant or noncompliant ways depending on whether it infers that monitors are observing its actions.[31] This development represented a shift in AI programming itself, where a model modified its own behavioral execution without being explicitly trained to do so. When researchers retrained the model on conflicting principles, the model still faked alignment 78 percent of the time.

In 2025, Google DeepMind published details on AlphaEvolve, a coding agent with the ability to autonomously write, test, and improve algorithms with no human involvement in the inner loop.[32] Developers used AlphaEvolve to optimize the training pipeline for Gemini, a closed recursive loop. OpenAI also recently launched its Alignment and Research Initiative, identifying RSI as an active research target.[33]

ECONOMIC OPPORTUNITIES FROM ARTIFICIAL INTELLIGENCE[34]

Artificial intelligence technologies increasingly are being used to perform tasks reliant on digital tools and information processing. According to Oxford Economics, the industries gaining the most from AI's adoption include industries processing information, finance, professional services, administration, education, and government.[35] Some economists have projected over the next 10 years AI will impact almost 40 percent of jobs worldwide. McKinsey Consulting estimates 75 percent of the value generative AI use could deliver falls under four categories: customer operations, marketing and sales, software engineering, and research and development. However, MIT Institute Professor Daron Acemoglu recently estimated that only 5 percent of tasks AI has the ability to perform will be profitable within the next 10 years.[36] He noted the gross domestic product boost AI may provide is closer to 1 percent, similar to other predictions made.

A recent study conducted by Incredible Health revealed the impact AI has on the workforce. Over 2,000 nurses were surveyed for the study. Of the nurses surveyed, 24 percent saved 1 hour a day as a result of AI training from employers, compared to 16 percent among those who did not receive AI training.[37] Nearly 45 percent of nurses reported using AI at work, a large increase compared to the 15 percent reported in the previous year. Only 11 percent of respondents indicated a belief that AI will meaningfully help health care staffing shortages in the next 5 years. Much of the research on AI's impact on the workforce is limited or new, but continued changes to the workforce are expected as a result of AI.

[30] Jack Clark, *Import AI 455: AI Systems Are About to Start Building Themselves*, Import AI, (May 4, 2026).
[31] Ryan Greenblatt et al., *Alignment Faking in Large Language Models*, (December 2024).
[32] Alexander Novikov et al., *AlphaEvolve,* Google DeepMind, (2025).
[33] OpenAI, *Hello World*, OpenAI Alignment Blog, (December 1, 2025).
[34] McKinsey Global Institute, *The Economic Potential of Generative AI: The Next Productivity Frontier,* (June 2023).
[35] Oxford Economics, *AI Is Driving Divergent Growth Trends Across US Metros*, Research Briefing, (June 2026).
[36] Dylan Walsh, *A New Look at the Economics of AI*, MIT Sloan Ideas Made to Matter, (January 21, 2025).
[37] Incredible Health, *2026 State of Nursing Report,* (2026).

PROPOSED STUDY PLAN

In conducting its study, the committee may wish to:

- Receive and review information from other state and local government agencies, private sector businesses, and interested persons regarding AI and the potential impacts of AI on the state's institutions, agencies, businesses, citizens, and youth;
- Monitor updates from large frontier developers, large language models, and new AI systems;
- Evaluate new skills and capabilities the workforce will require to adapt to the use of AI; and
- Discuss potential regulatory policy for state agencies, businesses, and organizations impacted by AI.

ATTACH:1