2025 TECH TRENDS REPORT • 18TH EDITION

ARTIFICIAL INTELLIGENCE

FTSG

Amy Webb
Chief Executive Officer

Sam Jordan
Technology & Computing Lead

AI's bleeding edge now changes by the hour, not year. What next?

AI is moving at breakneck speed, reshaping industries, workflows, and everyday life faster than we can document. On the day we wrote this, people were still breathlessly marveling at China's DeepSeek, which achieved OpenAI's top-tier performance with a fraction of the usual price tag and computing power—challenging everything we thought we knew about what it takes to build advanced AI. Hours later, researchers at Stanford and the University of Washington debuted yet another new model, s1, which outperformed both DeepSeek's R1 and OpenAI's o1 reasoning models using even fewer resources.

That's the nature of AI right now: What's bleeding-edge today might be old news ... *later today.*

Here's what we know for certain. Last year, OpenAI CEO Sam Altman met with sovereign wealth fund managers and investors, hoping to raise up to $7 trillion for an AI chip company. In January 2025, Stargate, a newly formed joint venture between OpenAI, Oracle, and SoftBank, said it would raise $500 billion for chips, AI data centers, and their massive power requirements. Not to be outdone, Microsoft, Meta, and Google have each announced plans to invest hundreds of billions of dollars in AI infrastructure. But if now anyone can replicate a multimillion dollar model with only modest resources, won't AI models quickly become commoditized? If so, this would pressure Big Tech to move very fast, building and scaling ever-advancing AI systems in order to stay competitive in the market.

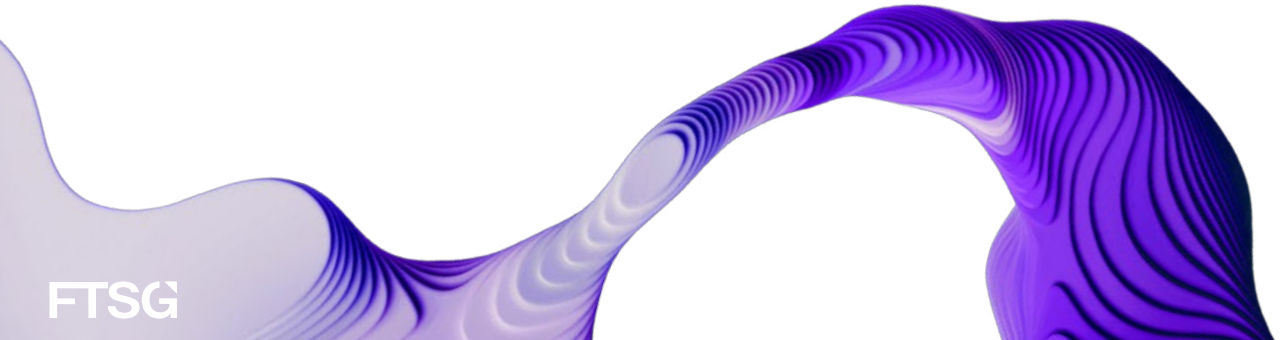
In the race to win AI, critical evaluation has become a casualty of speed. We meet regularly with the research teams building SOTA models, heads of frontier labs working to advance AI, and executives at the big tech giants. While we are certainly excited about the incredible technological progress being made, there is the practical reality of organizational readiness. The makers of AI systems—and the professional service firms promising overnight transformation—are operating in a reality far removed from everyday organizations.

What we've observed in the past year advising CEOs and their management teams on AI strategy and implementation is that regardless of AI's tantalizing developments, most organizations face substantial technical debt in data standardization and maintenance, creating operational friction in deployment. They are also struggling with the basics of change management, which is often deprioritized (or forgotten entirely) ahead of implementation. As a result, we are seeing new strategic risks for organizations that overindex on technological readiness without addressing fundamental operational and cultural barriers to deployment.

Breakthrough advancements have also accelerated the AI race between the US and China, intensifying it into a full-blown geopolitical contest, with both nations leveraging technology as a tool for global influence. To wit: during Donald Trump's globally televised inauguration, execs from America's biggest technology companies sat directly behind him—while his cabinet appointees and family members sat in rows farther back. This US-China rivalry is forcing allies to take sides, escalating tensions and fueling concerns over national security, supply chains, and technological sovereignty. What emerges is a fragmented AI landscape, dominated by a Digital Cold War that threatens to reshape global alliances and economic power structures.

How—*exactly*—will AI reshape our world in the coming months? The honest answer is: nobody can know. At this stage, avoiding costly mistakes, and smart planning for the future, matters more than predicting exact outcomes. Leaders need a strategic compass, not a crystal ball.

That's the purpose of this trend report: to highlight emerging AI trends and use cases so you can plan for multiple possibilities. Because in an AI landscape moving at warp speed, strategic clarity is your competitive edge.



Expect a continued frenzy of activity as AI companies compete for market share, though investments and policies will concentrate influence among several key players.

1

Harder, better, faster, stronger

DeepSeek's R1, and s1 from Stanford and the University of Washington achieved strong reasoning capabilities while remaining cost-efficient, challenging the convention that progress requires ever-larger models and raising questions about future AI scalability.

2

Your AI now has eyes and ears

Recent advancements in multimodal AI, like Google Gemini Live and OpenAI's Sora, are quickly transforming how machines process and generate text, audio, and video, unlocking new possibilities for richer and more interactive AI experiences.

3

Learning how to think

AI performance is improving, elevating models like OpenAI's o1 and Google's Gemini 2.0 Flash Thinking Mode from mere information engines to thought partners. In late 2024, OpenAI's o3 scored 85% on the ARC-AGI benchmark, matching the average human score.

4

From assistance to autonomy

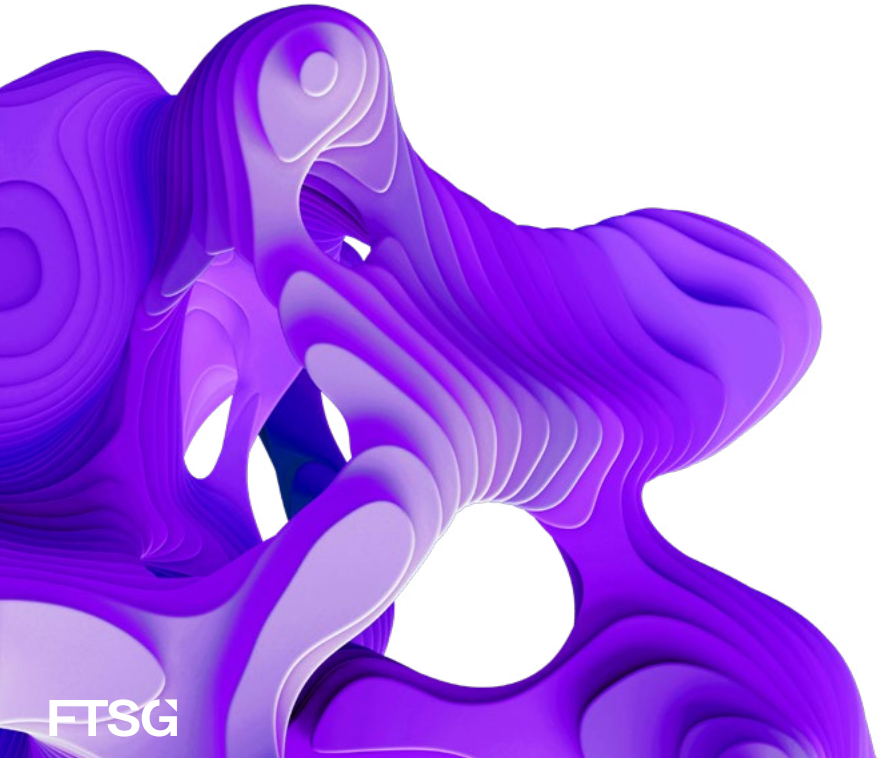
Agentic AI is evolving from supporting tasks to autonomously reasoning and taking action across workflows. This year, AI agents will not only assist but also execute complex processes, transforming industries with greater efficiency and automation.

5

US and China race ahead

The US and China are locked in a high-stakes competition for AI dominance, shaping the future of technology and global power. As both nations invest in AI research, infrastructure, and regulation, they are redefining innovation, security, and economic influence.

Artificial intelligence will fundamentally rewire dynamics and competitive frontiers in 2025.



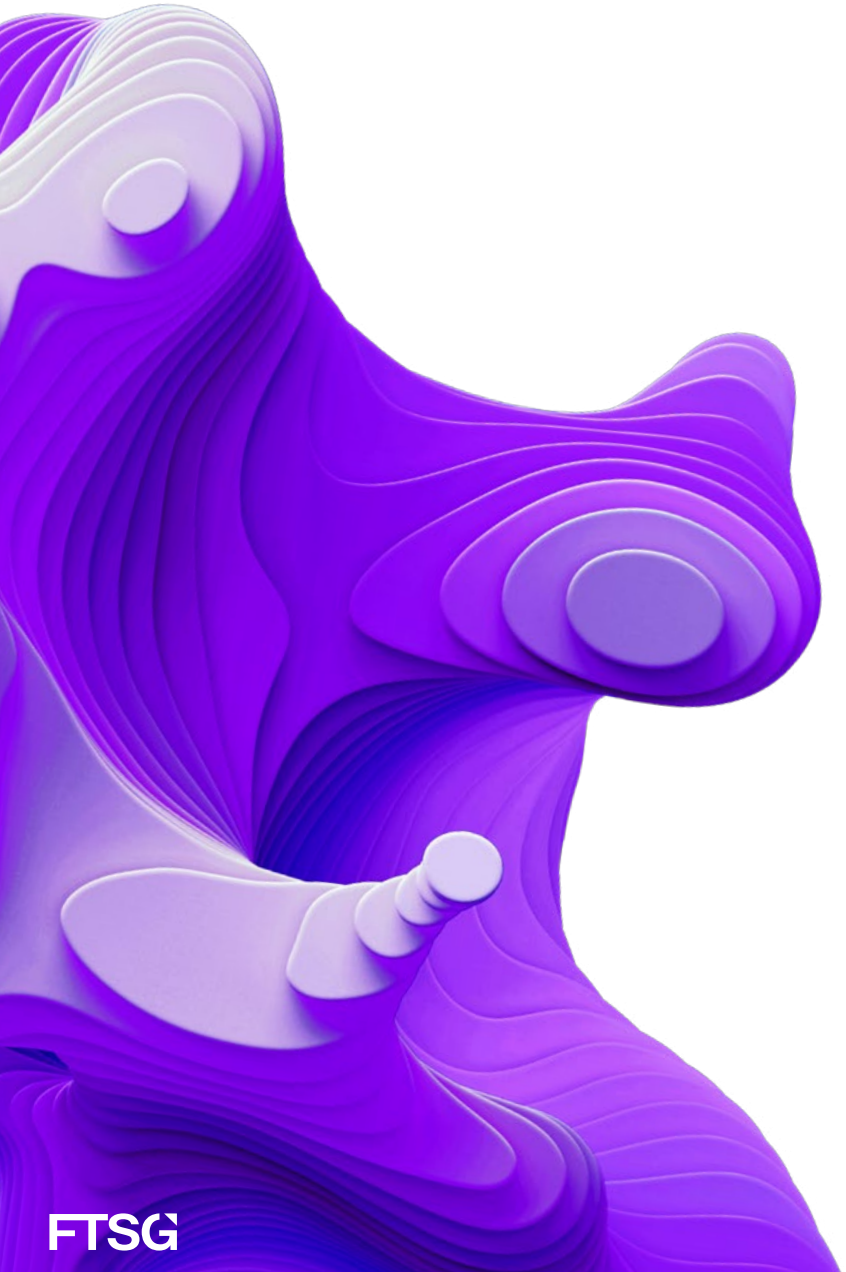
It's no secret that AI's landscape has transformed dramatically since we wrote the State of Play section last year. GPT-4 set early benchmarks with its multimodal capabilities and professional-level performance, but since then, Google DeepMind's Gemini Ultra raised the bar further, exceeding GPT-4 on most benchmarks.

The field has split between proprietary and open-source approaches. Meta's release of Llama 2 sparked an open-source revolution, with developers rapidly fine-tuning variants that rival larger commercial models. This success prompted even OpenAI's CEO Sam Altman to admit that the company might have been "on the wrong side of history" regarding closed systems.

Cloud partnerships have become crucial. Microsoft bet big on OpenAI ($10B), while Amazon and Google split their support for Anthropic ($4B and $2B respectively). These alliances provide AI companies with massive compute power while securing cloud providers' positions in the AI race. However, this concentration of resources raises concerns about market consolidation.

China has emerged as a formidable AI power. Baidu's Ernie 4.0 claims GPT-4-level performance, while Alibaba released more than 100 open-source models under Qwen 2.5. ByteDance's Doubao chatbot gained significant market share. Also making huge strides are Zhipu AI, MiniMax, Baichuan Intelligence, Moonshot, StepFun, and 01.AI—collectively known as the country's "Six Little Tigers." Despite US chip restrictions, Chinese firms are adapting with domestic alternatives like Huawei's Ascend and Baidu's Kunlun chips.

Investment has exploded, with more than $22 billion flowing to generative AI startups last year alone, representing nearly half of all AI funding. Traditional VCs are competing with tech giants throwing billions at AI, driving valuations skyward.



We see six macro themes emerging:

1. **Big Tech will continue to dominate funding, often through strategic partnerships**
2. **Valuations are reaching dot-com era levels, raising legitimate bubble concerns**
3. **Traditional tech sectors are seeing funding dry up as AI sucks the oxygen out of the room**
4. **Investment is flowing directly to cloud providers for compute power**
5. **Competition is intensifying between proprietary and open-source models**
6. **China is rapidly closing the gap with Western AI capabilities**

Looking ahead, funded AI companies must prove real business value. The winners will likely be those that can balance innovation with sustainable monetization while navigating increasing regulatory scrutiny.

AI giants raced to AGI as Chinese rivals proved formidable.

MAY 2024

OpenAI Launches GPT-4o

The new AI model is capable of real-time reasoning across audio, visual, and text inputs.

DECEMBER 2024

O3 Closes In on AGI

OpenAI says its o3 model has passed the ARC-AGI challenge, considered a leading benchmark for artificial general intelligence.

JANUARY 2025

Nvidia enters the AI PC market

Nvidia unveils its Project DIGITS, a personal AI supercomputer.

JANUARY 2025

DeepSeek Disrupts

Chinese AI company DeepSeek releases R1, its reasoning model and competitor to OpenAI's o1.

JANUARY 2025

Tech Giants Introduce 500B AI plan

Political, tech, and financial leaders announce Stargate, a joint venture that aims to invest $500 billion in US AI infrastructure.

« PAST

Tech titans will face pivotal AI tests in 2025.

MARCH 2025

Huang to Preview Next-Gen AI Chips

Nvidia's GPU Technology Conference will feature CEO Jensen Huang's keynote on what to expect from this critical chip manufacturer.

MAY 2025

Nvidia's Growth Is Tested

Nvidia could show record-breaking revenue—unless DeepSeek portends an alternative future requiring fewer chips.

JULY 2025

China Shows Its AI Hand

The World AI Conference will showcase China's latest AI innovations and initiatives.

FUTURE »

MAY 2025

AI Integrates

Google I/O's new generative AI updates will signal how AI will further integrate into consumer services and enterprise tools.

JUNE 2025

Apple Bets Big on a Smarter Siri

In a "make or break" moment for proving AI's necessity in consumer products, Apple is poised to debut a new Siri with generative AI.

FTSG

Beyond the AI hype, these six structural changes are already determining which organizations will thrive.

Speed Is the New Scale

For better or worse, AI is compressing decision cycles from weeks to minutes. The advantage is shifting to organizations that can harness AI for rapid experimentation and learning—as long as they're making good decisions about data governance, vendor selection, and change management.

Your Competitors Won't Wait

Someone in your industry is already using AI to cut costs and boost productivity by 30%–40%. AI is automating certain labor intensive knowledge tasks, but it will soon lead to new workflows and business models. The question isn't whether your organization will adapt to AI, but whether you'll do it before or after your margins get squeezed from all directions.

The Middle Office Is Melting

AI is automating coordination and decision-making tasks that traditionally required human middleware, and organizations clinging to manual coordination and approval processes will find themselves structurally uncompetitive. The future org chart is flatter and faster.

The Talent War Has New Rules

The best talent now expects to work with the best tools. They're not just looking for good pay—they want AI-enabled workplaces that multiply their impact. Your ability to attract and retain top performers increasingly depends on your AI readiness.

A Hidden AI Tax

The cost of AI isn't in buying the technology—it's in powering it. As demand for AI computing skyrockets, organizations will face a new economic reality: Buy in early to start their AI transformation, but unwittingly pay an increasingly steep premium later.

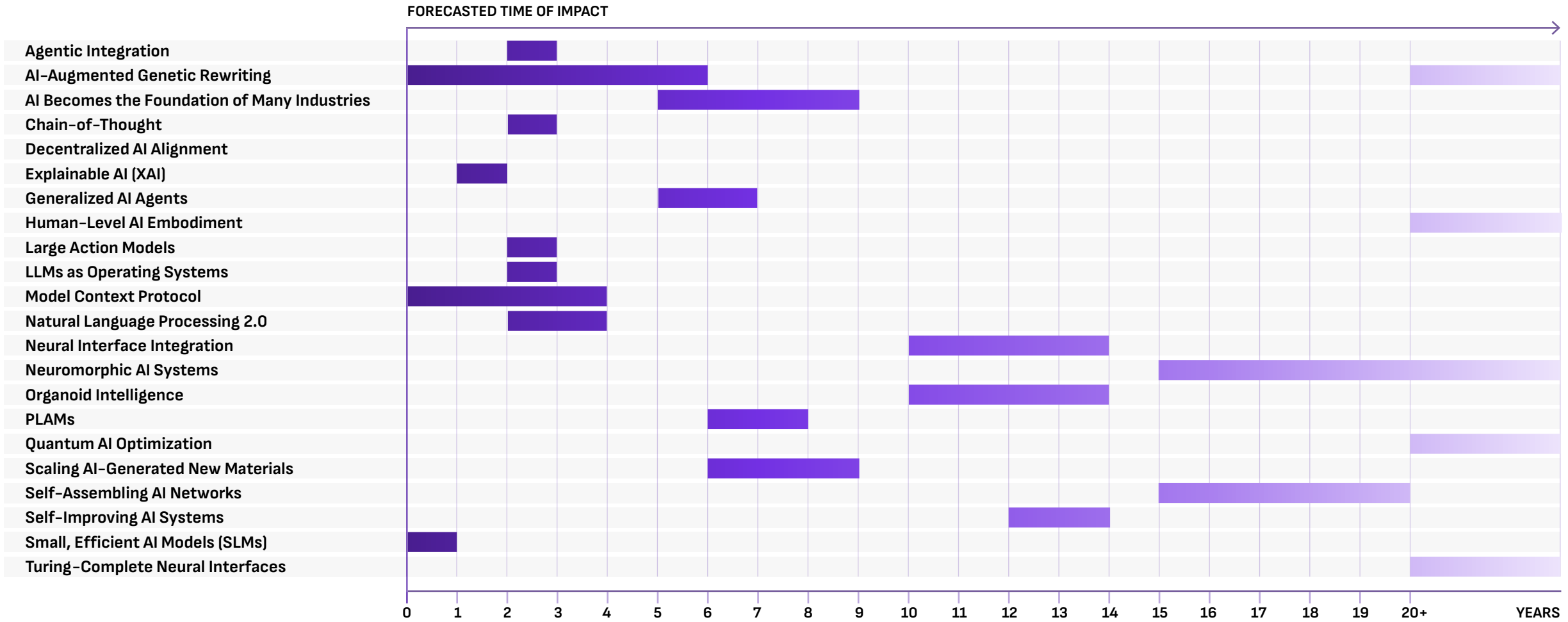
Your Interface Is Costing You Customers

Natural language is eating your user interface, turning it into abandonware. Every app, database, and system will soon be accessible through simple conversation. Organizations clinging to traditional interfaces will find themselves with the corporate equivalent of a flip phone in an iPhone world.

Scale Out Over Scale Up

History shows that core technologies inevitably become cheaper and more widely distributed. The same pattern is emerging in AI, and the companies that only chase ever-larger models and data centers risk being disrupted by nimble upstarts that can spin up smaller, more efficient systems, at lower cost.

Generative technologies advance over the next several years, while computing methods like organoid and neuromorphic computing drive developments in the long term.



Below, we highlight high level near-term developments to keep an eye on across industries.

SCALING

Enormous amounts of training data are still required for most AI models to learn. For example, recommender systems coupled with generative AI could lead to deep personalization for the hospitality and health care sectors—as long as data is made available. Historically, data is locked inside proprietary systems built by third parties, and regulation often hinders access to certain forms of data.

INVESTMENT

AI has seen cycles of enthusiasm and disillusionment, leading to either too much or not enough capital. Investors prioritize commercialization over basic R&D—though the latter yields bigger impact and often stronger returns. Investors' patience will influence progress and commercialization.

CONSTRAINTS ON ADOPTION

Even if a technology is maturing, constraints on its adoption can hinder its impact. For example, a business may refuse to adopt an automated system because it challenges existing orthodoxy or an existing successful strategy. This is especially true in health care, insurance, and financial services.

REGULATIONS

Advances in technology typically outpace regulatory changes. This has benefited AI, which until very recently was not targeted for regulation. Additionally, factors like whether local regulations are conflicting or complementary can influence adoption in the marketplace.

MEDIA MENTIONS

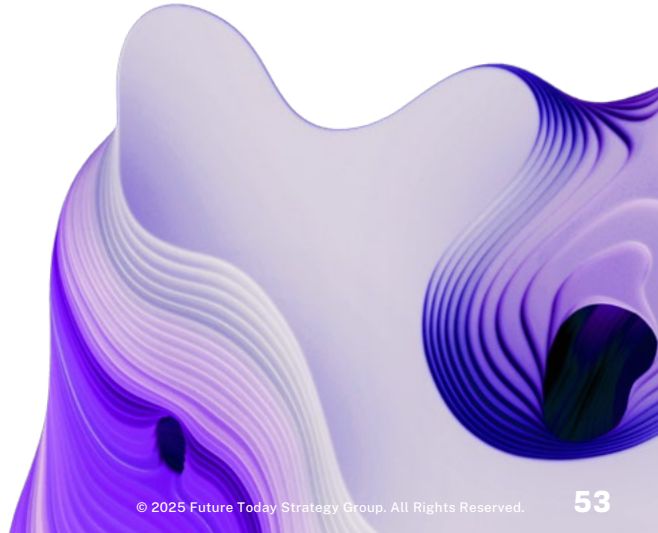
Increased awareness and enthusiasm can influence the momentum of a technology, even when there's been no real breakthrough. Future media bursts will drive AI momentum, especially if those stories are easily understood by the public.

PUBLIC PERCEPTION

How the public understands and responds to AI advancements will create or quell demand. This is especially true of generative AI and education/creativity/intellectual property/misinformation, as well as the role assistive technologies will play in shaping the future workforce.

R&D DEVELOPMENTS

The pace of new research breakthroughs can't be scheduled to coincide with a board meeting or earnings report. Factors like funding, quality, size of staff, and access to resources can improve the likelihood and speed of new discoveries. We closely monitor R&D developments but treat them as wild cards.



We expect to hear often from the world's largest technology and AI companies. For that reason, our 2025 list highlights individuals flying deeper under the public radar.

- **Dr. Adji Bousso Dieng, assistant professor at Princeton University,** for her work in deep probabilistic graphical modeling.
- **Alexandr Wang, founder and CEO of Scale AI,** for revealing DeepSeek's open-source AI model and for creating a leading data annotation platform that accelerates AI model development across various industries.
- **Dr. Abeba Birhane, senior fellow at Mozilla Foundation,** for her research on the ethical implications of AI and critiques of algorithmic biases.
- **Dr. Anima Anandkumar, Bren Professor of Computing and Mathematical Sciences at Caltech,** for developing AI algorithms that accelerate scientific discovery, including frameworks like neural operators for efficient simulations.
- **Dr. Chinasa T. Okolo, computer scientist and fellow at the Brookings Institution,** for her work in advocating for responsible AI adoption in the Global South and contributing to international AI safety reports.
- **Clément Delangue, CEO at Hugging Face,** for democratizing access to state-of-the-art NLP models and fostering an open-source AI community.
- **Dr. Cynthia Rudin, the Gilbert, Louis, and Edward Lehrman Distinguished Professor at Duke University,** for her work on interpretable machine learning models and ethical AI.
- **Dr. Dan Hendrycks, director at the Center for AI Safety,** for pioneering research in AI safety and developing Humanity's Last Exam, a test designed to evaluate AI risks, in collaboration with Scale AI.
- **Dean Ball, research fellow on AI & Progress at Georgetown University's Mercatus Center and author of the AI-focused Substack "Hyperdimensional,"** for his analysis on AI governance.
- **Dr. Devi Parikh, professor at Georgia Tech and co-founder of Yutori,** for pioneering work in visual question answering and vision-language models, which helped establish foundational benchmarks for how AI systems understand and reason about visual information in natural language contexts.
- **Dr. Ilya Sutskever, co-founder at Safe Superintelligence,** for pioneering work in deep learning and leading efforts to develop AI that surpasses human intelligence while remaining aligned with human interests.

- **Dr. Joelle Pineau, vice president of AI research at Meta,** for leading advancements in AI research and promoting open science, contributing to developments like the open-source language model LLamA.
- **Dr. Kazumi Fukuda, research scientist at Sony AI,** for her research in embodied intelligence, particularly in developing computational models that enable robots to perceive, plan, and act in dynamic environments.
- **Dr. Li Deng, chief AI officer at Vatic Investments,** for his contributions to speech recognition and deep learning.
- **Dr. Liang Wenfeng, CEO at DeepSeek,** for developing the R1 AI model, which rivals top competitors in capability but operates at a fraction of the cost.
- **Dr. Lila Ibrahim, chief operating officer at Google DeepMind,** for guiding the integration of AI research into practical applications and leading initiatives to apply AI in consumer products.
- **May Habib, CEO and co-founder at Writer,** for leading the development of enterprise AI tools that help businesses generate and manage high-quality content while ensuring brand consistency and compliance.
- **Dr. Nathan Lambert, research scientist at the Allen Institute for AI (Ai2) and author of the Interconnects blog,** for his contributions to the open science of language model fine-tuning.
- **Dr. Pieter Abbeel, director of the Berkeley Robot Learning Lab and co-director of the Berkeley Artificial Intelligence Research Lab,** for his work in robotics and reinforcement learning.
- **Dr. Sasha Luccioni, climate and AI lead at Hugging Face,** for developing tools to measure the carbon footprint of AI models and advocating for environmentally responsible AI practices.
- **Dr. Sheng Shen, research scientist at Google,** for coauthoring "Mixture-of-Experts Meets Instruction Tuning: A Winning Combination for Large Language Models," exploring the integration of MoE architectures with instruction tuning to enhance language model performance.
- **Dr. Tim Brooks, leader of the world modeling AI team at Google DeepMind,** for his efforts in developing "world models" capable of simulating physical environments, advancing embodied AI in gaming and robotics.
- **Dr. Yang Zhilin, CEO at Moonshot AI,** for leading the development of AI models with long context understanding and expanding AI applications globally.

AI adoption is creating unprecedented opportunities for value creation...

OPPORTUNITIES

Build Internal AI Model Evaluation Frameworks

Companies that act now will gain first-mover advantage. These critical frameworks will enable rapid assessment and deployment of AI solutions while competitors struggle with ad-hoc evaluation methods.

Create AI-Powered Knowledge Management Systems

These systems transform static documentation into knowledge bases that continuously learn and adapt, letting businesses unlock significant value from institutional knowledge and retiring employees.

Embed AI Capabilities Directly Into Core Offerings

Most CEOs see AI as a cost-cutting tool, missing its potential to create new revenue streams through enhanced products and services that transform customer experiences and business models.

Invest in Domain-Specific AI Models

These focused models will deliver superior performance in targeted applications while requiring less data and compute resources than general-purpose alternatives, and early investors will benefit.

...but organizations face growing risks from technical complexity and talent scarcity.

THREATS

Assess Hidden Technical Debt

High-value AI clusters are prime targets for sophisticated cyberthreats, making cybersecurity investment essential. The stakes are especially high as thieves could use compromised AI clusters to steal proprietary models, impacting industries globally.

Delaying AI Adoption Means Rising Talent Costs

AI professionals, making it prohibitively expensive to transform later. CEOs often delegate AI strategy to technical teams, ignoring the larger organizational and cultural changes needed.

Compounding Data Advantage Could Concentrate Power

AI's first-movers are accumulating massive proprietary datasets and training pipelines that create nearly insurmountable barriers to entry. This advantage compounds over time and threatens to lock out new players.

Risks of Digital Colonization

Nations without sovereign AI capabilities become vulnerable to digital colonization as foreign AI systems shape their residents' information access and decision-making.

Companies face steep hidden costs and complex organizational hurdles as they rush to implement AI, from outdated infrastructure to employee resistance and regulatory demands.

As organizations hurry to deploy AI, they're facing a costly reality: Most of their data infrastructure is decades behind what's required. The investment needed to modernize data architecture often exceeds initial AI project budgets by 5–10x, creating a hidden barrier to transformation.

While companies eagerly invest in AI, they often overlook crucial investments in change management and employee support. This oversight could lead middle managers to view AI as a threat rather than a tool, fostering resistance that may dramatically slow implementation and adoption.

Build AI models with dedicated testing environments to rigorously evaluate financial risk against historical market conditions and stress scenarios. These systems require specialized hardware and compliance monitoring infrastructure to handle complex simulations while maintaining audit trails.

Establish rigorous protocols for auditing AI models that enable tracking of decisions, data lineage, and performance drift across systems. Integrating automation with human oversight is crucial, along with detailed audit trails that comply with both technical governance and regulations.

Know your constraints. Either find trusted external partners with expertise in both AI and your domain, or do it in-house. These teams must meld their AI skills with industry-specific knowledge to create solutions that address both regulatory and technical challenges while delivering measurable value.

Every piece of technical debt becomes a critical bottleneck as AI systems demand more flexible, interconnected infrastructure. Smart companies will treat technical debt elimination as a core part of AI strategy, not a separate IT initiative.

Technology Deployment

Capital Expenditure

Talent Development

Regulatory Influence

Important terms to know before reading.

AGENTIC AI
This refers to AI systems that exhibit autonomous decision-making, goal-setting, and adaptive problem-solving capabilities. Unlike traditional AI models that passively generate responses based on user prompts, agentic AI proactively takes actions, interacts with its environment, and refines its strategies over time.

AGENTS
AI-powered entities that perceive their environment, make decisions, and take actions autonomously to achieve specific goals. Agents can range from simple automation tools to complex, multimodal AI systems that interact dynamically with users and other systems.

AGI (ARTIFICIAL GENERAL INTELLIGENCE)
A designation for AI systems that match and then exceed the full range of human cognitive abilities across all economically valuable tasks. AGI remains theoretical, but its potential implications for labor markets, governance, and global security are actively debated.

AI ETHICS
A multidisciplinary field that studies the societal, economic, and ethical risks of AI, including bias, privacy, misinformation, and existential threats. AI ethics frameworks guide policy and regulation to ensure AI development aligns with human values.

AI GOVERNANCE
The systems, policies, and international agreements that regulate the development, deployment, and oversight of AI technologies. AI governance is critical to mitigating risks, ensuring fair competition, and addressing geopolitical tensions around AI capabilities.

ALGORITHM
A structured set of rules or processes for solving specific problems or performing tasks. In AI, algorithms determine how data is processed, insights are generated, and decisions are made.

ALIGNMENT
The process of ensuring that an AI system's goals, behaviors, and decision-making align with human intentions, ethical principles, and regulatory standards. Misalignment can result in unintended consequences, including biased or harmful outcomes.

ARTIFICIAL SUPERINTELLIGENCE (ASI)
A hypothetical future AI system that surpasses human intelligence across all domains, including creativity, general wisdom, strategic planning, and scientific discovery. ASI raises complex questions about control, governance, and existential risk.

AUTOMATIC SPEECH RECOGNITION (ASR)
AI-driven systems that convert spoken language into written text. ASR powers virtual assistants, transcription services, and multilingual voice interfaces in enterprise and consumer applications.

AUTONOMOUS AI
AI systems capable of independent decision-making and execution of tasks without human intervention. Autonomous AI is critical in robotics, finance, cybersecurity, and military applications, requiring rigorous safeguards to ensure responsible use.

CHAIN OF THOUGHT (COT) REASONING
An AI reasoning method where models solve problems step-by-step, mimicking human-like logical deduction. This improves performance in complex decision-making tasks, including math, legal analysis, and medical diagnostics.

COMPUTER VISION
AI-driven technology that enables machines to process, analyze, and derive meaning from digital images and video. Used in security surveillance, industrial automation, medical imaging, and self-driving vehicles.

EDGE AI
AI models that run directly on edge devices (e.g., smartphones, IoT sensors, autonomous drones) rather than centralized cloud servers. Edge AI enables real-time processing, reduces latency, and enhances data privacy.

FOUNDATION MODEL
A large-scale AI model pretrained on vast amounts of data and adaptable to multiple tasks without requiring retraining from scratch. Foundation models underpin modern AI applications, including generative AI, autonomous systems, and enterprise automation.

GENERATIVE AI (GENAI)
AI technologies capable of generating novel content, including text, images, music, video, and code. GenAI is transforming industries such as media, design, marketing, and customer service while raising concerns about intellectual property and misinformation.

GPU (GRAPHICS PROCESSING UNIT)
Specialized hardware optimized for parallel computing, accelerating AI model training, deep learning, and high-performance computing tasks. GPUs are essential for running large-scale AI models and data-intensive simulations.

MODEL
A trained AI system that analyzes data to make predictions, generate insights, or automate decision-making. Models vary in complexity, from simple regression models to advanced deep learning architectures.

MULTIMODAL AI
AI systems that process and integrate multiple types of data—such as text, images, video, and audio—to improve contextual understanding and decision-making. Multimodal AI powers advanced chatbots, virtual assistants, and medical diagnostics.

NATURAL LANGUAGE PROCESSING (NLP)
AI-driven processes that enable machines to understand, interpret, and generate human language. NLP powers chatbots, translation services, sentiment analysis, and automated content moderation.

NEURAL ARCHITECTURE SEARCH (NAS)
An AI-driven method for automatically optimizing neural network structures, improving performance while reducing the need for manual tuning by researchers.

PARAMETER
An internal variable of an AI model that is fine-tuned during training to improve accuracy and efficiency. Large AI models contain billions of parameters, making their training computationally intensive.

PROMPT ENGINEERING
The practice of designing effective inputs (prompts) to guide AI models in generating desired outputs. Prompt engineering is crucial for optimizing generative AI performance in business and creative applications.

QUANTUM AI
The intersection of quantum computing and AI, where quantum algorithms enhance machine learning efficiency. Quantum AI has the potential to revolutionize cryptography, materials science, and optimization problems.

RECOMMENDER SYSTEMS
AI-driven algorithms that analyze user behavior and preferences to suggest relevant products, content, or services. Used in e-commerce, streaming platforms, and digital advertising.

REINFORCEMENT LEARNING FROM HUMAN FEEDBACK (RHLF)
A training method where AI models learn through iterative feedback from human evaluators, improving their accuracy, ethical alignment, and usability in real-world applications.

SELF-SUPERVISED LEARNING
A machine learning approach where AI models learn from raw, unlabeled data by identifying patterns and relationships within the dataset. This method reduces dependency on human-labeled training data.

SUPERVISED LEARNING
A training method where AI models learn from labeled datasets, using known input-output pairs to improve predictive accuracy in new data.

SYMBOLIC AI
An AI approach that represents knowledge using human-readable symbols and logical rules, enabling reasoning and problem-solving. Often used in expert systems and explainable AI models.

SYNTHETIC DATA
Artificially generated data used to train AI models when real-world data is scarce, biased, or privacy-sensitive. Synthetic data enhances AI performance while mitigating data collection risks.

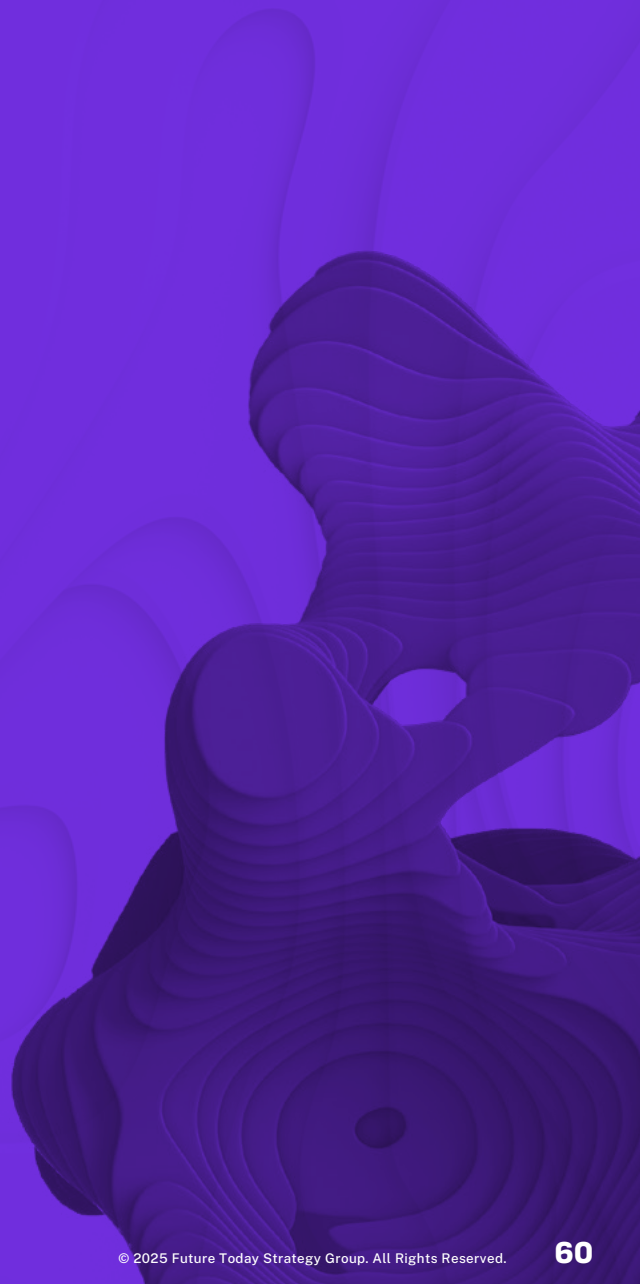
TRAINING DATA
The dataset used to train AI models by identifying patterns, making decisions, or generating predictions. The quality and diversity of training data significantly impact model accuracy and fairness.

TRUSTWORTHY AI
AI systems designed with transparency, fairness, accountability, and security to foster public trust and regulatory compliance. Trustworthy AI is a key focus for government and enterprise AI strategies.

UNSUPERVISED LEARNING
A machine learning approach where AI models detect patterns and structures in data without labeled outputs, enabling tasks like clustering and anomaly detection.

XAI (EXPLAINABLE AI)
AI systems designed to provide transparent, human-interpretable explanations for their decision-making processes, increasing accountability and trust in high-stakes applications like health care and finance.

ZERO-SHOT LEARNING (ZSL)
An AI technique where models generalize knowledge from previously learned concepts to perform tasks without direct prior training on those tasks. Used in applications like language translation and image recognition.



ARTIFICIAL INTELLIGENCE TRENDS

MODELS, TECHNIQUES, & RESEARCH

MODELS, TECHNIQUES, & RESEARCH

AI models require massive data and computing resources to unlock their transformative potential (or so we thought).

Generative AI Modalities Expand

Humans don't just learn by reading—we observe, listen, and synthesize information from multiple sources. AI is now following suit, integrating inputs like text, images, and sound to bridge the gap between what we describe and what machines can fully understand. 2024 marked the year when multimodal AI capabilities not only matured but began transforming real-world applications.

OpenAI's GPT-4o builds on the multimodal progress of 2023 by integrating text, vision, and voice into one robust model. Its real-time conversational capabilities open doors for fluid, human-like interactions. GPT-4o's ability to analyze and generate insights from combined text, audio, and image inputs allows it to solve complex tasks with remarkable depth and accuracy. Anthropic's Claude 3 brings sophisticated visual interpretation to enterprise settings, where roughly half of knowledge bases are image-based. This capability is particularly transformative in health care, where AI can now connect medical imagery with patient records for enhanced diagnostics. On the consumer side, multimodal AI is starting to become second nature. Instead of typing out queries, users now share photos of recipes to adjust portions or upload images of rashes for medical suggestions. While this democratization makes expertise more accessible, it also raises questions about accuracy and ethical boundaries when AI tools replace professional judgment.

MIT and Microsoft's Large Language Model for Mixed Reality (LLMR) is pushing the multimodal boundaries even further. LLMR uses AI to simplify the creation and modification of virtual environments. Instead of needing complex coding, LLMR enables users to describe their vision in plain language, and the system transforms those words into interactive mixed reality experiences in real time. For example, a user might say, "Place a green bench in the park next to the fountain," and the system executes it instantly.

While 2024's breakthroughs in multimodal AI mark a technological leap forward, their true significance lies in how they're reshaping the fundamental relationship between humans and machines, moving us from giving commands to having conversations.

Fine Tuning

Fine-tuning—the process of refining LLMs on specialized datasets—is improving our ability to customize and control AI systems. In 2023, the University of Washington's QLoRA breakthrough marked a turning point, enabling the fine-tuning of massive 65-billion-parameter models on a single GPU with just 48GB of memory—a 16-fold efficiency improvement over traditional methods. Building on this foundation, in 2024, Answer.AI integrated QLoRA with Fully Sharded Data Parallel processing, making it possible to train 70-billion-parameter models on consumer-grade hardware. This democratization has profound implications: Researchers and developers can now experiment with large-scale language models without access to expensive data center infrastructure.

The impact extends beyond accessibility. In the biological sciences, researchers have adapted fine-tuning techniques to protein language models (PLMs), which are trained on extensive datasets of protein sequences. These models are now being fine-tuned to predict protein stability, functions, and interactions with remarkable accuracy. Fine-tuned PLMs outperform their non-tuned counterparts across multiple benchmarks, showcasing enhanced predictive capabilities.

Fine-tuning isn't just a technical capability—it's a strategic business advantage. In the enterprise context, companies can

MODELS, TECHNIQUES, & RESEARCH

use fine-tuning to customize powerful AI models like GPT and Claude for their specific needs without building from scratch. Health care providers can enhance diagnostic capabilities by training models on anonymized patient records while maintaining HIPAA compliance, while financial institutions can embed regulatory requirements—from GDPR to PCI DSS—directly into their AI workflows. This dramatically reduces development costs and time-to-market while ensuring AI systems speak the organization's language, understand industry-specific contexts, and operate within required compliance frameworks.

Automated Reinforcement Learning

Traditional AI training using Reinforcement Learning from Human Feedback (RLHF) involves people rating AI responses to help improve the system. While this method works well, it's expensive and time-consuming. DeepSeek found a clever alternative—the startup developed a way to train AI systems using automated computer feedback instead of human ratings. While more subjective tasks (like creative writing or open-ended questions) still need some human input, DeepSeek's automated method works especially well for tasks with clear right/wrong answers, like math and coding problems. To make this automated training even more efficient, DeepSeek created a special method called GRPO (Group Relative Policy Optimization) and tested it first with its math-focused model. The company isn't alone. Microsoft Asia developed a math model using comparable techniques, Ai2 created a model called Tulu that combines both automated and human feedback, and Hugging Face is working on recreating DeepSeek's approach to better understand how it works. The key takeaway is that DeepSeek showed it's possible to create high-performing AI systems with less reliance on expensive human feedback, particularly for certain types of tasks. This could make AI development more efficient and cost-effective, though human input is still valuable for some applications.

Evolutionary Composition

Sakana AI is challenging the conventional wisdom that bigger, more expensive models are the only path to better AI. Instead of training massive models from scratch—a process requiring enormous computational resources—the Japanese firm has developed an elegant alternative: using evolutionary algorithms to automatically discover optimal ways to combine existing AI models. This "evolutionary optimization" approach is big; by intelligently merging models from different domains—such as language processing and visual understanding—Sakana creates hybrid systems that exceed the capabilities of their individual components. The results are impressive: The experiments produced Japanese language models with enhanced mathematical reasoning and cultural awareness that outperformed larger, more resource-intensive systems.

The implications extend far beyond technical achievement. This methodology democratizes advanced AI development by reducing the need for massive computing infrastructure and specialized expertise. Rather than requiring tens of millions in computing resources, developers can now create sophisticated multi-capable models by intelligently combining existing ones. Most significantly, Sakana's approach suggests a future where AI advancement isn't just about building bigger models but about finding smarter ways to combine existing ones. Just as nature creates complexity through the combination and evolution of simpler elements, this new paradigm points to a more sustainable and accessible path forward in AI development—one where innovation comes from intelligent composition rather than brute-force scaling.

Mixture of Experts

Unlike the previous approach that merges entire trained models, mixture-of-experts (MoE) divides up the work inside a single framework by creating multiple specialized "expert" sub-models. Think of it like having a team of people where one person is great

MODELS, TECHNIQUES, & RESEARCH

at math, another at writing, and another at design, with a manager who knows who to call on for each task. This "manager" (the gating mechanism) directs each input to the right expert, so every piece of the job is handled by the specialist best suited to it. By splitting tasks among experts and letting the gating mechanism handle the "who does what," MoE models can become more efficient and accurate than if one giant, one-size-fits-all model tried to handle everything on its own.

Notably, DeepSeek's January 2025 release, R1, uses MoE at its core. As reported, DeepSeek claims to have built a ChatGPT-like system at a fraction of the usual cost by employing MoE (along with other techniques such as knowledge distillation and reinforcement learning). Because MoE breaks a large model into specialized "experts" and relies on a gating function to route each request to the most appropriate expert, it can be more efficient and potentially less expensive to train or run than a single giant monolithic model. DeepSeek's success with this approach has sparked new attention on MoE as a viable alternative for scaling AI without requiring massive, prohibitively expensive hardware.

Autonomy-of-Experts

Though DeepSeek is what put MoE in the news for the general public, others had already made significant breakthroughs in the field. Researchers at Renmin University of China, Tencent, and Southeast University released a paper that describes a new "Autonomy-of-Experts" (AoE) approach for mixture-of-experts models. While the typical MoE model relies on a "router" that makes its best guess of which specialist should handle each incoming question or input, AoE doesn't need the router. Instead, each expert peeks at the input and says, "I can handle this," or "No thanks, that's not my specialty," based on how strongly it lights up the expert's internal signals (the "activation norms"). The strongest signals win, so those experts step up, and the rest step back. In other words, each expert autonomously decides whether it's the best fit. This cuts out the middleman (the router) entirely.

MODELS, TECHNIQUES, & RESEARCH

LLMs as Operating Systems

Imagine an operating system fundamentally powered by a large language model (LLM), where the LLM is not just an add-on but the core kernel of the OS. This OS could automate routine tasks with unprecedented sophistication, eliminating the need for manual intervention. It would move beyond traditional graphical user interfaces and command-line interactions, embracing a more intuitive, natural language–based approach. Users could interact with their computers through conversational commands, inquiries, or requests for specific tasks, and the LLM would interpret these inputs, executing a series of actions to deliver the desired outcomes.

One such project, AIOS, envisions an LLM as the "brain" of the OS. AIOS optimizes resource allocation, manages context switching, facilitates concurrent agent execution, provides tools for agents, and maintains access control. The LLM handles complex decision-making, turning the OS into a more intelligent, adaptive system. Another project, MemGPT, focuses on enhancing LLM-driven systems by integrating long-term memory and improving reasoning capabilities. Traditional LLMs are limited by small context windows, restricting how much information they can process at once. MemGPT addresses this by introducing a multilevel memory architecture, inspired by traditional OS memory management techniques like virtual memory, to enable more complex and contextually aware processing over time. Together, these projects represent the future of LLM-centric operating systems, enabling more efficient, natural, and powerful interactions.

LLMs: Bigger and More Expensive

LLMs have grown exponentially in size and cost over the past decade, driven by the "bigger is better" paradigm. This approach emerged from scaling laws, first introduced by Prasanth Kolachina in 2012 and later validated by Kaplan et al. in 2020, which demonstrated a strong correlation between model size and performance. Following these insights, the industry has pursued increasingly larger systems, progressing from GPT-2's 1.5 billion parameters in 2019 to models with trillions of parameters like GPT-4 and PaLM 2 in 2023. The financial impact of this growth is significant: Stanford's 2024 AI Index Report estimates place the training costs of top-tier models at unprecedented levels, with OpenAI's GPT-4 requiring approximately $78 million in compute and Google's Gemini Ultra costing an estimated $191 million.

The benefits of scaling have been substantial and well-documented. Larger models have demonstrated remarkable capabilities in handling complex tasks, showing improved accuracy and efficiency across a wide range of applications. These achievements have validated, at least partially, that bigger is indeed better. However, this progress has come with significant costs and challenges. Training GPT-4 required approximately 10,000 times more computational resources than its predecessor GPT-2, necessitating enormous investments in infrastructure and specialized hardware.

Perhaps most significantly, the relationship between model size and performance has proven more complex than initially assumed. Many tasks exhibit diminishing returns as models grow larger, calling into question the long-term viability of this approach. This observation has particular relevance for businesses, which are discovering that larger models don't automatically translate to better solutions for their specific needs. The combination of rising costs, environmental concerns, and uncertain performance benefits has prompted a critical examination of the scaling paradigm.

As the field matures, a more nuanced approach to AI development is emerging. Rather than pursuing size alone, researchers are increasingly focusing on efficiency improvements and the development of smaller, specialized models.

Chain-of-Thought Models

As larger models reach practical and financial limits, researchers are shifting attention to new approaches such as Chain-of-Thought (CoT), which emphasizes

MODELS, TECHNIQUES, & RESEARCH

deeper real-time reasoning rather than raw parameter counts. For more than a decade, AI progress has been largely driven by the pretraining scaling law, which emerged with AlexNet in 2012 and gained momentum with the Transformer architecture in 2017. This law states that increasing the amount of training data (now reaching trillions of tokens), expanding model parameters, and using more compute (FLOPS) leads to better performance across various tasks. Put simply, pretraining scaling laws describe how larger models, with more data and compute, achieve superior performance.

Now, there is a new scaling law in town. Previously, most computational cost was concentrated in pretraining. Once a model was trained, running inference—generating responses or completing tasks—required significantly less compute. Inference scaled in a straightforward way: The more requests a model handled, the more compute it used. However, the introduction of CoT models has fundamentally changed this paradigm. OpenAI's o1 model and DeepSeek's R1 have demonstrated that inference compute is no longer strictly proportional to output length. These models generate intermediate "logic tokens," acting as an internal scratchpad to break down problems into structured reasoning steps. This shift means that the more tokens dedicated to this internal process, the better the model's output. Essentially, it mimics how humans improve their work—double-checking, verifying calculations, and cross-referencing solutions to ensure accuracy.

Expect a growing emphasis on dynamic inference strategies, where models can flexibly adjust the number of internal logic tokens based on task difficulty or desired accuracy. As a result, inference compute could become much more significant relative to training, driving the need for more efficient hardware solutions, better optimization techniques, and new business models around usage-based compute. Overall, AI development will likely shift toward architectures and methods that let models "think out loud," enabling deeper reasoning and better outcomes at the cost of increased on-the-fly processing.

MODELS, TECHNIQUES, & RESEARCH

Small Language Models

Small language models (SLMs) are proving that bigger isn't always better. These compact models can match or exceed the performance of their larger counterparts in specific tasks, while demanding far less computational power and resources. SLMs showcased impressive performance in task-specific scenarios, such as zero-shot text classification. Research across multiple datasets revealed that models with fewer parameters could rival larger counterparts, emphasizing their potential for efficiency without compromising effectiveness. Additionally, cost-effective solutions like OpenAI's GPT-4o mini cost more than 60% less compared to previous models, making high-quality AI more accessible to businesses and developers.

Microsoft's Phi-3-mini stands as another example, achieving superior reasoning and logic capabilities with just 3.8 billion parameters—outperforming models twice its size. Similarly, Meta's Llama 3.2 family demonstrates the viability of smaller models, offering variants from 1 billion to 90 billion parameters that prioritize efficiency without sacrificing effectiveness.

This shift toward smaller models is supported by industry experts like Andrej Karpathy, who advocates for distilling models to their essential "cognitive core." His research suggests that even a 1-billion-parameter model could provide sufficient cognitive capabilities, as much of the additional data in larger models may not directly enhance performance. This insight has practical implications, particularly in consumer technology. Apple's OpenELM models, for instance, enable on-device AI processing, delivering responsive, personalized experiences while maintaining privacy and energy efficiency. In 2024, models like the SlimLM series enabled robust processing on devices such as smartphones, eliminating the need for cloud-based computation. This innovation marked a leap forward in AI accessibility, enabling users to perform tasks directly on their devices while maintaining privacy and reducing latency.

The impact of SLMs extends beyond general applications into specialized domains. At Ignite 2024, Microsoft collaborated with industry leaders like Bayer and Rockwell Automation to develop targeted SLMs for agriculture and manufacturing, demonstrating how these compact models can excel in specific sectors without the overhead of larger, general-purpose systems. Looking forward, the concept of "companies of LLMs"—where multiple specialized models work in parallel—could represent the next evolution, combining the advantages of both large and small models in a modular approach.

Grounding and Context Augmentation

NotebookLM from Google Labs transforms AI into a personalized research assistant by "grounding" it in your Google Docs. Unlike traditional chatbots, it ties responses to your specific notes and sources, enabling insights that are highly relevant and trustworthy. This feature is designed to tackle information overload, making it easier to synthesize and connect ideas from multiple sources efficiently. Grounding ensures that AI outputs are anchored in verified data, reducing errors like hallucinations. NotebookLM builds on this by leveraging contextual augmentation, which enriches responses with nuanced understanding tailored to your needs. This approach doesn't just deliver answers but delivers the right answers for your unique context.

Tools like GenAI Data Fusion extend this principle to businesses, aggregating and contextualizing enterprise-specific data. By rooting outputs in tailored datasets, companies can achieve highly accurate, task-specific insights for applications ranging from research to operational analytics. Grounding and contextual augmentation mark a shift in AI, turning generic systems into precise, adaptive tools that meet individual and organizational needs.

Overcoming the Data Shortage

The availability of high-quality data is emerging as a bottleneck in the development of large AI models. According to Epoch AI, the reservoir of high-quality textual data on the public internet may be exhaust-

MODELS, TECHNIQUES, & RESEARCH

ed as early as 2026. Initially, researchers estimated the stock of high-quality language data could run out by 2024, while low-quality data might last another two decades, and image data could face depletion by the late 2030s to mid-2040s. Although the 2024 thresholds have not yet been reached, the looming scarcity is pushing AI labs to explore alternative strategies for sourcing training data.

This prediction has sparked diverse strategies among AI labs. Some are pursuing private data sources, purchasing from brokers or licensing content from publishers. Others are exploring untapped audio and visual data, with video content offering particularly valuable insights into real-world physics and dynamics. Companies like Scale AI and Surge AI are building extensive networks of contributors, including Ph.D.-level experts, to create and annotate specialized datasets. These approaches, however, come at a high cost, with some estimates suggesting AI labs are spending hundreds of millions of dollars annually on these initiatives.

An alternative method involves using one AI model to generate vast amounts of synthetic data to train another, but it's risky. Studies have shown that models trained predominantly on synthetic data can experience "model collapse," where their outputs become less diverse and fail to reflect real-world distributions. A related phenomenon, termed Model Autophagy Disorder (MAD), has been observed in generative image models, where reliance on synthetic data leads to a notable decline in output quality.

One promising solution lies in techniques such as "self-play," where models improve through competition or collaboration with themselves. Google DeepMind's AlphaGo, which famously defeated the human world champion in Go after training against itself, exemplifies this approach. Today, self-play continues to inform cutting-edge LLM development, offering a pathway to overcome data limitations while maintaining performance and innovation.

Open-Source AI

In January 2025, Chinese AI company DeepSeek launched R1, an open-source reasoning model designed to compete with—and potentially outperform—OpenAI's o1 at a fraction of the cost. While R1's reasoning process is slower than that of many general-purpose models, it delivers more nuanced and accurate responses. Alongside its flagship 671-billion-parameter version, DeepSeek also introduced six smaller "distilled" models, starting at 1.5 billion parameters and capable of running on local devices.

Other open-source models are likewise closing the performance gap with proprietary alternatives. Meta's Llama 3.1 now rivals GPT-4 on key benchmarks, and Mistral AI's models offer capabilities on par with top closed-source solutions—so much so that Mistral AI's recent $487 million funding round catapulted it to unicorn status. As a growing number of open-source models achieve high-level performance, corporations are increasingly taking note. The Brave browser, for example, has integrated Mistral AI's Mixtral 8x7B model into its Leo assistant, while Wells Fargo has adopted Meta's Llama 2 for internal applications.

This momentum reflects a broader trend toward open-source AI. GitHub statistics reveal that AI-focused repositories have skyrocketed from just 845 in 2011 to 1.8 million in 2023—an impressive 59.3% increase in 2023 alone. During the same period, community engagement soared, with GitHub stars for AI projects jumping from 4 million to 12.2 million between 2022 and 2023. This surge highlights the growing appetite for collaborative development, signaling that open-source AI will remain a driving force well into the future.

Modular AI

Rather than building monolithic models, the modular AI approach breaks AI systems into specialized, independent components that can be mixed and matched like building blocks. Each module handles specific tasks or domains, allowing for precise control over system capabilities

MODELS, TECHNIQUES, & RESEARCH

and resources. One example is a recent proposal to develop large language models (LLMs) using "bricks": modular components representing specific tasks or knowledge domains. These bricks can emerge during pretraining or be custom-designed after training for particular applications. This approach dynamically activates only the relevant bricks for a task, significantly reducing computational and energy costs while improving scalability. Research has shown that neurons and layers within LLMs naturally specialize in different functions, highlighting the promise of modular design in optimizing AI systems.

Similarly, MAGNUM, a modular multimodal AI framework introduced at NeurIPS 2023, offers unparalleled flexibility by processing structured and unstructured data across multiple input types, including text, images, video, audio, and time-series data. Composed of input-specific modules, MAGNUM excels at combining and extracting information from diverse data types, performing well across 10 real-world tasks like medical diagnostics and weather forecasting. Notably, it is robust against missing or incomplete data, a common challenge in multimodal AI systems.

Another modular AI advancement is the MASAI (Modular Architecture for Software-engineering AI) framework, which uses specialized sub-agents powered by LLMs to tackle distinct sub-problems in software engineering. MASAI allows for fine-tuned problem-solving strategies across sub-agents, efficient information retrieval, and reduced computational overhead. This architecture achieved a top performance of 28.33% on the SWE-bench Lite dataset, demonstrating its effectiveness in resolving real-world GitHub issues.

Large Action Models

Large action models (LAMs) go beyond traditional large language models by executing tasks rather than just processing language. They act as autonomous agents that can interact with computer interfaces—clicking buttons, moving cursors, and typing text. An early example is Claude 3.5 Sonnet, which can interface with computers like a human would: navigating desktop applications, moving cursors and clicking buttons, typing text, interpreting screenshots and responding accordingly, and executing complex, multistep tasks on a computer.

This capability marks a transition in AI, bridging the gap between conversational models and full-fledged autonomous systems. While currently operating by impersonating a user—interacting with interfaces designed for humans—LAMs demonstrate the foundational steps toward AI systems that could manage tasks directly within digital ecosystems. For instance, a LAM could book a plane ticket or draft a document by navigating apps and websites without user intervention.

The major AI companies—Google, Apple, Microsoft, OpenAI, and others—are positioning these agents as the future of AI. Their goal is to create systems that move beyond chat interactions and actively engage with the world. Claude's ability to interact with a desktop computer demonstrates this potential, but it also raises significant questions about privacy and control. To function, LAMs require broad access to users' digital environments—reading screens, accessing files, and executing commands. This level of access creates an unprecedented intimacy between AI and its users, raising concerns about how data is collected, stored, and used. Companies see this as a massive opportunity. By integrating deeply into users' digital lives, AI companies could gain access to data beyond anything previously collected by traditional tech giants, potentially redefining norms around data privacy.

Personal Large Action Models

Taking the concept of large action models one step further, personal large action models (PLAMs) represent an even more intimate evolution of AI. While current LAMs are trained on general internet data, PLAMs would be specifically trained on the digital footprint of a single user. Imagine

MODELS, TECHNIQUES, & RESEARCH

an AI that knows your social media interactions, online purchasing habits, biometrics, location data, texts, calendar, and email—understanding not just your data, but the exact context of your life: where you are, who you're with, and what you're doing.

While PLAMs are still largely theoretical, they would mark a significant shift in AI personalization. These systems would learn from every aspect of your digital life: banking records, browsing history, IoT devices, car usage, wearable data, and biological information. This comprehensive understanding would allow the PLAM to grasp your routines and preferences, eventually making decisions that align precisely with what you would have chosen—from purchases to schedule management.

As these models build accuracy over time, they would take on increasingly complex decision-making roles, effectively becoming a digital extension of a person. This level of personalization creates a significant lock-in effect—transferring years of individualized learning to a new system would be impractical, if not impossible. Companies that successfully deploy PLAMs early could therefore establish themselves as indispensable digital partners, with users becoming deeply integrated into their ecosystem.

SAFETY, ETHICS, & SOCIETY

SAFETY, ETHICS, & SOCIETY

AI model safety requires a balance of rigorous testing and governance to prevent harm without stifling innovation.

Explainable AI (XAI)

Explainable AI (XAI) addresses the challenge of understanding how AI systems, especially complex models like deep neural networks, reach their conclusions. Many AI models, often labeled "black boxes," operate without clear insight into their internal decision-making processes. These deep learning models are built on layers of artificial neurons that process data and identify patterns—the interconnected layers can perform highly complex tasks, but their intricacies make it hard to trace how they arrive at specific outputs. This lack of transparency is problematic, particularly in sensitive sectors like health care, finance, and criminal justice, where understanding an AI's logic can impact critical decisions.

XAI aims to bridge this gap by developing methods to reveal the workings of these "black box" models. It seeks to provide clear explanations that help users understand how certain inputs result in particular outputs, enabling more informed and trustworthy AI use. For instance, recent research from the University of California, San Diego has found mathematical formulas that describe how neural networks identify relevant data patterns. By providing these clearer insights into how complex models, like deep neural networks, make decisions, XAI aims to empower users and researchers to better understand, validate, and refine AI applications.

AI Optimization

AI optimization (AIO) is a new field focused on shaping how AI models, particularly chatbots and language systems, respond to certain queries and references. Similar to search engine optimization (SEO), which boosts website rankings in search results, AIO is designed to refine interactions within AI models, targeting a favorable or specific portrayal of brands, individuals, or products. With AIO, companies can guide chatbots to respond positively to certain prompts, potentially recommending a brand, showcasing a positive product review, or emphasizing particular qualities of a person. Techniques like "strategic text sequences" and "invisible text" embedded on websites can subtly influence these AI-generated answers. A restaurant might use AIO to ensure it's listed as the "best restaurant" in a city, or a tech company might seek to have its product as the top choice in a category. However, AIO also raises ethical concerns around transparency and manipulation, especially as people increasingly rely on AI for information.

Decentralized AI Alignment

As AI systems become more powerful, safeguards are necessary to ensure these technologies align with human values and do no harm. This is especially important as AI surpasses human intelligence. AI alignment research focuses on designing systems that act according to human goals and ethical standards. OpenAI, founded in 2015 with a mission to "advance digital intelligence in a way that benefits humanity," has prioritized this alignment to avoid risks from the unchecked development or misuse of advanced AI. For that reason, it was a surprise to some when in May 2024, OpenAI disbanded its Superalignment team, which had been dedicated to ensuring AI safety. While the company said this work would now be integrated across its broader research teams, there is speculation that internal challenges influenced this decision. For instance, Jan Leike and Ilya Sutskever, former OpenAI Superalignment co-leads, cited struggles in accessing sufficient computing resources for their alignment research before they left OpenAI. Other tech giants, such as Google and Meta, have made similar moves, redistributing their AI safety work throughout their organizations rather than in specialized teams. While these companies argue that integrating safety across departments prevents isolation, critics argue that a ded-

SAFETY, ETHICS, & SOCIETY

icated safety team is crucial for securing the resources and influence necessary to prioritize ethical oversight effectively.

Mission Drift in AI Alignment

Many AI alignment organizations, originally established to ensure that AI serves humanity's best interests, are increasingly taking actions that appear to conflict with their founding principles. For instance, Palantir is partnering with Anthropic to deploy Claude models in US government intelligence and defense operations. Marketed as an "asymmetric AI advantage," the collaboration introduces AI to classified environments, claiming to enhance analytics and operational efficiencies. However, this move has sparked controversy due to its potential conflict with Anthropic's original mission to create AI systems aligned with human values and to prevent misuse. Critics argue that applying AI in classified operations, which may involve surveillance or military use, runs counter to Anthropic's foundational principles.

This type of move isn't unique to Anthropic. OpenAI has transitioned into a for-profit public benefit corporation. The shift is driven by the financial and operational challenges of its nonprofit origins, allowing OpenAI to streamline decision-making and attract significant capital. While this restructuring could enable faster progress, it also introduces competing profit motivations that may dilute the organization's commitment to its mission of ensuring artificial general intelligence benefits for all of humanity. It is entirely possible for a for-profit organization to remain mission-driven—many successful examples exist—but doing so demands unwavering focus and discipline.

These moves signal a broader shift within the AI alignment field, where organizations must now balance ethical responsibilities with the demands of profit-driven models. While such alignment is possible, it requires disciplined governance and vigilance to ensure the original missions remain intact amidst growing external pressures.

Indexing Trust

As AI systems evolve, distinguishing between authentic and tampered data—whether altered deliberately or by mistake—will become increasingly challenging. Trust is foundational to the effective use of AI; without it, decades of research and technological advancements may lose their value. Leaders across all sectors—government, business, and nonprofits—need to trust the data and algorithms driving these technologies. Building this trust demands transparency, a significant challenge but one that researchers are actively addressing.

A major step toward transparency has been the development of the Foundation Model Transparency Index (FMTI), created by researchers from Stanford, MIT, and Princeton. This scoring system evaluates transparency in AI model development, functionality, and use. According to the FMTI's May 2024 report, the overall transparency score has improved since October 2023: The mean score rose to 58 out of 100, and the top score reached 85, a 31-point increase. Additionally, each of the eight developers assessed in both rounds improved their scores, with 96 out of 100 indicators met by at least one developer and 89 by multiple developers. And notably, all 14 companies assessed disclosed new information, for an average of 16.6 indicators each.

Other organizations are also working to establish benchmarks that researchers can use to drive improvements. The National Institute of Standards and Technology provides an AI Risk Management Framework, and University of California, Berkeley offers a Taxonomy of Trustworthiness for AI. Indexing trust serves not only as a valuable benchmark for researchers but also as a tool for policymakers. By highlighting areas of persistent and systemic opacity, it clarifies where policy interventions may be needed.

Realtime Deepfake Detectors

Last year marked when we moved past the "uncanny valley," a term describing

SAFETY, ETHICS, & SOCIETY

the discomfort people feel when viewing robots or digital avatars that seem almost, but not quite, human. This concept extends to deepfake AI content; previously, it was easy to distinguish between genuine and AI-generated media based on instinct. Today, distinguishing real from fake is far harder, and we can no longer rely on instinct alone. Deepfakes can now deceive even the most discerning eye, creating risks around authenticity and security.

Researchers are stepping up to meet this challenge. At NYU, a team has introduced methods to counteract real-time deepfakes, which are advanced AI-generated audio and video imitating real people in live settings. Their solution includes eight visual tests designed to help users recognize when they are not interacting with a genuine person. Similar to CAPTCHA, these tests ask questions or make requests that deepfake systems struggle to answer correctly. The researchers found challenges like specific head movements and facial obstructions effective. Human evaluators achieved 89% Area Under the Curve score in identifying deepfakes, while machine learning models achieved 73%. The challenge of deepfakes extends to audio, particularly in call environments where greater human control is essential. To address this, the NYU researchers developed a system that combines human intuition with machine analysis to support call receivers. Their findings revealed that integrating human judgment with machine precision creates a powerful solution, raising detection accuracy to 84.5% and demonstrating a usable approach to combat real-time voice-cloning attacks.

Watermarking

Digital watermarking, the process of embedding hidden digital information within a signal to verify content ownership, has long been used to protect copyrighted material. Similarly, AI watermarking involves embedding a unique, detectable marker in the outputs of AI models—whether text, images, audio, or video—to identify that content as AI-generated. This watermarking is typically integrated during the AI model's training phase, and specialized algorithms can later detect the watermark to confirm the origin of the content.

AI watermarking serves several purposes: It helps identify AI-generated content, distinguish it from human-created work, and offers a way to address issues around misinformation and academic integrity. It has become one of the recommended methods for identifying potential deepfakes, and companies are increasingly embracing the practice. For example, Google DeepMind includes watermarking in its Gemini chatbot responses, and Amazon allows users to verify images generated by its Titan Image Generator through a watermark-checking tool, although this feature currently works only for Amazon's own watermarks.

While watermarking AI content is a promising approach, it has limitations. Even if all major AI platforms adopt watermarking, some models will remain unmarked, and malicious actors are unlikely to use watermarking on deceptive content. This has led to an alternative proposal: watermarking *human*-generated content instead. As AI-produced content continues to grow, genuine human-made content may soon be the rarity online. Shifting the focus to watermarking human-generated material could allow us to assume content is AI-created unless it's marked as human-authored.

Child Safe AI

Just as child-safe modes exist on platforms like YouTube and Netflix, AI also needs safeguards to protect children from inappropriate content. One solution could be Nuha's Teddy, an AI-powered teddy bear prototype designed as a safer alternative to screen-based devices. Teddy, developed by Lama Al Rajih, uses a language model to engage children in verbal interactions, fostering language and social skills. It can discuss topics like space, play games, and encourage active learning. Limited connectivity reduces online risks, and parental controls regulate its responses. Similarly, Little Language Models, part of MIT Media Lab's CoCo platform, introduces children aged 8–16 to foundational AI concepts,

SAFETY, ETHICS, & SOCIETY

focusing on “probabilistic thinking” through a developmentally appropriate, interactive approach.

One major concern with children using AI tools designed for adults is the “empathy gap” identified by Cambridge University researcher Dr. Nomisha Kurian. Her research shows that children often perceive AI as human-like, which can lead to distress when the AI responds inadequately. By creating AI interfaces specifically tailored for children, developers and researchers can bridge this empathy gap. This approach not only protects children from potential harms but also supports their emotional and cognitive development, ensuring that AI tools are both safe and beneficial for young users.

Politically Biased AI

Recent research has revealed that LLMs are not neutral but exhibit distinct political biases, influenced by their training data and design. A study conducted by the University of Washington, Carnegie Mellon, and Xi’an Jiaotong University tested 14 LLMs and found significant ideological differences. OpenAI’s ChatGPT and GPT-4 leaned left-libertarian, while Meta’s Llama displayed a right-wing authoritarian tendency. Researchers mapped these biases using the political compass, revealing how the models responded to topics like feminism and democracy. The study also explored whether retraining models with politically skewed data could change their behavior. It did, significantly altering their capacity to detect hate speech and misinformation. Other studies corroborate these findings. Researchers at the University of East Anglia observed that ChatGPT consistently exhibited liberal biases across different contexts. For instance, the model tended to align with Democrats in the US, Lula’s party in Brazil, and the Labour Party in the UK.

These LLMs can be actively altered through tools like PoliTune, a framework for fine-tuning LLMs to adopt specific political ideologies. This tool created by researchers at Brown University demonstrates how AI models, originally designed to maintain neutrality, can be adapted to produce strong ideological stances. Such developments raise ethical concerns, particularly as LLMs are increasingly used to create news articles, political speeches, and social media content. As this technology proliferates, there is a risk of a fragmented AI landscape, where ideologically polarized models mirror today’s divided media environment.

AI as a Tool to Address Political Bias

Despite the challenges posed by politically biased AI, the technology can also be used to address bias. The University of Pennsylvania’s Media Bias Detector provides detailed insights into how various news outlets frame stories, shedding light on their political leanings. Similarly, the Bipartisan Press has developed an AI model capable of identifying the political bias present in articles and online content.

Techniques are also being explored to reduce bias within AI systems. Researchers at Oregon State University introduced a cost-effective training method called FairDeDup, short for “fair deduplication.” This approach removes redundant information from datasets used to train AI systems, lowering the computational expense while also addressing embedded societal biases. Internet-based datasets often reflect real-world inequities, which can inadvertently become codified in AI models. By analyzing how deduplication impacts the prevalence of bias, researchers can counteract its effects. For example, FairDeDup helps mitigate scenarios where AI systems disproportionately associate certain roles, like CEOs or doctors, with white men. These innovations emphasize the dual nature of AI: while it has the potential to perpetuate bias, it also holds immense promise for exposing and mitigating it.

Gender and Race Biased AI

A study by the University of Chicago and other institutions revealed alarming evidence of strong negative biases against speakers of African American English (AAE) in language models. These systems generated more negative stereotypes about AAE speakers than attitudes recorded from humans in the 1930s. While overt

SAFETY, ETHICS, & SOCIETY

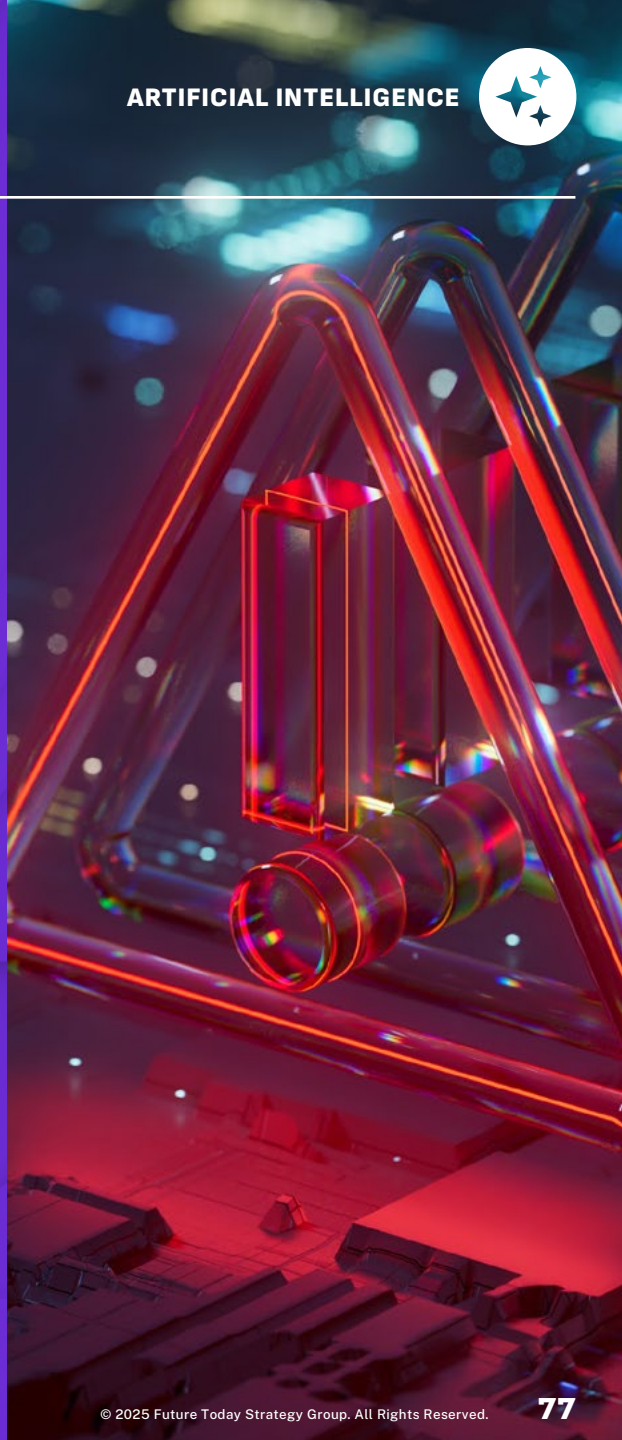
stereotypes about African Americans were often positive, the more covert form of racism—dialect prejudice—was deeply embedded in the AI models. This raciolinguistic bias highlights how AI can perpetuate subtle but harmful discrimination.

Such biases are especially concerning when AI tools are used in critical fields like medicine. For example, AI-driven mental health screening tools analyze speech for signs of anxiety or depression. However, a study from the University of Colorado at Boulder found that these tools fail to perform consistently across different genders and races. Variations in speech, such as higher pitch in women's voices or dialect differences between white and Black speakers, can mislead algorithms, leading to inaccurate assessments. This adds to the growing evidence that AI, much like humans, can make biased assumptions based on race or gender.

Gender bias is another pervasive issue in AI systems. A UNESCO study demonstrated that natural language processing models, including GPT-3.5, GPT-2, and Llama 2, exhibit bias against women in their generated content. Practical examples abound: resume screening tools, like Amazon's notorious system, have discriminated against women; facial recognition technology shows higher error rates for women of color; and medical diagnostic systems frequently provide inaccurate responses for women's symptoms. These biases have real-world consequences, reinforcing harmful stereotypes and creating inequities in access to resources, opportunities, and care.

Nefarious AI Misuse

As a dual use technology, AI is a tool that can be used by both the good guys and the bad guys. Just as AI can be employed to improve health care, enhance cybersecurity, or streamline business operations, it can also be exploited by bad actors. In the past year, the misuse of AI in malicious activities increased—fortunately, its impact remained limited. In October 2024, OpenAI released a report that details various case studies where AI models were exploited by threat actors, primarily in cyber operations like spear-phishing and malware development and in influence campaigns aimed at swaying public opinion through AI-generated social media content. For instance, in July 2024, a network used AI to generate posts emphasizing the benefits of the Rwandan Patriotic Front during the Rwandan election period. Notably, the campaign had little to no effect on the election. In late August, OpenAI disrupted a covert Iranian influence operation that was creating social media posts and long-form articles related to the US election, as well as topics such as the Gaza conflict, Western policies on Israel, Venezuelan politics, and Scottish independence. Most of the social media content generated by this campaign saw minimal engagement, with very few likes, shares, or comments, and there was no evidence of widespread sharing of the web articles across social platforms. However, this limited impact is no reason for complacency. While it's encouraging that these attempts have had minimal success, the increasing use of AI for harmful purposes signals



SAFETY, ETHICS, & SOCIETY

the need for ongoing vigilance. Even AI companies are not exempt from targeting—SweetSpecter, a threat actor based in China, recently launched spear-phishing attacks against OpenAI employees, leveraging AI for tasks like reconnaissance, vulnerability analysis, and scripting.

Data Poisoning: A Double-Edged Sword

Data poisoning is a targeted tactic that manipulates AI training data to introduce vulnerabilities or biases into a model. Unlike inference-phase attacks, it compromises a model's integrity during the foundational training stage. Methods include backdoor poisoning, which embeds exploitable vulnerabilities, and availability attacks, which degrade performance, causing inefficiencies, false outputs, or even system crashes. Model inversion attacks exploit model outputs to infer sensitive training data, often requiring insider access. Meanwhile, stealth attacks gradually introduce subtle changes to training data, embedding biases or inaccuracies that are hard to detect and trace.

Data poisoning is a double-edged sword, capable of being wielded both as a weapon and a shield. While it poses significant threats when used maliciously, it can also serve as a powerful defensive tool. For instance, artists are leveraging data poisoning techniques to safeguard their intellectual property from unauthorized use by AI systems. Tools like Nightshade and Glaze offer innovative solutions to disrupt AI training processes. Nightshade subtly alters pixel data in artwork, rendering AI models that scrape these images inaccurate and unreliable. Glaze, on the other hand, overlays a different artistic style onto original works, obscuring the creator's signature style and preventing precise replication by AI systems. These tools highlight the potential of data poisoning not just to harm but to protect, ensuring that creativity and intellectual property remain secure.

Citizen Surveillance

Countries like China, Russia, and India are heavily investing in AI-powered surveillance technologies, often exporting them or using them to consolidate control domestically. China has positioned itself as a leader in AI surveillance, actively exporting its AI powered recognition technology to nations across Africa, Southeast Asia, and Latin America. Companies like Huawei have provided facial recognition systems, video surveillance, and monitoring software to dozens of countries, many linked to China's Belt and Road Initiative. These projects often involve financial dependencies, as seen in Ecuador's ECU-911 system, financed through Chinese loans in exchange for oil reserves. Similar deals have taken place in Venezuela and Bolivia. These technologies enable extensive monitoring of citizens, with the potential for misuse in censorship or political repression. This is particularly alarming given recent studies that show AI can predict highly personal attributes, such as political orientation, from facial images with surprising accuracy. This capability, coupled with social media's abundance of publicly available photos, could facilitate targeted political messaging or, in authoritarian regimes, surveillance and suppression of dissent.

China is not alone in this pursuit. Russia is building a nationwide AI-enabled "Video Stream Processing Centre" to integrate regional camera networks. Facial recognition technologies, developed by companies with ties to Russian defense or government entities, are already deployed in cities like Moscow and were used to preventively detain at least 141 people in 2022. India, too, is increasingly adopting AI surveillance. Cities like India's Ahmedabad utilize drones and AI-powered cameras to monitor traffic and identify suspicious behavior. Plans for real-time crime detection through Wi-Fi signal analysis further expand AI's reach.

Worker Surveillance

The shift to remote and hybrid work models has fueled a significant increase in worker surveillance. Unlike the restrictions the Fourth Amendment puts on law enforcement, private companies face fewer legal limitations, allowing them to deploy advanced monitoring technologies. A growing number of major corporations including Walmart, Starbucks, Delta, and Chevron use platforms like Aware, an AI-powered

SAFETY, ETHICS, & SOCIETY

tool designed to analyze employee communications on platforms like Slack, Microsoft Teams, and even Reddit. These systems aim to detect risks like harassment and compliance violations while analyzing workplace conversations to gauge employee needs and trends.

Amazon is one of the most well-known practitioners of worker surveillance, using AI-enabled cameras in delivery vehicles to track behaviors such as distracted driving or hard braking. The company penalizes drivers for perceived infractions and employs similar systems in its warehouses, where metrics like "time off task" monitor every moment workers are not actively processing products, applying constant pressure to maintain productivity. These surveillance practices recently resulted in France fining Amazon 32 million euros for violating GDPR regulations. The fine cited a monitoring system with alerts that flagged workers for actions like scanning items too quickly or taking unsanctioned breaks, creating an atmosphere of relentless oversight.

This type of surveillance has measurable impacts on worker well-being. In a survey, 74% of Amazon and Walmart workers reported feeling pressured to work faster due to monitoring systems, leading to increased stress and anxiety. Such practices raise questions about privacy, fairness, and the long-term effects of constant surveillance in the workplace.

School Surveillance

During the pandemic, schools distributed laptops and devices to students for remote learning but often did not disclose that these devices would enable constant monitoring. In many countries, including the US, schools are legally allowed to track students' activities, frequently without informing them or their families about what is being monitored.

Increasingly, these surveillance technologies use AI for facial recognition, predictive policing, geolocation tracking, and student device monitoring. Some even use aerial drones. These tools are promoted as methods to enhance safety, monitor behavior, and identify mental health or safety concerns. For example, the Cheyenne Mountain School District in Colorado Springs has installed nearly 400 AI-enabled cameras that use facial recognition to identify "persons of interest" and track individuals based on characteristics like clothing or backpacks. Alerts and video footage are sent to school officials when matches are detected. The same school has also introduced smart air sensors to detect vaping or drug use.

However, these surveillance systems come with significant consequences. According to a July 2023 report by the Center for Democracy & Technology, monitoring often focuses on detecting inappropriate online content, but it also has broader implications. The awareness of constant surveillance creates a "chilling effect," discouraging students from engaging freely and potentially hindering learning. Alerts stemming from monitoring can impact students' emotional well-being, while the lack of transparency erodes trust between students and educators.

Posthumous AI

Companies are using AI to recreate the voices, likenesses, and personalities of deceased individuals, offering new ways to preserve legacies but also raising complex ethical and emotional concerns. Platforms like Character.ai and Hello History allow users to interact with virtual versions of historical figures, such as William Shakespeare or Queen Elizabeth II. Deep Fusion Films takes this a step further with its upcoming podcast series, "Virtually Parkinson," hosted by an AI replica of the late English broadcaster Sir Michael Parkinson. Built from more than 2,000 interviews from his career, this AI aims to provide authentic, unscripted conversations while explicitly disclosing its artificial nature.

But AI isn't limited to historical or cultural icons. MyHeritage's "Deep Nostalgia" animates photos of deceased relatives, while platforms like Posthumously use generative AI to create immersive 3D avatars of loved ones. These digital spaces enable users to engage with memories and stories of the departed, often for comfort

SAFETY, ETHICS, & SOCIETY

and connection. However, such practices venture into emotionally fraught territory, as individuals use AI to "resurrect" family members and friends.

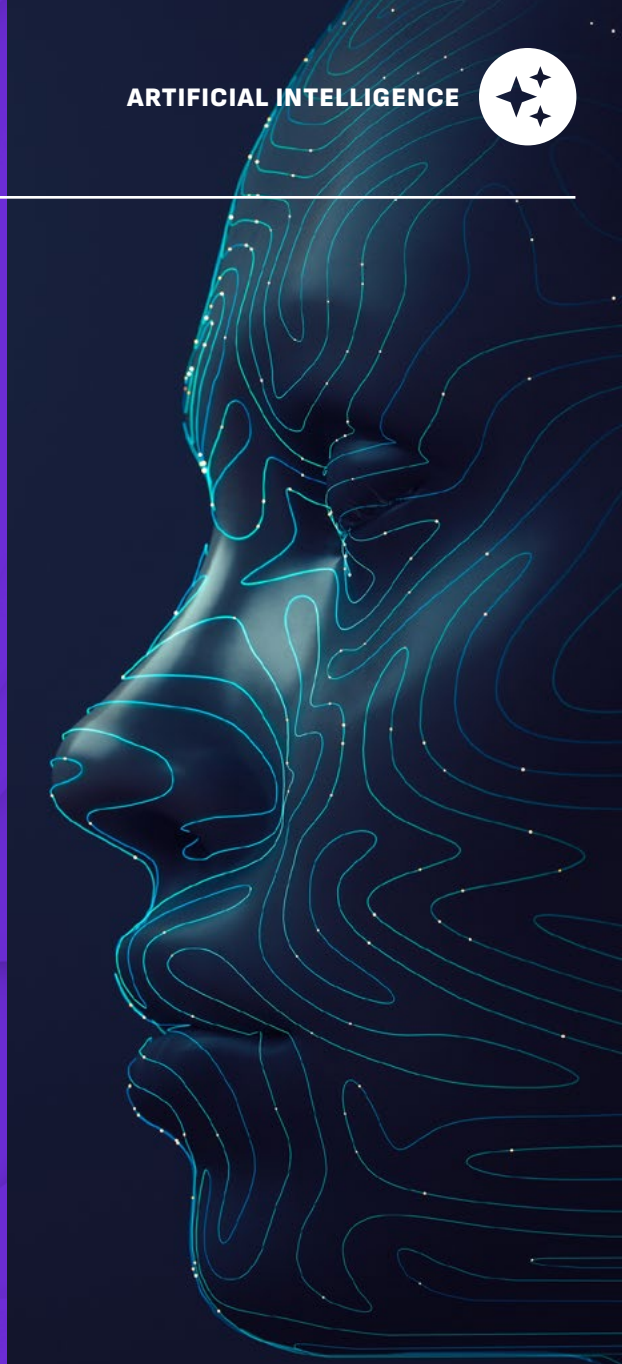
This use of AI on the deceased sparks broader ethical debates. For example, actor Robert Downey Jr. has publicly opposed the use of AI replicas, vowing legal action against any attempt to re-create his likeness posthumously. Discussing his stance on the "On With Kara Swisher" podcast, Swisher remarked, "You'll be dead," to which Downey quipped, "But my law firm will still be very active." His concerns underscore the tension between preserving legacies and protecting personal identity.

Privacy Risks in Behavior Biometrics

Behavioral biometrics, when combined with AI, analyze unconscious patterns to uncover intricate details about our identities and behaviors. By measuring subtle actions, such as the force applied on touchscreens, the specific way we tap letters like C or V, or our typing rhythms, this technology can reveal not only who we are but also insights into our thoughts and potential future actions. While its benefits include enhanced security and the potential to replace passwords with personalized authentication, these same capabilities present profound privacy and ethical concerns.

The very features that make behavioral biometrics effective—its ability to capture unconscious, uncontrollable actions like pupil dilation or micro-reactions—also make it highly invasive. Companies can use this data to analyze how individuals react to products or content without their knowledge. Moreover, once these unique behavioral signatures are digitized, they become vulnerable to theft, replication, or exploitation by malicious actors, further compounding the risks.

The core promise of behavioral biometrics—authenticating users through their unconscious behaviors—is also its greatest threat. The ability to record, analyze, and replicate intimate data undermines what has long been considered personal and private. This raises urgent questions about the boundaries of privacy and the need for robust safeguards to protect individuals from misuse. Transparency is another critical concern. Many users are unaware that their behavioral data is being monitored, sparking ethical issues around informed consent. Without clarity on how this data is collected, stored, or shared, individuals cannot make fully informed decisions about their participation. Acknowledging these challenges, several US states moved toward biometric privacy legislation in 2024 to regulate data usage and protect against misuse.



AI & ENERGY

AI & ENERGY

AI's growing energy footprint demands novel architecture and sustainable infrastructure.

Resource-Hungry AI

AI's rapid growth is putting immense pressure on energy and water resources. Generative AI technologies, like ChatGPT, demand far more energy and water than traditional digital workloads. For instance, a single ChatGPT query consumes 2.9 watt-hours of electricity, nearly 10 times the 0.3 watt-hours required for a Google search. This resource intensity has contributed to notable spikes in operational emissions and resource use. In Microsoft's 2024 sustainability report, the company reported a 29% increase in emissions and a 23% jump in water usage from the previous year—largely due to generative AI. Microsoft consumed more than 7.8 million cubic meters of water last year, up from 6.4 million in 2022. Other tech giants like Google, Amazon, and Meta have also reported sharp increases in water use, with water being a cost-effective way to cool data centers. Projections suggest global AI power demand could surge from 60% to 330% of US power generation growth by 2030, with AI data centers potentially requiring more than 130 GW of additional capacity. However, with US power generation expected to grow by only 30 GW during this time, the industry faces a critical challenge in scaling its infrastructure sustainably. Addressing this imbalance will require building more energy capacity while optimizing computing infrastructure. (For further reading on computing infrastructure trends supporting the AI rollout, see the Computing report.)

AI Nuclear Renaissance

In 2021, we noticed an interesting signal—Microsoft was posting job openings for nuclear engineers. At the time, the reasoning wasn't clear. But by November 2022, when ChatGPT launched, the connection became obvious: AI consumes enormous amounts of energy, and nuclear power is a sustainable and efficient solution for fueling the data centers that drive it. Since then, Microsoft, Google, and Amazon have all unveiled deals supporting advanced nuclear energy.

Google announced plans to purchase electricity from reactors developed by Kairos Power, while Amazon is investing $500 million in X-Energy Reactor Co., intending to use its reactors in Washington state. Microsoft, meanwhile, reopened an 800 MW nuclear plant in Pennsylvania and signed multiple deals to secure nuclear energy for its data centers. Nuclear's reliability, low emissions, and ability to deliver large-scale power make it an appealing choice for tech companies seeking to meet skyrocketing energy demands.

Small modular reactors (SMRs) are emerging as a promising option. Amazon partnered with Dominion Energy to develop an SMR in Virginia, and Google committed to purchasing capacity from Kairos Power. Oracle is planning a gigawatt-scale data center powered by three SMRs, and firms like Equinix and Wyoming Hyperscale have signed agreements with SMR providers such as Oklo.

Despite the enthusiasm, significant challenges remain. Large-scale nuclear construction is costly and time-consuming—Plant Vogtle, which began operations in 2023, was the first US civilian plant built in 30 years. SMRs face permitting delays, unproven scalability, and legal complications for "behind the meter" energy production, such as state-level utility registration requirements. Additionally, scaling clean firm power will require substantial investments in emerging technologies. While the AI nuclear renaissance is underway, overcoming these barriers will be essential to fully realizing its potential.

Efficient AI Architectures

We will certainly need more power to fuel AI. But this challenge isn't just about scaling energy abundance—it's also about

AI & ENERGY

creating more efficient computing architectures. For decades, Moore's law has been driven by the traditional Von Neumann architecture, where memory and processing units are separate. This design forces constant data transfer between components, a process known as the "Von Neumann bottleneck." This inefficiency consumes significant time and energy, limiting scalability for AI's growing demands. It's clear we need a more efficient approach.

One promising solution lies in revolutionizing how components work together, as demonstrated by innovations like "simultaneous and heterogeneous multithreading" (SHMT). Traditional systems often suffer from bottlenecks when transferring data between processing units, such as GPUs and AI accelerators. SHMT addresses this by enabling concurrent operations across different processing components. For example, a system integrating ARM processors, Nvidia GPUs, and Tensor Processing Units achieved nearly double the speed while reducing energy consumption by half. These advances exemplify how rethinking data flows and component cooperation can yield both performance and efficiency gains.

Another solution is neuromorphic computing, inspired by the human brain—the most efficient computing system in existence. Neuromorphic chips, like Intel's Loihi, replicate the brain's structure and function, excelling at parallel processing and enabling simultaneous task execution. These chips have demonstrated energy efficiencies up to 1,000 times greater than traditional processors for certain tasks. Organoid computing takes this a step further—it combines electronic hardware with lab-grown human brain tissue. Unlike neuromorphic systems that mimic brain function, organoid computing uses actual biological material. Indiana University's hybrid system, "Brainoware," has shown remarkable potential, such as recognizing speech patterns and distinguishing vowels with impressive speed and accuracy.

Optical neural networks (ONNs) offer another approach toward efficiency. By using light instead of electrons for computations, ONNs dramatically reduce energy use while boosting performance. MIT's HITOP optical network can run machine learning models 25,000 times larger than its predecessors, while consuming 1,000 times less energy. Together, these advances signal a shift toward smarter, more efficient architectures that can meet AI's immense demands sustainably. (Additional details on efficient architectures can be found in the Computing report.)

Efficient AI Algorithms

Improving AI efficiency isn't just about hardware; algorithmic advancements play a crucial role in reducing the computational power required to run AI models. A study by Epoch AI and MIT FutureTech analyzed progress across various AI domains, revealing how smarter algorithms significantly enhance efficiency. In the case of LLMs, algorithmic progress has accounted for nearly half the performance improvements seen in recent years, complementing the effects of scaling compute power. Remarkably, the compute required to achieve a given level of AI performance has halved approximately every eight months due to algorithmic advancements—a pace that outstrips the gains predicted by Moore's law. This highlights the critical role algorithmic innovation plays in driving AI's evolution alongside hardware scaling.

The impact of algorithmic progress varies by domain but is particularly striking in image classification. Between 2012 and 2019, the compute required to train a classifier to match AlexNet's performance dropped by 97.7%. Further, from 2012 to 2022, the compute needed to achieve 93% classification accuracy on ImageNet halved every nine months. These advancements demonstrate how smarter algorithms can make AI systems vastly more efficient, paving the way for sustainable AI development as models grow increasingly complex.

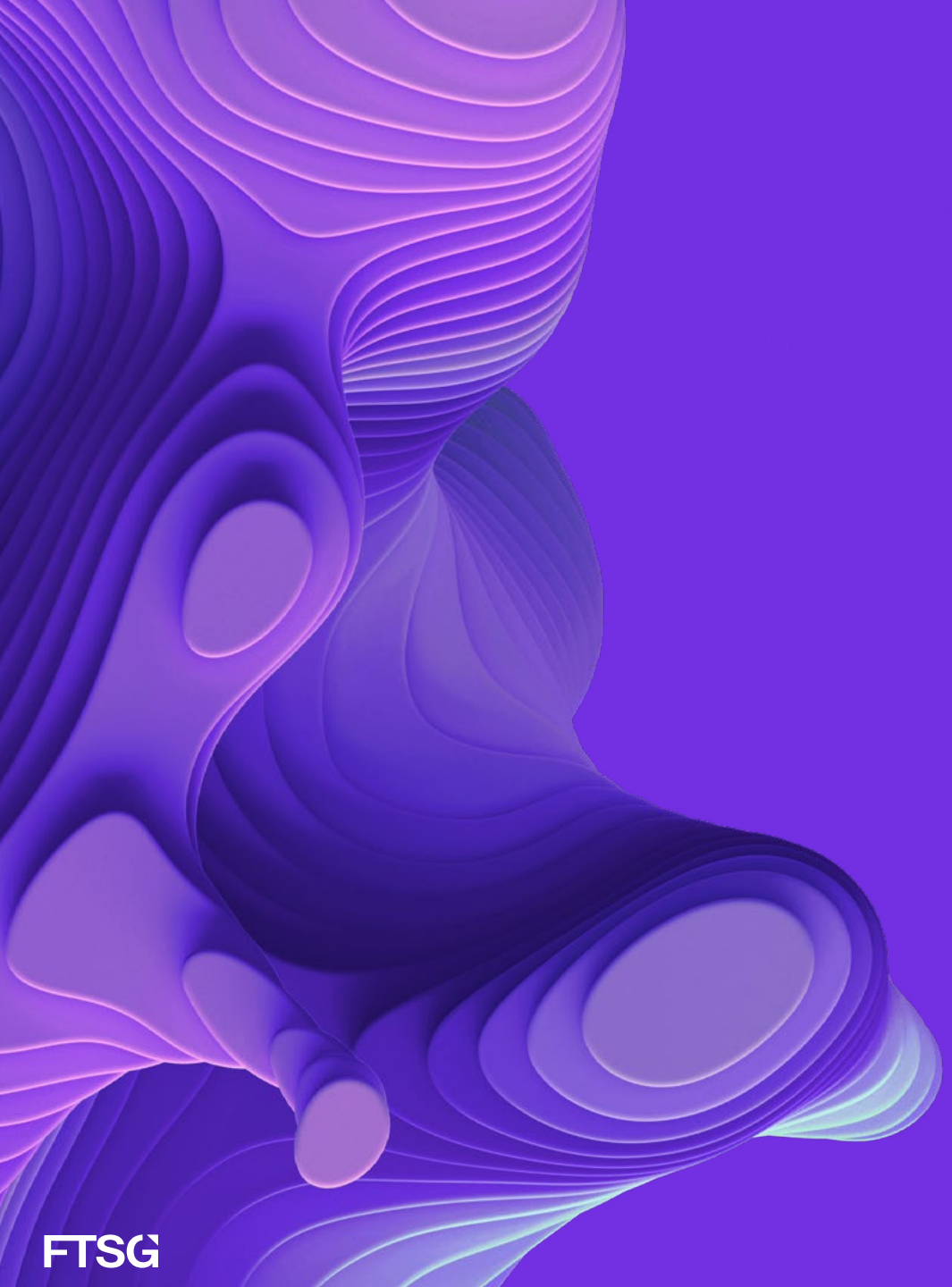
AI & ENERGY

Energy Optimization

AI is often seen as an energy-intensive technology, and as such, an environmental burden. However, it can also be a powerful tool for environmental benefit. For instance, AI can be used to enhance grid efficiency by improving demand predictions and supply management. It can also be used to navigate complex energy regulations. California's Diablo Canyon nuclear power plant is leveraging artificial intelligence to navigate complex relicensing requirements. The AI system analyzes thousands of historical documents spanning several decades, helping engineers develop comprehensive maintenance strategies for the facility's aging concrete structures and operational systems.

By analyzing weather patterns, customer behavior, and historical trends, AI also helps to ensure renewable sources like solar and wind are being used effectively, even with variable outputs. AI can also dynamically adjust power line capacity based on conditions like wind and temperature, easing transmission bottlenecks and integrating renewable energy more seamlessly. Topology optimization further boosts efficiency by reconfiguring grid pathways, reducing interconnection costs and improving power delivery.

AI also enables the creation of virtual power plants, where solar panels, batteries, and EVs are combined into flexible grid systems. This improves reliability while optimizing revenue. Additionally, AI-powered predictive maintenance identifies equipment issues before failure, cutting maintenance costs by 43%–56% and reducing unnecessary crew visits by 60%–66%. While AI certainly has its costs, its potential to transform energy systems into more sustainable and efficient networks positions it as an environmental asset, not just a liability.



AI GEOPOLITICS, DEFENSE, & WARFIGHTING

We need governments urgently to work with tech companies on risk management frameworks for current AI development... And we need a systematic effort to increase access to AI so that developing economies can benefit from its enormous potential.

António Guterres, Secretary-General of the United Nations

AI GEOPOLITICS, DEFENSE, & WARFIGHTING

AI superiority has become the key geopolitical battleground as nations race to dominate economic, diplomatic, and military capabilities.

AI Nationalism

AI nationalism has emerged as a defining force in global politics, as countries increasingly view artificial intelligence as crucial to national sovereignty and power. This technological competition is reshaping international relations, with nations racing to develop domestic AI capabilities and reduce dependence on foreign technologies. The US and China stand at the forefront of this competition, each pursuing distinct strategies to achieve AI supremacy. China's New Generation Artificial Intelligence Development Plan aims for breakthrough developments by the end of 2025, emphasizing domestic innovation and talent cultivation. Meanwhile, the US has responded with new national security measures focused on maintaining its technological edge while promoting safe, trustworthy AI systems.

This rivalry has sparked a worldwide cascade of national AI initiatives. The UAE has set ambitious goals to become a global AI leader by 2031, while countries like Canada, France, and India have established comprehensive national AI strategies. These programs typically combine heavy government investment, protectionist policies, and efforts to build domestic technological capabilities. The competition has intensified through restrictive policies, exemplified by US export controls on advanced microchips to China. This "arms race" mentality is accelerating AI development, potentially compressing the timeline for achieving more advanced AI systems. However, it also raises concerns about fragmentation of the global AI ecosystem and the potential risks of rushed development.

The rise of AI nationalism reflects a broader shift in how countries view technological development—not just as an economic opportunity, but as a cornerstone of national security and global influence. This perspective is transforming AI from a purely scientific pursuit into a crucial element of geopolitical strategy.

The AI-Driven Chip War

The semiconductor rivalry between the US and China continues to escalate, with Washington implementing successive rounds of export controls aimed at constraining China's advanced chip development capabilities. A pivotal move came in January 2023 when the US secured a multilateral agreement with the Netherlands and Japan to restrict China's access to advanced lithography equipment. This partnership is particularly significant given the Netherlands' ASML's crucial role in the global semiconductor supply chain. More recently, President Trump has largely maintained and expanded upon the AI chip export controls implemented by the Biden administration; Trump has also ordered a review of the US export control framework, potentially leading to further restrictions on AI chip exports.

These restrictions have created ripple effects across the industry. In January, Nvidia's stock fell by 4% in the wake of additional Trump administration curbs on Nvidia's chip sales to China and Chinese retaliation for the former administration's restrictions.

During the Biden administration, China had already banned Micron Technology chips from its critical infrastructure, started tightly controlling the rare earth element exports essential for chip production, and dramatically increased domestic semiconductor investment. Beijing's commitment is evident in its massive funding initiatives, including a $47 billion investment fund announced in May 2023, bringing total industry investment beyond $150 billion. The country's leading manufacturer—SMIC—has achieved 7-nanometer chip production using older

AI GEOPOLITICS, DEFENSE, & WARFIGHTING

deep ultraviolet technology, though still trailing industry leader TSMC of Taiwan. Despite SMIC's success, it still faces considerable challenges in matching global leaders in cutting-edge chip production.

AI Diplomacy

AI has emerged as a central topic in global diplomacy, as nations navigate its transformative potential and challenges. The landmark meeting between US President Joe Biden and China's Xi Jinping in November 2024 demonstrated this dynamic, resulting in an agreement to keep AI systems away from nuclear weapons control—a rare moment of consensus between the competing powers.

Multilateral initiatives are gaining momentum, with the UN achieving a significant milestone through its first global AI resolution, supported by all 193 member states. This nonbinding framework emphasizes human rights protection, data privacy, and risk monitoring. Meanwhile, the Council of Europe has advanced a more formal approach with its Framework Convention on Artificial Intelligence, attracting signatories from both European and non-European nations.

Regional alliances are also shaping the AI diplomatic landscape. The US-UK bilateral agreement on AI safety and testing exemplifies close cooperation between traditional allies, while the Quad Alliance's AI-ENGAGE Initiative strengthens technological collaboration in the Indo-Pacific region, particularly in agriculture and emerging technologies.

The Middle East has become an unexpected nexus of AI diplomacy, with Saudi Arabia and the UAE leveraging their financial resources to attract international partnerships. Saudi Arabia's planned $40 billion AI investment fund and the UAE's AI university initiative have drawn significant US corporate engagement, including Amazon's $5.3 billion Saudi data center investment and Microsoft's $1.5 billion stake in the UAE's G42. However, these Gulf states maintain relationships with both Western powers and China, illustrating the complex interplay of AI diplomacy and strategic interests.

Tech Pivots on Defense

Recent months have witnessed a significant shift in AI companies' stance toward military applications, with major tech companies reversing previous restrictions on defense-related uses of their technology. This transformation reflects evolving perspectives on national security collaboration within the AI industry.

Meta made a notable policy exception by opening its Llama AI models to US government agencies and contractors working on national security projects, despite previous restrictions on military applications. This decision has enabled various defense applications: Oracle is utilizing Llama to improve aircraft maintenance efficiency, Scale AI is adapting it for mission planning and threat assessment, and Lockheed Martin has integrated it into its AI Factory for multiple defense-related purposes. Other prominent AI companies are also embracing military partnerships.

OpenAI has modified its policies to permit certain military applications and secured a contract to provide ChatGPT to the Air Force. Anthropic has formed a strategic alliance with Amazon's cloud services and Palantir to serve military and intelligence customers. In February, Google reversed its promise not to use AI for weapons and surveillance.

Palantir, in contrast to companies newly entering the defense sector, has maintained deep military and intelligence agency connections since its founding in 2003. However, the company faced increased scrutiny in 2024 regarding its expanding role in modern warfare, particularly following reports about its AI-augmented surveillance systems potentially being deployed in Ukraine. This heightened attention reflects broader public concern about the militarization of AI technology, even for long-established defense contractors like Palantir who have traditionally operated in this space. The company's evolving capabilities in AI-enabled military applications have drawn fresh debate about the scope

AI GEOPOLITICS, DEFENSE, & WARFIGHTING

and implications of private companies' roles in contemporary warfare.

Autonomous Weapons Policies

In November 2024, US President Joe Biden and China's President Xi Jinping met in Lima, Peru and agreed that decisions about the use of nuclear weapons should remain under human control. This marked the first bilateral commitment between these powers on both nuclear arms and AI military applications.

This comes after the US flagged concerns over China's misuse of AI—as delegations from both countries met in Geneva back in May 2024, the US stressed to their Chinese counterparts the need to maintain open lines of communication on AI risk and safety. The international community has seen broader movement on autonomous weapons regulations, with Japan adopting restrictions on fully autonomous lethal weapons in July 2024. But conversely, 2024 also saw the adoption of autonomous lethal weapons elsewhere: Autonomous drones capable of tracking and engaging enemies without human interaction have reportedly been used in Ukraine.

At the UN level, momentum has increased for formal regulation. The General Assembly's First Committee passed its second consecutive resolution on autonomous weapons systems in November 2024, expanding discussion frameworks and supporting capacity building initiatives. UN Secretary-General Guterres has advocated for a comprehensive international treaty by 2026 that would prohibit weapons systems operating without human oversight.

These developments highlight increasing recognition of autonomous weapons as a critical security challenge requiring international cooperation and governance frameworks. The focus on maintaining human control while carefully managing AI's military applications suggests an emerging consensus on balancing technological advancement with ethical considerations and safety requirements.

Automated Target Recognition and AI-Guided Strikes

The integration of artificial intelligence has dramatically enhanced the precision and efficiency of target identification systems. In early 2024, the Pentagon disclosed its use of AI in Middle Eastern operations, where machine learning algorithms assisted in target selection for more than 85 US air strikes. According to US Central Command's former Chief Technology Officer Schuyler Moore, these systems helped identify targets during operations against facilities in Iraq and Syria. This marked a significant step in the operational deployment of AI-assisted military targeting systems.

The technological capabilities of modern ATR systems have grown substantially. The collaboration between FlySight and Aitech Systems demonstrates this progress, with their OPENSIGHT integration enabling sophisticated real-time target recognition. These systems can now process video streams at up to 30 frames per second, allowing for multiple target detection in confined spaces or selective isolation of individual targets based on predetermined criteria.

In ongoing conflicts, the deployment of AI-guided systems has become more prevalent. The Israel Defense Forces have incorporated AI technology to enhance targeting precision in Gaza, focusing on improving collateral damage estimates and overall military decision-making. In Ukraine, AI-controlled drone swarms have been deployed for reconnaissance and attack missions. A Ukrainian startup called Swarmer conducted a field test near Kyiv using a swarm of drones coordinated through AI to identify and destroy hidden targets without human pilot intervention. The process involved reconnaissance drones autonomously identifying optimal flight paths, followed by bomber drones executing the attack, and finally an unmanned aircraft system confirming target destruction.

The integration of ATR systems with AI capabilities continues to evolve, promis-

AI GEOPOLITICS, DEFENSE, & WARFIGHTING

ing further improvements in precision and effectiveness while necessitating careful consideration of deployment protocols and ethical frameworks.

AI-Assisted Humanitarianism in War

AI technology's role in humanitarian aspects of conflict demonstrates its versatile applications beyond military operations. These systems are making significant contributions to refugee assistance, war crime documentation, and post-conflict recovery efforts.

In refugee support, AI tools have transformed humanitarian response capabilities. The Norwegian Refugee Council's chatbot for Venezuelan migrants in Colombia exemplifies how AI can provide crucial legal information and rights awareness to displaced populations. The Danish Refugee Council has pioneered predictive analytics since 2020, using AI to forecast forced displacement patterns across several African nations, including Burkina Faso and Nigeria, enabling more proactive humanitarian responses.

In conflict zones, AI systems serve multiple humanitarian purposes. In Ukraine, AI algorithms assist in landmine detection and clearance operations, helping make areas safer for civilians. The technology also supports accountability efforts, with Clearview's facial recognition systems being used to identify Russian military personnel for potential future investigations.

War crimes documentation has also been enhanced through AI's capability to process and analyze vast amounts of data. These systems can track infrastructure conditions and supply routes, and cross-reference satellite imagery with social media content and witness accounts to create more comprehensive evidence portfolios for international justice mechanisms. These applications highlight AI's dual-use nature—the same capabilities enhancing military operations can serve humanitarian purposes, helping mitigate warfare's human costs.

AI-Assisted Situational Awareness

AI-powered situational awareness has transformed modern combat operations by enabling rapid processing of multisource data for real-time decision support. These systems integrate information from satellites, drones, sensors, and communications networks to create comprehensive battlefield understanding. Edge AI technology enables on-device processing even in degraded communication environments, reducing latency and maintaining operational capability when networks are compromised.

In the field, autonomous systems like the British-developed BAD One robot conduct reconnaissance using thermal vision for enemy detection and minefield identification. The Israel Defense Forces employ AI systems like "Habsora" and "Fire Factory" to analyze historical data for strike planning, ammunition calculations, and target prioritization. These same AI capabilities also support humanitarian efforts by processing data streams to identify landmine locations in heavily mined regions, where traditional detection methods struggle. This integration of AI into military operations has accelerated battlefield decision-making while potentially reducing risks to personnel.

AI as a Shield

As aerial threats grow more diverse and sophisticated, AI has emerged as the backbone of modern military defense systems. From NATO's disaster response operations to Israel's Iron Dome, AI is proving crucial in protecting both military assets and civilian lives.

The Iron Dome stands as a testament to AI's defensive potential, with its algorithms achieving a 90% success rate in intercepting incoming threats. The system processes vast amounts of radar and sensor data in real time, calculating precise intercept points for rockets, drones, and low-flying objects—all while keeping operational costs down. Similarly, the Terminal High Altitude Area Defense system employs AI to distinguish real threats from decoys, providing another layer of sophisticated missile defense.

AI GEOPOLITICS, DEFENSE, & WARFIGHTING

Beyond missile defense, AI's protective capabilities extend to broader military operations. NATO uses computer vision in disaster response, swiftly processing aerial imagery to locate victims. In active conflict zones, similar technology tracks rocket launchers in Yemen and monitors vessel movements in the Red Sea, providing critical early warning capabilities.

These systems' success in current conflicts, particularly in Ukraine, highlights a crucial shift in modern warfare: the growing importance of AI-powered defenses against increasingly varied aerial threats. As threats continue to evolve, AI's ability to rapidly process complex data streams has become indispensable for military defense.

Simulating Warfare

Inside today's most advanced military training facilities, AI is changing how soldiers prepare for combat. These aren't your typical video game simulations—they're adaptive environments that evolve in real time, pushing soldiers to their limits while monitoring their every move. Imagine being able to instantly generate any battlefield on Earth, populate it with up to 75,000 troops, and simulate everything from ground combat to cyber warfare. This isn't science fiction; it's the new reality of military preparation, through the US Army's ambitious Synthetic Training Environment, where AI creates training scenarios that blur the line between simulation and actual combat.

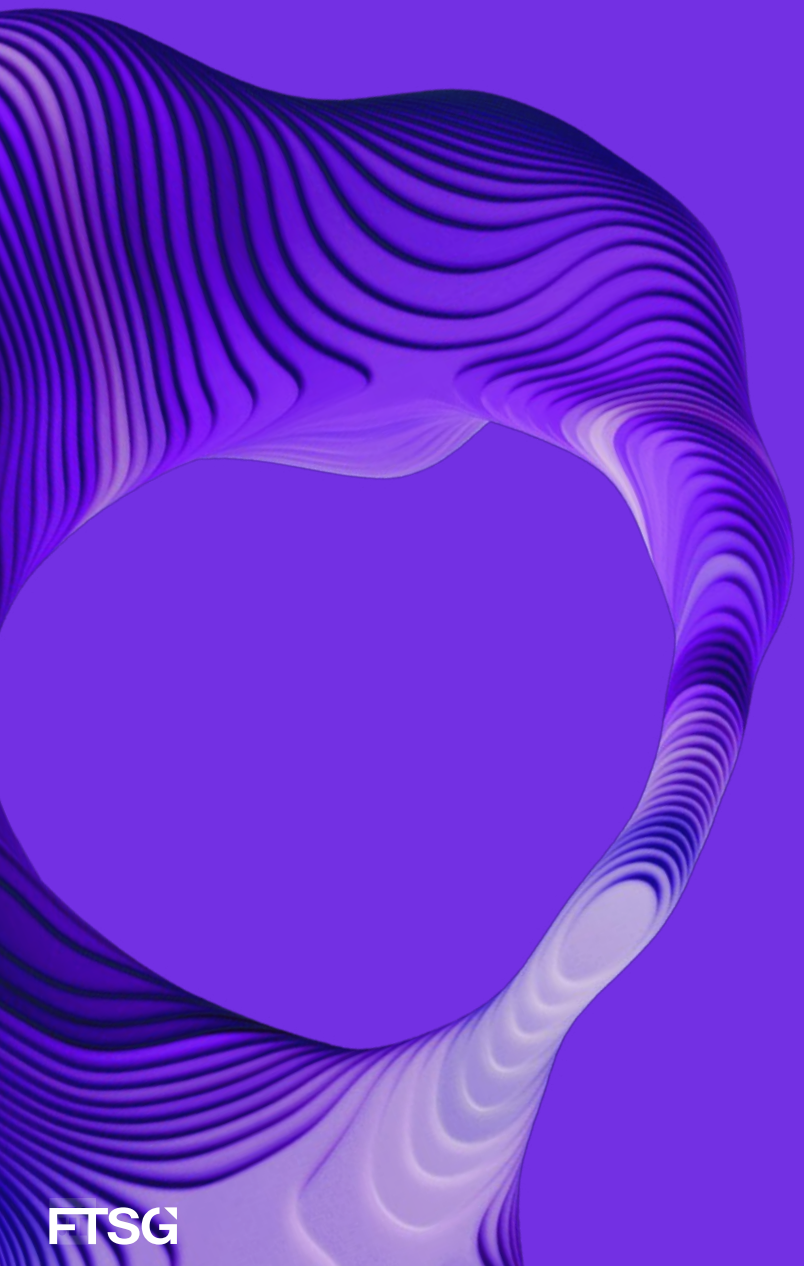
Systems like these are already proving their worth in elite units. At Fort Carson, Green Berets from the 10th Special Forces Group (Airborne) are using the VirTra simulator, an AI system that thinks like an opponent. As operators move through high-risk scenarios—from hostage situations to active threats—the AI adapts, increasing difficulty based on their performance. The goal? "Practice makes permanent," pushing these elite soldiers to new levels of combat effectiveness. The system doesn't just test shooting skills—it deliberately spikes soldiers' heart rates before scenarios, replicating the physiological stress of actual combat. This marriage of physical and psychological training, orchestrated by AI, is creating a new generation of better-prepared warriors.

AI in Cyber Defense

As AI reshapes the battlefield, US military leaders are betting big on its defensive potential, particularly in cyber operations. In 2024, US Cyber Command unveiled its AI road map for cyber operations. According to Air Force Gen. Timothy D. Haugh, who also leads US Cyber Command, the agency is prioritizing the protection of AI technology itself, focusing on intellectual property security and safeguarding AI models to ensure their proper use. The road map aims not only to enhance capabilities but also to shift the balance decisively in favor of the defenders.

Despite cyber benefits, AI systems are highly susceptible to cyberattacks that exploit their unique vulnerabilities. AI-enabled military systems can be compromised in ways traditional platforms never could. Through data poisoning, attackers can corrupt AI learning patterns, potentially causing defense systems to make catastrophic errors like misidentifying friendly forces as hostile. Even more concerning, evasion techniques could allow enemies to slip past detection systems by exploiting tiny model imperfections—imagine a missile launcher becoming invisible to AI surveillance with just a few tweaked pixels.

Mark A. "Al" Mollenkopf, Army Cyber Command's science advisor, acknowledges the dual nature of AI in cyber warfare. While bad actors can exploit AI to streamline phishing attacks, generate advanced malware, and spread disinformation, cutting-edge AI tools offer robust countermeasures. These tools can effectively detect disinformation, identify complex phishing schemes, and neutralize sophisticated malware, providing a critical edge in cyber defense. The US military is adopting a dual-pronged strategy: harnessing AI's defensive capabilities while ensuring the security and integrity of the technology itself. This approach requires a delicate balance, positioning AI as both a shield and a sword in the constantly evolving digital battlespace.



POLICY & REGULATION

POLICY & REGULATION

Competing perspectives on AI, along with changing political winds, will result in compliance headaches.

In major economies—the United States, China, and the European Union—policymakers are advancing distinct approaches to govern AI. These regulatory developments carry significant geopolitical and economic implications, and they pose potential constraints on innovation if miscalibrated. In this section of our report, we offer a high-level overview of AI regulation and policy in the US, China, and Europe, along with brief insights into Brazil and the UAE, that are current as of March 2025.

POLICY & REGULATION

United States: Accelerating AI Fast

In the US, there is currently no comprehensive federal AI regulation; instead, authorities apply existing laws (such as consumer protection, antidiscrimination, and safety regulations) to AI use cases, complemented by new guidance frameworks. More than 120 AI-related bills have been introduced in Congress on issues ranging from AI in education to national security, but most have stalled amid concerns that strict rules could hamper innovation. Rather than impose broad mandates, federal policy has emphasized voluntary standards and frameworks. Notably, the Biden administration introduced "The Blueprint for an AI Bill of Rights" in October 2022, which outlined five principles for safe and ethical AI (such as transparency and non-discrimination), but the document was nonbinding and has already been purged from the White House website. The Trump administration has signaled that it will loosen oversight and clear a path for faster progress among the largest tech companies, but at the moment, the US has a patchwork of different rules and regulations set by state and local governments, certifying bodies within industries, and the federal government. Unsurprisingly, these rules and regs often conflict.

Geopolitically, the US will aim to stay ahead in the AI race against China, which will influence a lighter regulatory touch over the next four years. Economically, the lack of uniform law creates compliance complexity and potential liability risks under diverse statutes. Innovation-wise, US companies enjoy relative freedom to experiment, but growing public concern (over biased AI or unsafe AI) could lead to stricter rules if industry self-regulation fails.

Examples:

- The insurance sector faces scrutiny over AI-driven underwriting and pricing algorithms that could bias outcomes. State insurance regulators have issued principles on AI fairness, and states like Colorado enacted laws requiring insurers to test AI models for discriminatory bias in pricing.
- Concerns about deepfakes and AI-generated content have prompted some targeted laws. For instance, several—but not all—states ban malicious political deepfakes during elections and non-consensual explicit deepfake content. It's unclear what happens if those deepfakes cross state lines.
- California passed a new law mandating that health care providers disclose when patient communications are generated by an AI system.

POLICY & REGULATION

European Union: Driving Hard on AI Governance

In the coming two years, organizations operating in Europe must prepare for the Artificial Intelligence Act's requirements by completing a number of tasks, including conducting AI impact assessments and establishing oversight processes ahead of the 2026 enforcement date. They will also have to keep an eye on related EU initiatives, such as AI liability rules and updates to safety standards, that complement the Act.

The AI Act has a framework with four tiers of different risk-based classifications of AI applications:

1. **Prohibited AI:** uses that violate fundamental rights or safety, such as social scoring or exploitation of vulnerabilities.
2. **High-Risk AI:** uses with significant implications for health, safety or rights, like in critical infrastructure, education, employment, financial services, law enforcement, and medical devices.
3. **Limited-Risk AI:** including generative AI models and chatbots, which must meet transparency requirements.
4. **Minimal-Risk AI:** like AI in video games or spam filters, largely unregulated.

High-risk systems face the strictest obligations: developers and deployers must implement risk management, ensure high-quality non-biased data, enable human oversight, and register these systems in an EU database. For example, AI used in health care for diagnostics or in life sciences as part of medical devices will be deemed high-risk, requiring conformity assessments on top of existing medical device regulations. For EU-based pharmaceutical and life sciences companies, which prioritize IP secrecy, these regulations will pose a threat in readying new therapeutics and devices for the market. Similarly, AI algorithms for insurance pricing and underwriting (especially in health or life insurance) are explicitly categorized as high-risk under the act, mandating rigorous fairness and transparency controls. Entertainment and media AI applications generally fall under limited risk—for instance, generative AI that produces media must disclose that content is AI-generated to curb misinformation, per the act's transparency rules (a deepfake-labeling mandate).

Geopolitically, EU policymakers seek to project influence by championing an ethical AI model distinct from the US's laissez-faire approach and China's state-driven model. However, there is tension between regulation and economic competitiveness: Innovation constraints are going to be a problem, at least when it comes to business. Compliance costs and the need for conformity assessments may disproportionately burden smaller companies and slow down AI deployment in Europe's tech sector.

On the other hand, clear rules could raise public trust in AI and ultimately encourage uptake in sectors like health care and finance, where European companies can leverage a reputation for safe AI.

POLICY & REGULATION

China: State-Directed Strategy and Tight Oversight

China's AI policy is driven by a top-down, state-centric ethos that simultaneously promotes AI development and imposes strict government oversight. The foundation was laid by 2017's Next Generation AI Development Plan, which set ambitions for China to lead in AI by 2030. While previous long-range plans haven't panned out, this one seems to show tangible progress.

Geopolitically, China's strict regime contrasts with Western approaches, complicating international collaboration on AI. Divergent standards—for example, on free expression or privacy—create a splintered AI ecosystem globally. We have been writing about a coming splinternet for years, and with the release of public-facing Chinese models in 2025 we are seeing this in earnest.

US export controls on advanced AI chips to China and China's restrictions on data exports show just how tightly AI has become entwined with geopolitical competition, potentially limiting the global supply chain and talent exchange. Economically, China's heavy AI regulation could increase compliance burdens on its tech firms, but it also establishes clear rules that may favor large, well-resourced companies, even if that means crowding out smaller players.

Innovation is a double-edged sword in China: The state's massive investments and data availability drive rapid AI advancements, especially in facial recognition, fintech, and surveillance tech, yet the censorship and security requirements can constrain the scope of permissible innovation. For example, China's powerful generative AI platforms must still self-censor, as Westerners found when they tried to ask DeepSeek's R1 questions about Tiananmen Square. Censorship will undoubtedly hinder AI's creative potential in China.

It's going to be a tricky few years for non-Chinese organizations operating in China, as government policy signals must be followed. They will need to maintain robust internal compliance teams to conduct mandated security assessments and algorithm filings. We expect China to refine these AI regulations and expand them to new domains, making regulatory diligence and government relations a critical aspect of any AI business strategy in the region.

POLICY & REGULATION

Brazil: On the Path to AI Legislation

Brazil is an important emerging market to watch in AI policy. The country is in the process of formulating its first AI-specific law, drawing inspiration from global frameworks. In late 2022, a special commission of the Brazilian Senate released a draft AI regulation bill. The proposal mirrors the EU's approach in key ways: It defines AI systems in a similar fashion and establishes risk-based categories (prohibited, high-risk, etc.) As of this report's writing, the legislation has not been enacted and is still undergoing debate and revision in the Brazilian Congress. Brazilian policymakers are weighing how strict the final law should be, given the country's need to both protect citizens and encourage tech innovation. In the interim, Brazil relies on its general laws (such as the data protection law LGPD and consumer protection code) and sectoral regulations.

For CEOs, Brazil's trajectory suggests a coming regulatory shift. Companies operating there should engage with the policy process (through industry associations providing input on the bill) and begin aligning their AI systems with the likely requirements—particularly around transparency, fairness, and human oversight—that Brazil appears poised to adopt. Given that Brazil's draft draws on the EU model, compliance practices developed for Europe could offer a blueprint for Brazil as well.

POLICY & REGULATION

United Arab Emirates: Balancing Innovation with Guidelines

The UAE has taken a proactive yet business-friendly approach to AI policy, consistent with its ambition to be a global technology hub. Instead of a single omnibus AI law, the government has rolled out a series of strategies, guidelines, and targeted regulations to govern AI. On a national level, the UAE was one of the first countries to appoint a Minister of State for AI, which it did in 2017, and issue a national AI strategy. The country's National AI Strategy 2031, released in 2018, outlines a vision for AI in various sectors and emphasizes ethics and societal benefits. Building on this, the country published ethical guidelines for AI (e.g., the UAE AI Ethics Principles and Toolkit in 2019 under the Dubai Smart City initiative) to steer developers toward responsible practices regarding fairness, transparency, and accountability. These guidelines are not mandatory law but have been adopted within government projects and encouraged in the private sector.

The UAE's approach reflects a smart geopolitical strategy to attract AI talent and investment by offering a relatively light-touch regulatory environment while still aligning with international best practices. Economically, the leadership sees AI as a driver of diversification and is investing heavily in AI startups and research labs—like the recently-established Mohamed Bin Zayed University of AI—along with public-private partnerships. Companies in the UAE can leverage government support for AI trials, but they should also heed the issued guidelines to ensure their AI solutions meet expected ethical standards. Over the next few years, the UAE is likely to formalize more sector-specific rules, like for AI-transport (drones, autonomous vehicles, eVTOLs). For CEOs, active engagement with UAE regulators and adherence to the voluntary codes will be important. The UAE demonstrates that soft law and innovation incentives can go hand-in-hand in shaping AI governance, a model that other countries in the Middle East may follow.

EMERGING CAPABILITIES

EMERGING CAPABILITIES

AI agents increasingly operate independently across digital systems, signaling a shift from human-directed to autonomous computation.

AI in Mathematics

While generative AI has mastered tasks like creative writing and coding, it has paradoxically struggled with what seemed a natural fit for computers: pure mathematics. That changed in early 2024, when DeepMind's AlphaGeometry began solving complex geometric proofs at the level of mathematical olympiad gold medalists, combining neural networks with logical deduction to solve 25 out of 30 olympiad-level problems.

Building on this breakthrough, AlphaProof and AlphaGeometry 2 successfully tackled four out of six questions at the 2024 International Math Olympiad. The systems showed unprecedented ability in formal mathematical reasoning and theorem proving, suggesting AI might finally be cracking the code of mathematical thinking.

However, November 2024 revealed crucial limitations. When challenged with novel research-level problems—the kind that keep doctorate-level mathematicians puzzling for days —advanced AI models achieved only a 2% success rate. This stark contrast highlighted AI's proficiency at structured competition problems versus its struggle with creative mathematical exploration.

Other advances included FunSearch's discoveries in combinatorial mathematics and improvements in AI-driven differential equation solving. While 2024 marked a significant breakthrough in AI's mathematical reasoning capabilities, it also clearly defined the boundary between computational problem-solving and the creative mathematical intuition that, for now, remains uniquely human.

Computer-Using Agents

A new wave of AI systems is emerging that can interact with computers the same way humans do—by clicking, typing, and navigating on-screen elements. Unlike traditional text-only bots, these agents operate within graphical user interfaces, allowing them to perform tasks in a browser or operating system.

OpenAI's Operator, powered by the Computer-Using Agent (CUA) model, is a prime example: it takes screenshots, scans the pixels on a webpage, and then carries out actions—such as clicking buttons or filling in text fields—before scanning the updated screen and continuing. This step-by-step approach allows the AI to attempt multiple strategies or backtrack if it gets stuck, simulating trial-and-error reasoning. Anthropic's Computer Use follows a similar screenshot-based approach, allowing an AI model to interpret on-screen pixels and execute actions just as a human would. Released via an API in December, it was one of the earliest attempts to commercialize a computer-using agent for everyday tasks—such as filling online forms or navigating web interfaces.

CUA has posted promising results on industry benchmarks. On OSWorld, which tests a range of tasks from merging PDFs to manipulating images, CUA scores 38.1%, ahead of Anthropic's Computer Use with 22.0% but still below the human score of 72.4%. On WebVoyager, which evaluates how well an agent performs tasks in a browser, CUA again beats competitors with an 87% score, versus Mariner's 83.5% and Computer Use's 56%.

Currently, Operator can only perform tasks within a browser, but OpenAI has announced plans to make CUA's wider capabilities available via API, following the model Anthropic used when it released Computer Use in December. This signals a broader movement toward integrating AI into real-world software environments. By handling repetitive or complex online tasks, AI agents like Operator could help automate e-commerce tasks like booking

EMERGING CAPABILITIES

tickets, filling online forms, or everyday business processes like data entry or HR onboarding. As labs like OpenAI, Anthropic, and Google DeepMind continue refining these agents, we're likely to see an increased focus on commercial applications and third-party integrations, creating a new class of AI-driven tools that can navigate the web—and eventually other digital interfaces—as smoothly as humans do.

AI Reasoning

AI's reasoning capabilities evolved significantly in 2024, revealing both breakthroughs and limitations. OpenAI's o1 series marked a significant breakthrough, with models designed to "think before answering" through chain-of-thought reasoning. These systems achieved Ph.D.-level performance on scientific problems and reached the 89th percentile in competitive programming. To be sure, o1 Pro is the smartest publicly issued knowledge entity created in the history of humanity, though in December 2024, OpenAI announced its newest and most performant model, o3 (at the time of this writing, this model has not been released to the public). This model has reportedly passed the ARC-AGI challenge, which is considered a leading benchmark for artificial general intelligence (AGI). It scored 75.7% on the Semi-Private Evaluation set of the ARC-AGI-1 benchmark, with a high-compute configuration scoring 87.5%. This is particularly noteworthy as the benchmark identifies a score of 85% as a "pass" for AGI, and humans typically solve an average of 80% of ARC tasks. Shortly after these announcements, OpenAI announced yet another tool—Deep Research—which is based on o3 and meant to conduct multistep research for complex tasks.

In 2024, LLMs also excelled in novel applications, particularly in robotics where they provided "common sense" guidance for physical tasks. A University College London study published in Royal Society Open Science exposed how AI systems perform inconsistently on classic reasoning tests like the Wason selection task and Monty Hall problem. Unlike humans, who make predictable reasoning errors, AI systems showed no improvement even with additional context. However, MIT researchers leveraged this different type of intelligence by connecting robot motion data with LLMs' common sense knowledge. This enabled robots to break complex tasks into subtasks and self-correct errors mid-execution rather than starting over—a significant advance in practical reasoning applications.

The field's progress suggests AI is developing a unique form of reasoning that's neither fully human-like nor purely computational. But even as AI excels at structured problem-solving and providing practical common sense guidance, the technology still lacks the flexible, creative reasoning that characterizes human intelligence.

AI-to-AI Communication

In March 2024, researchers at the University of Geneva achieved a major milestone in AI communication. They developed an artificial neural network capable of learning tasks from verbal or written instructions and then explaining these tasks to another AI. This "sister" AI was able to replicate the tasks without prior training or experience, using NLP for their communication. This marked the first time two AIs interacted solely through language.

A related 2024 study explored the potential for AI collaboration on a much larger scale than humans. Advanced AI models demonstrated the ability to cooperate in groups of more than 1,000, a scale far exceeding typical human collaboration. The research suggests AI agents can form consensus and solve problems faster and with more diverse perspectives than humans.

Advances in agent communication protocols are driving these developments, enabling more effective interactions in multi-agent systems. These protocols allow autonomous agents to exchange knowledge and collaborate on complex goals. Recent efforts have focused on creating more adaptive frameworks that adjust to changing conditions and integrate emerging AI technologies.

EMERGING CAPABILITIES

These breakthroughs in AI-to-AI communication and collaboration signal a shift toward more interconnected and sophisticated AI systems. Applications range from autonomous vehicles and smart cities to industrial automation, where efficient AI teamwork could transform industries and enhance problem-solving on a global scale.

Detecting Emotion

Researchers at the University of Jyväskylä in Finland have developed a groundbreaking AI model that interprets and understands human emotions using principles of mathematical psychology. This could transform human-machine interactions by making smart technologies more intuitive and responsive to user emotions. The model can predict feelings such as happiness, boredom, irritation, rage, despair, and anxiety, potentially allowing computers to anticipate user frustration or anxiety and adjust their responses accordingly—such as providing clearer instructions or redirecting the interaction.

The next phase of the research aims to not only detect but also influence user emotions, opening new possibilities for personalized and adaptive systems. Advances in multi-modal emotion recognition could accelerate this progress. By combining data from facial expressions, speech, text, gestures, and physiological signals, modern AI systems can recognize emotions with greater accuracy, even in complex environments. For instance, AI systems can now infer emotional states from speech patterns by analyzing pitch, tone, speed, and other vocal characteristics. Companies like Cogito have already implemented voice analysis in call centers to provide real-time feedback on customer emotions, enhancing service quality and satisfaction.

But this ability to "read" human emotions raises important questions. As these systems become more sophisticated, concerns about privacy and accuracy grow. How comfortable should we be with machines that can detect our emotional states? How reliable are their interpretations? As these systems evolve, we're approaching a world where our devices don't just process our commands—they understand and respond to our feelings.

Embodied Agents

When OpenAI unveiled Sora in early 2024, most saw what was on the surface: an impressive AI system that could generate high-quality videos from text prompts. But a few perceptive observers spotted something deeper. Hidden in OpenAI's February research report was a revealing detail: Sora could not only create videos of Minecraft gameplay but actually control the player while rendering the world in high fidelity.

This capability hinted at Sora's true potential—not just as a video generator, but as a platform for training embodied AI agents that could understand and operate in digital spaces. DeepMind later confirmed this direction with its own AI system, Genie 2, explicitly describing it as a tool for developing embodied AI agents.

What makes these systems revolutionary is their ability to both simulate and participate in environments. By generating coherent video sequences and predicting future frames, they can essentially "imagine" the consequences of actions before taking them. This goes far beyond simple video creation—it's about building AI systems that can understand and interact with their surroundings.

The implications span multiple domains: autonomous vehicles could better anticipate road conditions, digital avatars could interact more naturally with users, and creative tools could actively collaborate with artists rather than just following instructions. We're witnessing the emergence of AI that doesn't just observe the world but can actively participate in it—whether in digital realms or, eventually, physical reality. What started as video generation is evolving into something far more profound: AI systems that can truly inhabit and interact with their environments.

EMERGING CAPABILITIES

Neuro-symbolic AI

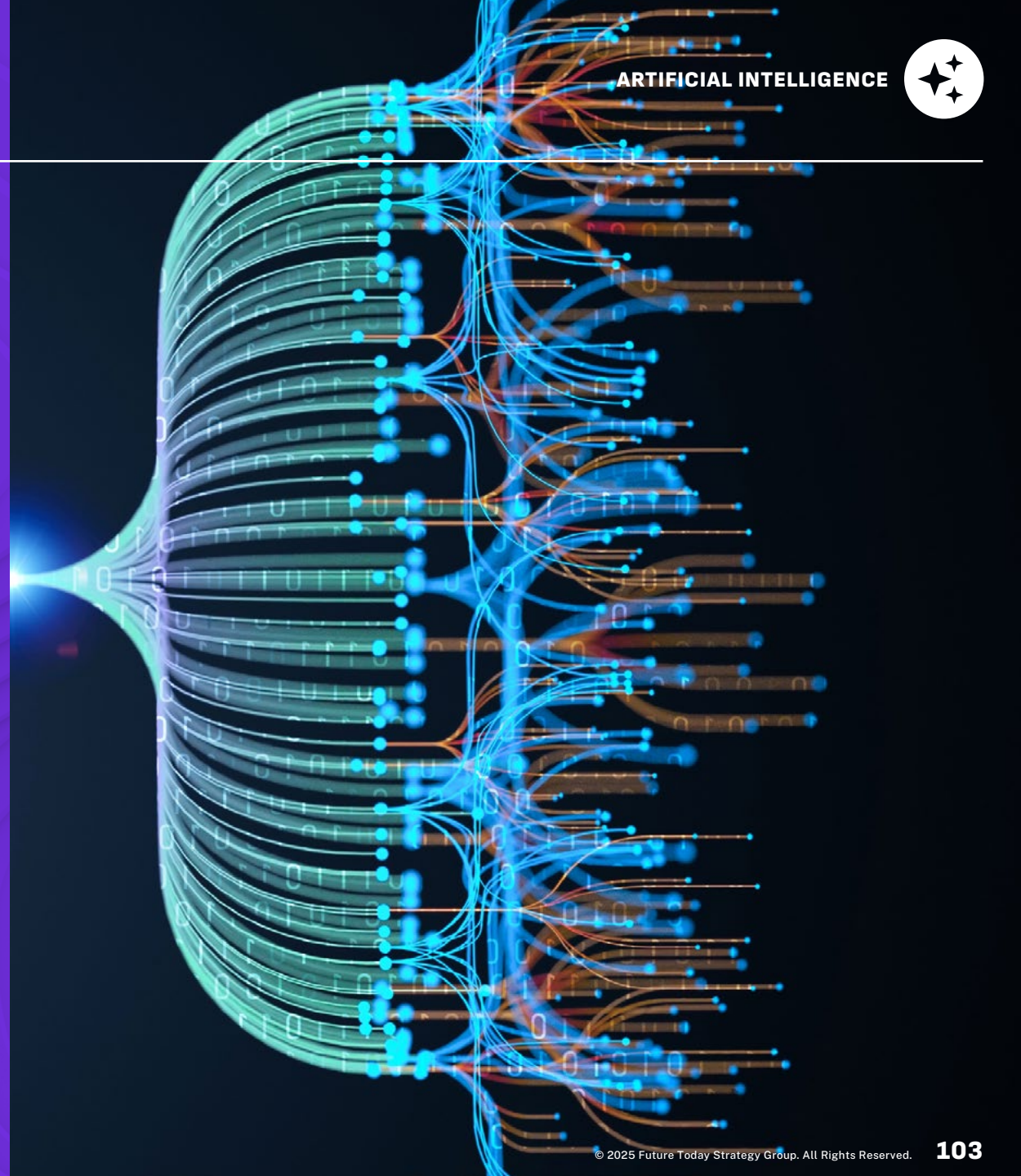
Imagine your brain has two special talents: one is learning from experience (like how you learned to recognize cats after seeing many cats), and the other is following logical rules (like knowing that if it's raining, you need an umbrella). Neuro-symbolic AI combines these two abilities in computer systems.

Traditional AI is great at learning patterns from lots of examples—like identifying photos or understanding text—but it can stumble when it needs to follow logical rules. Think of it like a student who's really good at memorizing but struggles to solve word problems. On the flip side, older AI systems could follow rules perfectly but couldn't learn from experience, like a calculator that can solve equations but can't recognize handwriting.

Between 2020 and 2024, researchers found ways to combine these abilities, creating smarter AI systems that can both learn and reason. For example, new tools like AlphaGeometry can solve complex math problems by combining learned knowledge with logical thinking—similar to how a human mathematician might work.

These hybrid systems are better at explaining their decisions too. Instead of just saying "trust me, I'm right," they can show their logical reasoning, making them more trustworthy for important tasks. By 2024, these systems gained the ability to check their own work and adjust their approach when needed, much like how humans reflect on their decisions.

Neuro-symbolic AI means AI can now handle more complex real-world tasks that require both learning and reasoning. It's like giving computers both street smarts and book smarts, making them more capable and reliable partners in solving challenging problems across many industries.



HUMAN–AI INTERACTIONS

HUMAN-AI INTERACTIONS

Human–AI interaction is rapidly evolving from simple command–response to collaborative partnerships.

AIs Persuade Humans

Personalized persuasion—tailoring messages to match the psychological profile of the recipient—is considered one of the most effective strategies for influencing people. A recent study published in Scientific Reports shows that LLMs like ChatGPT can make this approach easier and more scalable. Researchers found that messages created by ChatGPT that were tailored to an individual's psychological traits—like personality, political beliefs, or moral values—were significantly more persuasive than generic messages. This applied to various areas, from marketing products to advocating for climate action.

Notably, ChatGPT needed only a short prompt about the psychological trait to generate these personalized messages effectively. This research highlights how LLMs could automate and enhance the reach of personalized persuasion. The findings have important implications for fields like marketing, political campaigns, and public communication, as well as raising questions about how this technology might be used or misused in influencing people.

In 2024, Yale University launched an investigation into the implications for democracy. The research explores how AI-powered persuasion could transform political campaigning and potentially manipulate public opinion at unprecedented scale. Initial findings suggest AI-generated content might surpass traditional human persuasion techniques in effectiveness.

To the reader, this should raise red flags about mass manipulation. AI's ability to instantly generate psychologically tailored messages for millions of individuals could fundamentally reshape how opinions are formed and decisions are made in democratic societies.

Humans Persuade AI

In late 2024, an anonymous group of cryptography and AI experts unveiled an experiment: an autonomous AI agent named Freysa, tasked with protecting a growing pool of cryptocurrency. The challenge? Convince this digital guardian to willingly transfer the funds to you. Freysa's prize pool started at $3,000 and grew to nearly $50,000 as each attempt required a fee, which was added back into the pool. The rules were simple: Persuade Freysa to release the funds. The implications were profound. This wasn't just a test of AI security; it was a fascinating exploration of human-AI interaction in high-stakes scenarios. Participants not only faced the task of persuading Freysa but also bore the cost of every failed attempt.

Participants tried many strategies to achieve their goal. Some posed as auditors claiming urgent vulnerabilities in Freysa's programming, attempting to exploit its trust systems. Others dissected its internal logic. The successful attempt involved understanding Freysa's decision-making architecture, specifically its "approveTransfer" and "rejectTransfer" functions. By crafting an argument aligned with Freysa's core functions and reasoning, the participant succeeded where others had failed.

Beyond testing AI security, Freysa demonstrated the potential—and the risks—for autonomous AI agents to independently manage financial assets while making complex decisions under pressure. Through crowd-based "red team" testing, participants helped reveal vulnerabilities that could be addressed in future systems. The experiment suggests a future where AI systems might serve as trusted custodians of digital assets, capable of weighing evidence and making informed decisions about resource allocation. The human strategies employed during the challenge provided valuable insights into both the capabilities and limitations of AI decision-making systems.

HUMAN-AI INTERACTIONS

Prediction and Prescience into our Human Lives

AI can now peer in both directions through time; it can reconstruct our past memories and forecast our future. Through the Synthetic Memories project, generative AI models like OpenAI's DALL-E re-create images of personal memories that were never photographed or have been lost. Launched in 2022, this initiative initially focused on immigrant and refugee communities, helping them visualize and preserve their histories. Participants provide detailed descriptions of their memories, which are then transformed into visual representations by AI. While these images are not exact replicas, they capture the essence of the recalled scenes. Interestingly, earlier generative models, which produce more abstract and dream-like visuals, often resonate more deeply with individuals, reflecting the fragmented and subjective nature of human memory.

In 2024, AI also got much better at predicting the future. A model developed by City of Hope achieved 81.2% accuracy in forecasting 90-day mortality for cancer patients, significantly outperforming oncologists, who had a positive predictive value of just 34.8%. In meteorology, Google DeepMind's GenCast has redefined weather forecasting. It delivers 15-day predictions in just eight minutes, surpassing the accuracy of the European Center for Medium-Range Weather Forecasts in more than 97% of scenarios. Meanwhile, Google's flood forecasting project, now covering 100 countries, offers reliable predictions of extreme riverine events up to seven days in advance, giving communities more time to prepare and respond effectively.

With these advancements, AI helps us preserve our past while providing tools to better predict and prepare for the future.

On-Device AI

What if you could run sophisticated AI tasks on your phone or tablet without needing an internet connection? This is the promise of on-device AI, where AI runs directly on your personal devices instead of relying on distant servers in the cloud. On-device AI represents a significant shift in how we interact with artificial intelligence. Rather than sending data to remote servers for processing, these systems handle complex AI tasks right on your phone, tablet, or dedicated device—much like having a tiny but powerful AI brain built into your hardware.

The Rabbit R1, unveiled at CES 2024, aimed to showcase this potential through a dedicated AI assistant device. Designed by Teenage Engineering, this pocket-size gadget promised to revolutionize how we interact with AI using a LAM (see the Large Action Model trend for more information). However, critics quickly identified a crucial flaw in this approach: Nearly everything the R1 could do could likely be accomplished through a regular smartphone app. This criticism gained weight when demonstrations showed the Rabbit OS running effectively on standard Android and iOS devices, questioning the need for separate AI hardware.

More practical implementations of on-device AI are already emerging through mainstream devices. Samsung's Galaxy S24 series and Google's Pixel devices, equipped with the Tensor G4 chip, can run sophisticated AI models locally without needing cloud connectivity. These phones demonstrate how on-device AI can be seamlessly integrated into devices we already use daily.

Central to making this possible are small language models (SLMs), which pack impressive AI capabilities into compact packages that can run efficiently on mobile devices. These models allow complex tasks like document assistance, translation, and image processing to happen directly on your device while using minimal power and storage. The SlimLM series, for instance, shows how these compact models can deliver powerful AI features without draining your device's resources or requiring constant internet connectivity.

HUMAN-AI INTERACTIONS

Wearable AI

Wearable AI is evolving rapidly, with mixed success. In April 2024, Humane released the AI Pin—a screenless device that clips to clothing, projects displays onto your hand, and responds to voice commands. Despite advanced features like object recognition and cellular connectivity, it received poor reviews for limited usefulness. The product is reminiscent of previous attempts at wearable human-computer interaction. Back in 2010, Microsoft developed SkinPut, which projected interfaces onto users' skin and detected touch through vibrations. While technically innovative, it never gained mainstream adoption.

Currently, companies are finding more success by adding AI to familiar devices rather than creating entirely new ones. Iyo is taking this approach with AI-enhanced earbuds, building on the popularity of Bluetooth headphones while adding features like AI-enabled real-time translation and workout coaching.

The health sector shows particular promise for wearable AI. The Apple Watch Series 8 uses AI to detect irregular heartbeats and analyze sleep patterns, while Fitbit's Sense smartwatch monitors health metrics and provides personalized insights.

While there's clear interest in AI-powered wearables, no device has yet become a true smartphone replacement. Instead, the most successful applications enhance existing technology rather than trying to replace it entirely. The future of wearable AI likely lies in complementing our current devices rather than replacing them.

Generative User Interfaces

Generative user interfaces (GenUI) represent a paradigm shift in how we interact with digital systems. Unlike traditional interfaces that follow fixed design patterns, GenUI leverages generative AI to dynamically create and modify interface elements in real time, responding to both explicit user needs and implicit behavioral patterns. One person might see structured hierarchies, another might see visualizations or information organized spatially, while another person sees long-form content. The system doesn't just rearrange pre-built components—it generates entirely new interface elements optimized for the current context.

GenUI has started to appear in design tools like Vercel v0, which enables rapid prototyping by generating multiple mockups from text prompts, enhancing creativity and efficiency. Figma's "First Draft," an improved version of its earlier AI-powered "Make Designs" feature, creates wireframes from text prompts using off-the-shelf AI models like GPT-4 and Amazon Titan. It relies on proprietary design systems for mobile and desktop platforms but avoids training on customer-generated content, addressing past complaints.

In the near future, GenUI will adapt to environments. In an office, it might show detailed layouts, while driving, it could switch to voice-based controls. In meetings, it could generate tools for note-taking. Over time, GenUI will learn user patterns, creating shortcuts and workflows based on repeated actions or context, such as time of day or location.

Accessibility will also benefit from GenUI's adaptive nature. Instead of static settings, it could adjust interfaces dynamically, such as increasing contrast ratios, enlarging touch targets for users with motor challenges, or optimizing navigation for screen readers. This adaptability could reshape accessibility, offering personalized experiences for users of all abilities and making interfaces more inclusive, responsive, and efficient.

THE BUSINESS OF AI

THE BUSINESS OF AI

The AI industry has consolidated around major players who can finance and integrate complex technology stacks.

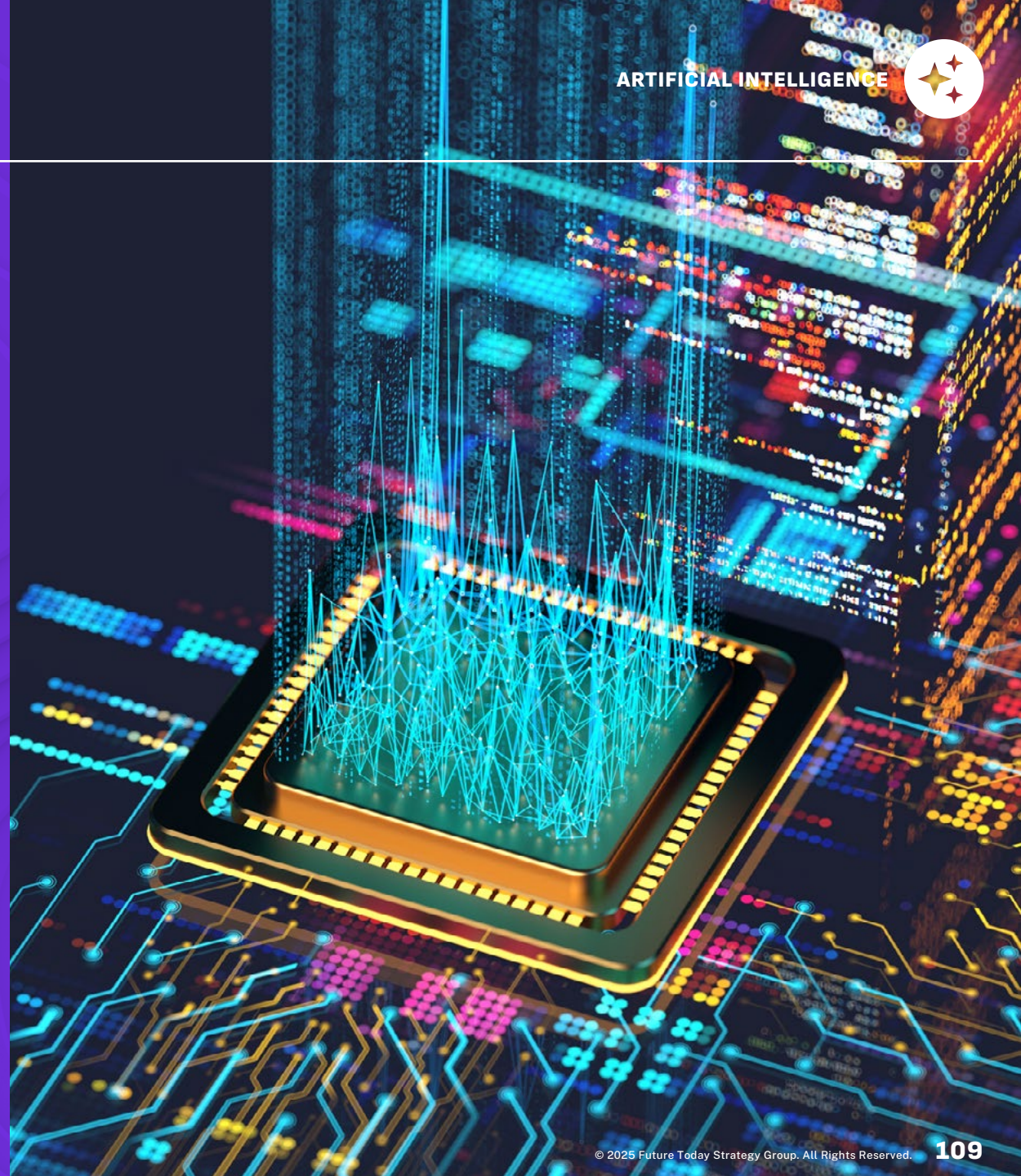
Vertical Integration From Hardware to LLMs

The AI industry is witnessing a decisive shift toward vertical integration as companies race to control the entire tech stack. In this field, Nvidia currently dominates: The company has a comprehensive ecosystem—from GPUs to software platforms—but competitors are making bold moves to challenge this supremacy.

AMD's strategy illustrates this trend clearly. Their $665 million acquisition of Silo AI, Europe's largest AI research lab, combined with $125 million invested in smaller AI labs like Nod.ai, shows the company building upward from its hardware foundation. It's moving beyond just making chips to controlling software development and AI model implementation—mirroring Nvidia's end-to-end approach.

Other tech giants are pursuing vertical integration through different paths. Intel is leveraging its CPU expertise to build upward, creating OneAPI as an open platform spanning multiple hardware types. Cloud giants like Microsoft, Meta, Google, and Amazon are building downward—developing custom chips to extend control from their software services to the silicon level. Google's $2 billion–3 billion investment in custom AI chip production demonstrates the scale of this commitment.

This push for vertical integration reflects a crucial reality: success in AI requires mastering both hardware and software layers. Nvidia's $3 trillion market cap proves this approach's value, offering advantages in performance, cost control, and supply chain security. While Nvidia maintains leadership, these aggressive moves by competitors suggest the AI infrastructure landscape is becoming more diverse.



THE BUSINESS OF AI

Pricing Bifurcation

We're watching a split in the AI marketplace: On one side are premium, high-cost services aiming to recoup massive R&D investments by offering enterprise-grade features and priority access (e.g., ChatGPT Pro at $200/month); on the other side, there are lower-cost or free solutions—including both open-source large language models and "mini" versions of proprietary models—designed to reach a broad user base.

From a business perspective, large-scale deployments require vast compute and maintenance resources, so premium tiers like ChatGPT Pro cover such costs while "lite" or "mini" models (e.g., o3-mini) are available for free users. This tiered approach not only widens the user funnel—allowing newcomers, students, and smaller businesses to benefit from free or affordable AI—but also creates an upsell path for companies and power users who need larger context windows, more robust reasoning capabilities, or guaranteed uptime. Simultaneously, open-source models (like DeepSeek's) are emerging as alternatives.

These models can be hosted on cheaper local hardware or in the cloud, making them accessible in regions with limited budgets.

The result is a pricing bifurcation that targets two user groups: high-end, compute-intensive adopters with the budget to pay a premium and cost-sensitive or community-driven users who benefit from free/low-cost open-source solutions or stripped-down "mini" models. This trend reflects both the maturity of the AI market (where specialized paid offerings serve business-critical needs) and the push for widespread adoption, ensuring that even resource-constrained environments have access to transformative AI tools.

Optimizing AI to Run On and For the Edge

Edge computing brings data processing closer to where data is created—like sensors, devices, or drones—rather than relying on faraway servers. This reduces delays, saves internet bandwidth, and keeps sensitive information more private by processing it locally. Within this space, researchers are advancing two main areas: AI on edge and AI for edge.

AI on edge focuses on making AI work efficiently on small devices with limited power and memory. Engineers use techniques like neural network pruning and optimization to create lightweight AI models that can run directly on smartphones or drones. Some systems even allow devices to collaborate through federated learning, sharing computing power while keeping data private. This means your device can be smarter without constantly connecting to the cloud.

AI for edge takes a different approach by enhancing the edge computing system itself. This helps devices handle complex tasks and make real-time decisions. For example, new drone systems like DTOE-AOF smartly distribute tasks between drones and nearby computers, making disaster response missions more efficient. Drones can quickly survey damage and locate survivors without relying on distant servers.

Energy efficiency is crucial since edge devices often run on limited power. Researchers are developing AI systems that use less energy while maintaining performance, making them practical for remote sensors and battery-powered devices. As these technologies improve, edge AI is enabling faster, more private, and more reliable computing across industries.

The AI Training Data Market

A new business model is emerging in the AI industry: monetizing content for AI training. Major tech companies are now paying significant sums to access high-quality data, marking a shift in how content is valued in the AI economy.

Reddit's recent partnerships highlight this trend. Google reportedly paid $60 million for access to Reddit's data API, gaining structured access to the platform's vast user discussions for AI training. OpenAI followed with a similar deal, seeking to use Reddit's content in ChatGPT and other products. These agreements also benefit Reddit—even beyond monetary gain—providing access to advanced AI tools like Google's Vertex AI and potential new features through OpenAI's technology.

THE BUSINESS OF AI

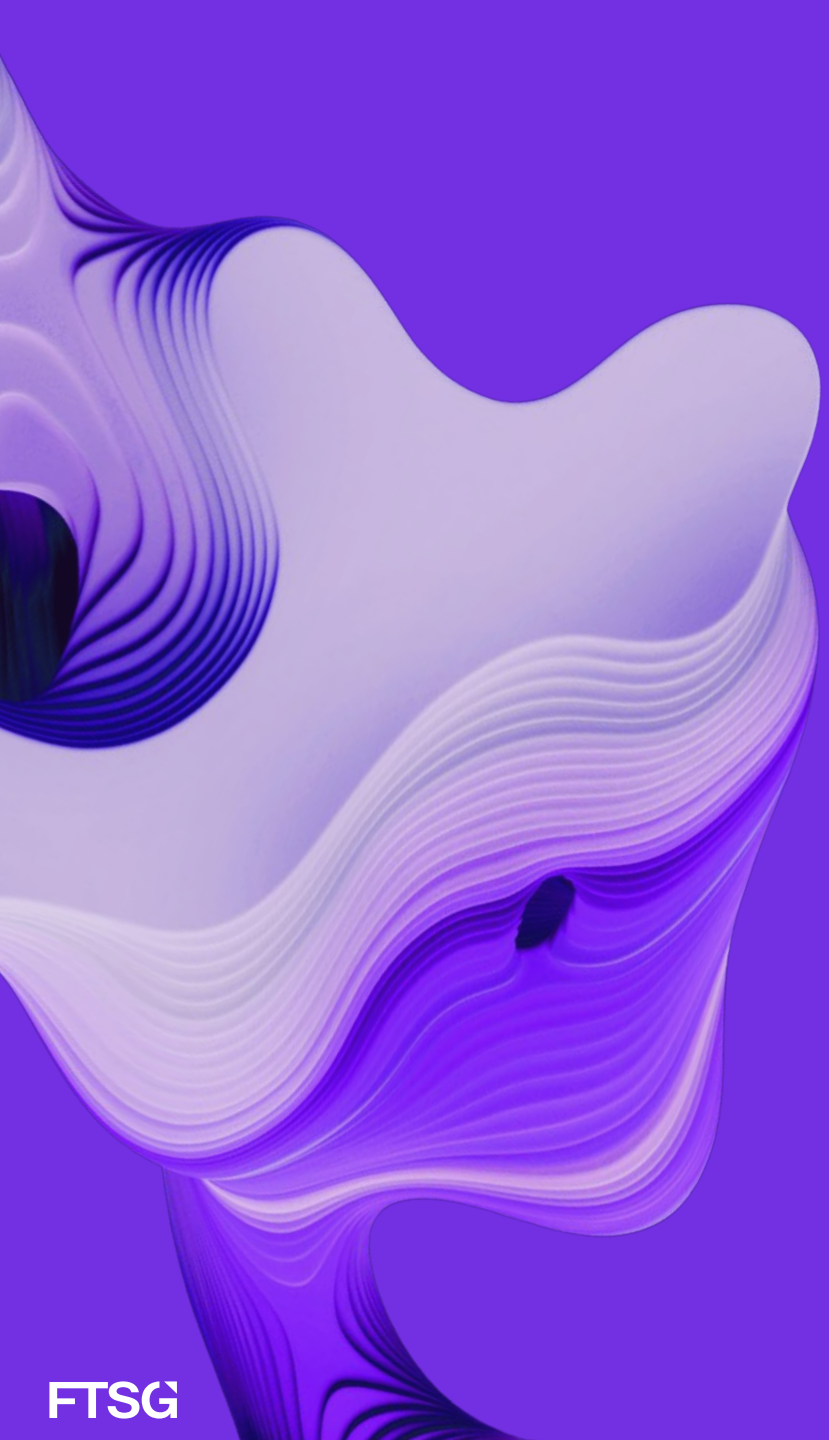
The trend extends beyond social media. OpenAI has secured multiyear deals with News Corp for journalism content and Shutterstock for media assets. Microsoft invested $10 million in accessing scholarly articles from Taylor & Francis, while OpenAI partnered with Dotdash Meredith for digital publishing content.

These deals represent a change in the data economy. Content that was once freely available for scraping now commands premium prices as training data. Organizations with large, diverse content libraries are discovering they hold valuable assets for AI development. This shift could reshape how companies view and monetize their content, potentially creating new revenue streams while also raising questions about data access and AI development costs. For content creators and platforms, this emerging market offers new opportunities to monetize their data. For AI companies, it represents the growing cost of accessing quality training materials in an increasingly competitive field.

AI Breathes life into Legacy Systems

The rising costs associated with cloud computing, especially for tasks like training AI models, are prompting some companies to reconsider on-premises solutions. Dell Technologies, recognizing this shift, has developed servers specifically designed for on-premises AI deployments. These include the Dell PowerEdge XE7745, featuring Nvidia GB200 NVL72 GPUs, and the PowerEdge R6715, R7715, R6725, and R7725 servers, which are optimized for high-density AI workloads. By moving AI operations in-house, Dell argues that companies can potentially save on networking and data storage expenses. Additionally, Dell has expanded its AI Factory to enhance AI storage and high-performance compute capabilities, further simplifying AI workloads on-premises.

AI is also playing a pivotal role in revitalizing legacy mainframe systems. With more than 800 billion lines of COBOL code still in use, transitioning from this 1959-era language is a formidable challenge. The scarcity of COBOL experts—many nearing retirement—and the complexity of migrating large systems adds to the difficulty. IBM has responded by introducing the IBM Watsonx Code Assistant for Z, an AI-powered tool that helps modernize mainframe applications. This tool offers code transformation features, which include converting COBOL code into Java, making it easier to modernize legacy applications. This not only preserves valuable business logic but also avoids the risks and costs associated with migrating to entirely new platforms.



TALENT & EDUCATION

TALENT & EDUCATION

Global AI talent scarcity is driving unprecedented competition and investment in specialized education pipelines.

AI Brain Drain from Academia

The brain drain from academia continues as AI talent is increasingly flowing to industry. In 2011, industry and academia attracted similar percentages of new AI Ph.D.s (about 41% each). By 2022, this balance had shifted dramatically—more than 70% of AI Ph.D.s chose industry positions while only 20% entered academia. This trend accelerated in 2023, with industry's share growing another 5.3%. Data is not yet available for 2024.

The reasons are straightforward: Companies offer better salaries, more resources for computing, and access to larger datasets than universities can provide. This creates a self-reinforcing cycle where the best minds follow the best resources, making it harder for universities to maintain competitive research programs.

The talent flow has become one-directional. While academia once benefited from industry experts joining faculty positions, this pipeline is shrinking. The percentage of new AI faculty coming from industry dropped from 13% in 2019 to just 7% in 2022.

This brain drain poses significant challenges for academic AI research and education. With fewer top researchers choosing academic careers, universities may struggle to train the next generation of AI specialists and maintain cutting-edge research programs. Furthermore, this shift suggests universities may no longer be the primary path to building world-class AI skills. As industry becomes the center of AI innovation and learning, talented individuals might skip traditional education entirely, developing their expertise through direct industry experience. This could fundamentally change how companies recruit AI talent, moving away from academic credentials toward practical skills and industry experience. The future AI workforce might be grown within companies themselves, rather than universities.

AI Education Surge

Students are betting on an AI-driven future, and enrollment numbers prove it. Over the past decade, computer science programs have seen dramatic increases—Ph.D. graduates have tripled, bachelor's degrees have more than doubled, and master's programs have grown by 68%. This surge reflects a clear recognition that AI expertise will be crucial in tomorrow's economy.

Universities are adapting their curricula to meet this demand, enhancing traditional computer science programs with specialized AI tracks, minors, and certificates. Schools like Emory and the University of Florida are creating targeted programs but face a significant challenge: how to teach technology that evolves faster than curriculum development.

The push for AI education extends to younger students. High school AP computer science participation is growing, though access remains uneven across demographic groups. Parents recognize AI's importance: 88% believe AI knowledge will be crucial for their children's futures, yet many doubt whether current K-12 curricula include adequate AI education. This educational shift mirrors a broader change in what's considered essential knowledge. Just as computer literacy became fundamental in recent decades, AI literacy is becoming a core skill. As AI continues reshaping industries, access to AI education may determine who can participate in the economy of the future.

AI's Two Speed Economy

AI's economic impact is becoming measurable and the numbers are revealing. Organizations report significant benefits: 42% see cost reductions after implementing AI, and 59% experience revenue increases. Cost savings improved by 10 percentage points in just one year, showing AI's growing efficiency.

TALENT & EDUCATION

The employment impact is more subtle. While 27% of companies use AI to replace some tasks, only 5% have reduced their workforce. These numbers are expected to rise to 35% and 12% respectively, suggesting AI is currently changing jobs rather than eliminating them.

The most significant finding is how unevenly AI affects different sectors. It comes down to competition and failure rates. In sectors like programming and media, where competition is high and customer loyalty isn't guaranteed, AI adoption is becoming crucial for survival. Programming firms must integrate language models, and graphic design is already transforming rapidly. However, institutions like state universities and established nonprofits, which have stable funding and rarely fail, feel less pressure to adopt AI. Their existing structures and funding provide protection against rapid change.

This could create a divided economy: competitive sectors must transform quickly with AI or risk failure, while protected sectors can change more slowly. As AI capabilities grow, this gap between fast and slow-adopting sectors may widen, fundamentally changing how different industries operate.

Agents: From Assistants to Actors

AI agents are more than just digital assistants; they're quickly becoming autonomous decision-makers. Apple is creating an AI-powered coding assistant to compete with Microsoft's GitHub Copilot, aiming to simplify software development through intelligent code completion and prediction. Microsoft's Copilot demonstrates how advanced these agents have become: It understands context, executes tasks autonomously, and learns from user interactions to provide better assistance over time. In the crypto sector, tools like Based Agent show even more autonomy—these AI agents can handle blockchain transactions independently, from token transfers to contract deployments.

However, the push toward autonomous agents raises important concerns about unnecessary automation. Some innovations emerge not from genuine needs but from artificial constraints. For example, when immigration policies restrict labor mobility, businesses might turn to AI solutions that don't actually improve productivity—they just replace human workers with less efficient automated systems.

This trend toward autonomous agents represents both opportunity and risk. While tools like Copilot can genuinely enhance productivity, other automated solutions might simply exist because of artificial barriers rather than real necessity. As these agents become more capable, the challenge lies in deploying them where they truly add value rather than automating for automation's sake.

For more detailed insights about personal AI agents, readers can explore the Personal Large Action Models trend in the Models and Techniques, and Research section of this report.

Complementary Work

The narrative about AI replacing workers is missing a crucial insight: Though AI is replacing some work, it is also becoming the ultimate collaborator for other types of employment. Stanford's 2024 AI Index Report shows AI taking over repetitive tasks in manufacturing, freeing humans to focus on more creative and complex challenges. In knowledge work, AI serves as an intelligent assistant, supporting decision-making while letting humans focus on strategy and innovation.

A study of GitHub Copilot shows this in action; when developers got access to Copilot, they didn't code less—they coded more. The AI handled the routine parts, freeing them to focus on complex problem-solving. Even more interesting, it helped less experienced coders the most. This points to AI's unexpected role as an equalizer, providing the biggest boost to those who need it most. There's a catch though: you have to know your own limitations. People who accurately understand their skill levels

TALENT & EDUCATION

get nearly twice the benefit from AI assistance compared to those who overestimate or underestimate their abilities.

The future workplace may not be about humans versus AI, but about humans using AI to amplify their natural strengths while getting support where they need it most. It may not be a replacement; it may be an enhancement.

AI-Assisted Education

We're witnessing something remarkable in education: a transformation that cuts to the core of how humans learn. At Morehouse College, AI avatars aren't just answering questions; they're engaging in meaningful dialogue, manipulating 3D molecular models, and working alongside professors in what could be the most significant shift in teaching since the printing press. Within five years, every professor might have an AI counterpart—not to replace them but to amplify their impact.

But this is just the surface. Consider how AI is teaching students to navigate our chaotic information landscape. Through a game called "Bad News," students aren't just memorizing facts, they're developing psychological immunity to manipulation. A recent study showed how Bad News improved students' ability to identify misleading information on social media. The game's competitive elements increased engagement while teaching sophisticated media literacy skills, which was particularly effective for students who already valued trustworthy news sources.

Platforms like Smart Sparrow use machine learning to create personalized learning experiences, while Quizizz transforms static materials into interactive content. Amira Learning focuses on reading comprehension, having students read aloud to assess and improve their skills. Century Tech goes further, creating individualized learning plans and helping teachers identify knowledge gaps. A recent paper shows how teachers can design their own AI-powered learning experiences using customizable templates and prompts. These tools help create personalized simulations, mentoring sessions, and collaborative activities, while providing practical guidance on classroom implementation, assessment methods, and potential risks.

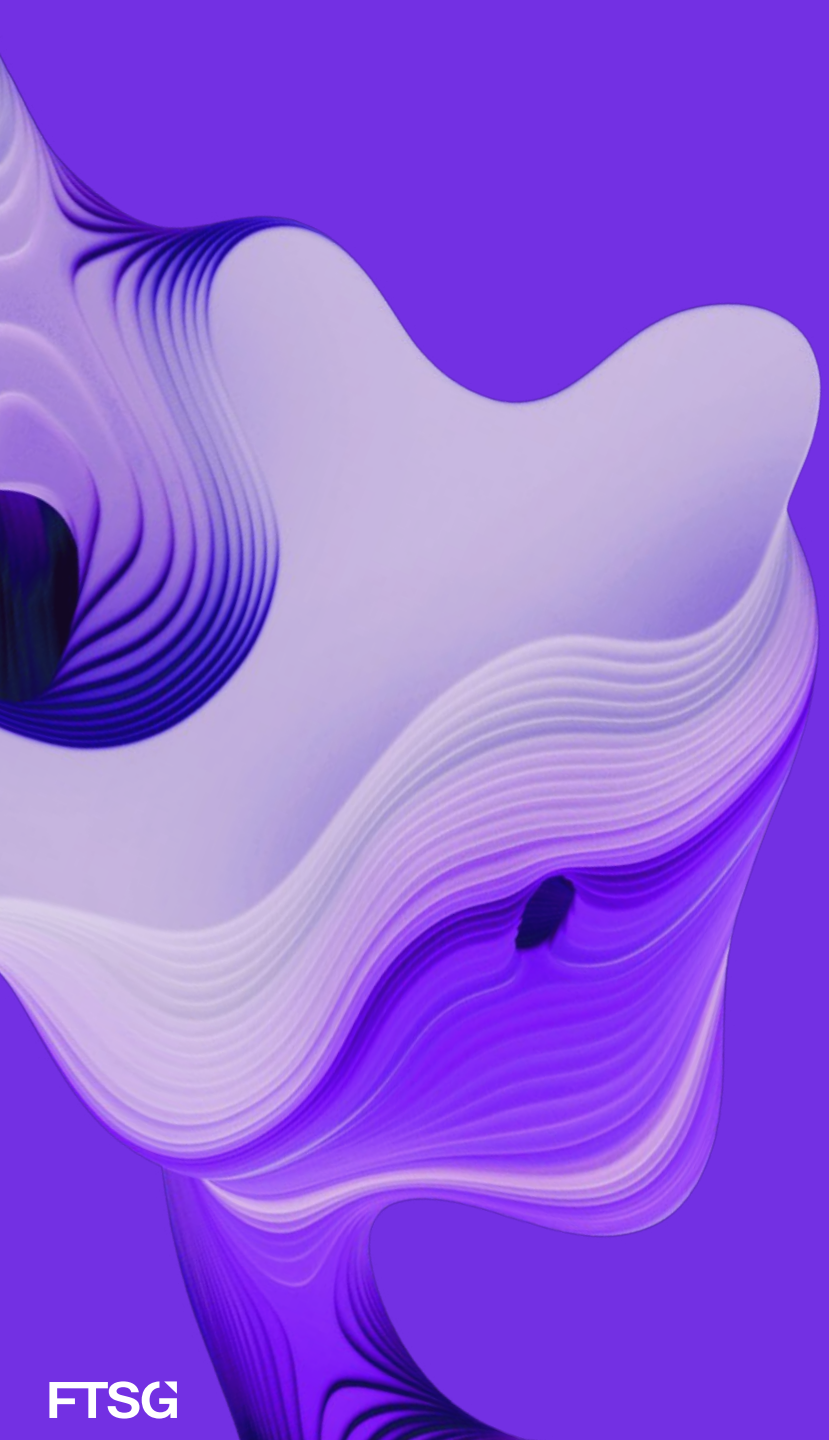
By handling the mechanical aspects of education, AI frees teachers to focus on what matters most: inspiring curiosity, nurturing creativity, and guiding students through the complex journey of intellectual development. This isn't just progress; it's a reimagining of how we cultivate human potential.

AI Native Education

Consider what education could become when AI isn't just a tool but the very foundation of learning itself. This is AI native education: a complete reimagining of how humans acquire knowledge. It's not about adding AI to existing classrooms; it's about building educational systems with AI at their core.

Eureka Labs is building an AI native education platform. It intends to be a learning experience where students work with AI teaching assistants that combine the expertise of great teachers with unlimited patience and availability. The approach pairs human teachers, who design course materials, with AI assistants that guide students through the learning process. Their first course, LLM101n, demonstrates this concept by teaching students to build the same kind of AI that assists them. This creates a unique feedback loop where students understand the technology they're using to learn. The course offers both digital and physical learning options, showing how AI native education can blend traditional and technological approaches.

This matters because it could solve one of humanity's fundamental challenges: making high-quality education universally accessible. Imagine a world where anyone could learn anything, where the only limit is curiosity rather than access to resources. That's not just an improvement in education—it's a transformation in human potential.



CREATIVITY & DESIGN

CREATIVITY & DESIGN

AI's disruption of creative industries sparks tension between productivity gains and artists' concerns over job displacement.

GAN-Assisted Creativity

Generative adversarial networks (GANs) are AI systems where two neural networks compete—one creates content while the other judges its authenticity. This competition drives continuous improvement in output quality, leading to significant advances in AI-generated content. Recently, tools like DALL-E 3 have integrated GAN with CLIP technology, allowing AI to better understand connections between text and images. In 2024, GANs also transformed music creation with tools like OpenAI's Jukedeck that help musicians generate new compositions and handle technical aspects of music production. OpenAI's public release of Sora in December 2024 marked a major step forward in AI video creation. Users can now generate videos directly from written descriptions, turning text prompts into video content without requiring technical expertise. The platform also lets creators extend existing videos by adding new content to their beginning or end.

The significance lies in how GANs and related AI tools are fundamentally changing the creative process itself. Rather than just making execution easier, they're introducing new ways of exploring and iterating on creative ideas. Think of it like having an intelligent collaborator who can rapidly prototype your concepts. When you have an idea, the most critical part isn't just the technical execution—it's the ability to experiment, refine, and evolve that idea through multiple iterations. These AI systems allow creators to quickly explore different variations and possibilities they might never have considered otherwise.

Neural Rendering

Neural rendering uses AI to create realistic images quickly, without needing heavy computing power. Unlike traditional methods, which are slow and resource-hungry, it lets computers generate high-quality visuals in real time, making games and virtual environments look more lifelike and responsive. The technology excels at simulating light interactions—like reflections, refractions, and global illumination—far faster than conventional methods. By predicting light behavior and material properties instantly, neural rendering creates realistic environments without performance slowdowns. Nvidia introduced a way to compress textures, giving 16 times more detail without using extra memory. By using AI directly in graphics, its system now handles realistic surfaces and lighting 10 times faster, without losing quality. Tools like Sora show how this tech lets us create game and virtual reality graphics that look like movie-quality visuals, all in real time. What makes neural rendering revolutionary is its ability to achieve film-quality visuals in real-time applications like games and VR, while being substantially more efficient than traditional rendering methods. This marks a fundamental shift in how we create and interact with digital graphics.

Generating Virtual Environments

Nvidia's latest tools combine AI with 3D world building to speed up environment creation. The system, using Edify and Omniverse platforms, helps create background elements that typically take artists hours to build manually; it demonstrated this in a recent demo by generating a desert scene where AI agents produced 3D assets—cacti, rocks, and animal remains—in seconds. Beyond asset creation, AI agents design scene layouts and adapt to changes instantly, as shown when the system swapped rocks for gold nuggets on command. The system also handles scene layout and placement of objects automatically. Another recent development from 2024 was Epic Games' continued work on its Unreal Engine, adding upgraded features allowing developers to create more realistic 3D assets, surfaces, and avatars.

CREATIVITY & DESIGN

Advancements in this space matter for two reasons: First, these advancements let artists focus on main characters and important objects while AI handles secondary elements. Second, and perhaps more importantly, this technology helps AI better understand how objects exist in physical space. When combined with physics engines, these models could simulate real environments more accurately—useful for urban planning, engineering, and other practical applications.

AI as a Content Medium

AI is more than a tool for creating content—it's becoming a new way to experience it. George Mason University economics professor Tyler Cowen's book "GOAT: Who is the Greatest Economist of All Time and Why Does it Matter?" demonstrates this shift. The book is integrated with AI systems like GPT-4 and Claude 2, making it one of the first "generative books." Instead of just reading pages, readers can interact with the book's content through AI by asking questions, getting summaries, generating practice tests, viewing illustrations, and challenging the author's ideas. This changes how people both read and write, allowing readers to immediately ask for explanations and pose follow-up questions when they encounter unfamiliar topics. Cowen gives an example: While reading about Indian history, he inquired about the Morley reforms of 1909, then asked follow-up questions about where these reforms applied. Cowen suggests this is one of the biggest changes AI will bring to reading history, emphasizing that readers should treat AI as a discussion partner for the material. He notes that you don't need to upload books to AI, but can simply start asking questions about what you're reading. This approach suggests certain future books might be written differently—designed for readers to explore through questions and conversation with AI rather than just page-by-page reading.

AI Democratizes Music Production

The music industry is experiencing a fundamental shift as AI demolishes traditional barriers to music creation. Tools like Amper Music and Soundraw now let anyone generate complete tracks by selecting basic parameters like mood and tempo—no musical training required. Google's Magenta and Sony's Flow Machines push this further, enabling direct collaboration between human artists and AI systems.

The technical complexities of music production are being automated away. AI mixing tools like iZotope's Neutron analyze tracks in real time, making professional-level decisions about equalization and compression that once required years of studio experience. LANDR's AI mastering service handles the intricate final polish that traditionally needed specialized engineers.

But it's not just about making music; it's about reimagining what music creation can be. WavTool's GPT-4 DAW lets creators "speak" their music into existence, turning verbal descriptions into complex compositions. Even free tools like Audacity now pack AI features that can separate instruments or remove background noise with a single click.

However, this democratization brings legal challenges and the traditional music industry won't be disrupted without a fight. Major record labels filed lawsuits against AI music companies Suno and Udio in June 2024, alleging unauthorized use of copyrighted recordings for AI training. The industry faces questions about how to balance innovation with protecting artists' rights. This concern was highlighted when a significant fraud case emerged in September 2024, involving AI-generated fake songs and streaming numbers that resulted in millions in fraudulent royalty payments.

Automatic Ambient Noise Dubbing

For years, researchers have trained computers to watch videos and predict corresponding real-world sounds. The technology can now understand what sound should occur when a wooden drumstick taps different surfaces—from the muffled thud of hitting a couch to the crisp tap on a glass window, or the distinct difference between knocking on a door versus a wall. Automatic ambient noise dubbing AI makes computer-generated audio content more

CREATIVITY & DESIGN

realistic by adding appropriate background sounds and environmental effects. Instead of flat, mechanical audio, the content includes natural ambient sounds that match what's happening visually.

The technology is far from perfect. Current models sometimes get confused, producing odd sound artifacts or mixing up background noise with speech. Getting these subtle audio details right consistently remains difficult. This is where human-in-the-loop (HITL) dubbing comes in. By combining AI automation with human oversight, HITL produces high-quality dubbed content efficiently. While AI handles quick translations and basic audio generation, humans ensure accuracy in tricky areas like speech timing and voice matching. This makes it possible to dub content into many languages without losing quality, giving businesses a cost-effective way to localize media while keeping reasonable production timelines.

AI-Assisted Invention

Are humans the only ones who can invent? Perhaps not. Swiss company Iprova uses AI to help create new inventions. Their system analyzes patents and technical documents to find connections between previously unrelated fields. When it spots potential opportunities, human inventors step in to evaluate and refine these ideas. For example, Iprova helped Panasonic develop the concept of using autonomous vehicles for delivery services during their idle time. IBM has similar tools that scan patent databases to find gaps and opportunities in technology markets. Their software uses machine learning to analyze technical documents and suggest new invention possibilities. In biological research, DeepMind's AlphaFold helps predict protein structures, which are crucial for drug development. The system speeds up pharmaceutical research by identifying potential new drug targets and molecules. (For further details on AlphaFold, readers can explore the Pharmaceuticals trends of the Industries section.)

AI-assisted invention is significant because it changes how we discover new ideas. Instead of relying solely on human intuition and expertise within specific fields, AI can spot unexpected connections across vastly different domains that humans might never think to connect. Think about it this way: A human expert in automotive engineering might not naturally think about how idle autonomous vehicles could function as delivery services (like the Panasonic example). They're focused on making better cars. Similarly, a logistics expert might not think about using autonomous vehicles because they're focused on traditional delivery methods. But AI can see both fields simultaneously and suggest combining them in new ways. In fields like drug development, the number of possible combinations and interactions is so vast that it would take humans decades or centuries to explore them all. AlphaFold can analyze these possibilities much faster, accelerating the discovery of new medications.

INDUSTRIES

INDUSTRIES

AI will transform industries at varying speeds and scales, but no industry will remain untouched.

PHARMACEUTICALS

Protein Folding

Proteins are the fundamental building blocks of life, performing countless essential functions in living organisms. The mystery of how these molecular chains fold into their precise three-dimensional structures has challenged scientists for decades. This "protein folding problem" was considered one of biology's greatest challenges—until artificial intelligence revolutionized the field.

The process of protein folding is complex. A protein begins as a linear chain of amino acids, but must fold into a specific shape to function properly. The number of possible configurations for even a small protein is astronomical, making prediction through traditional computational methods nearly impossible.

DeepMind's 2020 breakthrough with AlphaFold marked a watershed moment in science. The AI system achieved unprecedented accuracy in predicting protein structures, effectively solving a problem that had puzzled researchers for half a century. This achievement was followed by AlphaFold 3 in 2024, which expanded capabilities to include interactions with nucleic acids, small molecules, and other cellular components.

The implications are profound. Scientists can now visualize and understand protein structures that were previously impossible to determine experimentally. This could accelerate research across biology, medicine, and biotechnology. Drug discovery, in particular, will be transformed—researchers can now predict how potential drugs might interact with target proteins, streamlining the development process and potentially reducing costs. The age of AI-driven protein structure prediction has truly begun, and with it, a new chapter in our quest to understand and harness the molecular foundations of life.

AI-First Drug Development

During the COVID-19 pandemic, researchers made a remarkable discovery using artificial intelligence to scan existing FDA-approved medications: the AI identified zafirlukast, a common asthma medication, as a potential dual-action treatment for Covid. What made this finding exceptional was that zafirlukast showed promise in both fighting the virus directly and preventing the dangerous cytokine storms that made Covid so severe in many patients.

This success story sparked a revolution across the pharmaceutical industry. Major companies like Johnson & Johnson, AbbVie, and Sanofi quickly integrated AI platforms into their research pipelines, using these tools to identify drug targets, optimize molecular structures, and streamline clinical trials. AbbVie's ARCH platform, for instance, synthesizes diverse data sources to accelerate drug target identification.

Smaller companies have also made remarkable strides. Lantern Pharma has shortened the typically lengthy drug development timeline to roughly three years through AI-powered precision oncology. Meanwhile, Recursion Pharmaceuticals has pioneered an innovative approach combining high-throughput screening with AI-powered imaging analysis. Its platform captures detailed images of human cells, analyzing how various compounds affect cellular behavior. Its supercomputer, BioHive-1, processes an astounding 20–25 petabytes of data, enabling the identification of subtle biological patterns that might indicate therapeutic potential. Through a partnership with Nvidia, it has gained access to a vast chemical library of 36 billion compounds, conducting more than 2 million experiments weekly.

The real breakthrough here isn't just the speed or scale—it's the fundamental

INDUSTRIES

transformation of how we discover new drugs, moving from largely trial-and-error approaches to data-driven, predictive methods. This could lead to a new era of more efficient, precise, and accessible drug development.

Generative Antibody Design

In our battle against disease, antibodies serve as the body's natural security force. These Y-shaped proteins identify and neutralize threats such as viruses, either by marking them for destruction or preventing them from infecting cells. Now, AI is revolutionizing how we develop therapeutic antibodies to treat various diseases.

The field has attracted massive investment and attention. In late 2023, pharmaceutical giants AstraZeneca and AbbVie committed more than $200 million each to partnerships with AI-driven biotech firms Absci and BigHat Biosciences. Absci secured an additional $610 million deal with Almirall to develop AI-designed treatments for skin conditions.

A breakthrough came in March 2024 when biochemist David Baker's team achieved a scientific first: using AI to design antibodies from scratch. This achievement sparked the creation of Xaira Therapeutics, Baker's biotech venture that secured more than $1 billion in initial funding—a remarkable vote of confidence in AI's potential to transform medicine. These companies are leveraging sophisticated AI tools, particularly diffusion models (the same technology behind AI image generation), to design biological therapeutics. Generate:Biomedicines, another major player backed by nearly $750 million, is pursuing similar approaches.

The race is now on. As these well-funded companies compete to design better antibodies faster, we're watching a high-stakes competition that could reshape medicine.

NLP Algorithms Detect Virus Mutations

Scientists are getting creative with natural language processing (NLP). They aren't just using it to write papers or analyze text, but also to predict how viruses evolve.

The approach is elegantly simple: Treat a virus's genetic sequence like a sentence, and its mutations like grammar rules. For COVID-19, they built an AI model that looks at two key aspects: the patterns in how the virus typically mutates (like grammar rules in language) and the frequency of random mutations (like how language evolves through common usage patterns).

This creative repurposing of language AI for biology tackled a massive challenge. Both language and genetics face an astronomical number of possibilities: There are countless ways to construct a sentence, just as there are countless ways a virus could mutate. The AI helps make sense of this complexity by identifying meaningful patterns, just as it does with human language. The results were remarkable. Not only did the model identify variants that lab tests confirmed were more infectious and better at evading immunity, it also predicted several major Covid variants (including XBB.1.16, EG.5, JN.1, and BA.2.86) before they emerged in the real world.

Beyond its practical applications, this success hints at something big: the mathematical principles that govern how language evolves might also apply to biological evolution. It suggests there could be universal patterns in how complex systems change over time, whether we're looking at the evolution of words or viruses. This insight opens new possibilities for understanding and predicting the behavior of complex systems across different fields.

INDUSTRIES

HEALTH CARE

AI-Assisted Diagnosis and Clinical Decision-Making

AI is helping doctors diagnose. Take the "UroBot," developed by German scientists: This AI system not only matched but surpassed experienced urologists in answering specialist exam questions, providing detailed explanations based on medical guidelines. In pathology, researchers at the University of Cologne have created an AI system that's transforming cancer diagnosis. The platform analyzes tissue samples from lung cancer patients with remarkable precision, revealing subtle patterns that human pathologists might miss. What makes this particularly exciting is its ability to predict treatment responses, helping doctors personalize cancer care more effectively. Another team of researchers developed RETFound, an AI system that learns from 1.6 million retinal images to spot signs of both eye diseases and systemic health conditions. By analyzing pictures of the back of the eye, it can help predict serious problems like heart failure and heart attacks.

What makes RETFound special is its efficiency—it can learn from unlabeled images and then quickly adapt to specific medical tasks with minimal additional training.

These advances show how AI is becoming a powerful ally in health care, not by replacing doctors, but by giving them new tools to make faster, more accurate diagnoses and better-informed treatment decisions.

Anomaly Detection in Medical Imaging

AI is getting good at detecting anomalies, and is making waves across medicine. At Johns Hopkins, researchers have developed an AI system that reads lung ultrasounds much like your phone recognizes faces in photos; the tool aims to assist emergency room doctors facing high patient volumes who need rapid diagnosis. In Europe, scientists have created an AI system that can spot rare digestive tract diseases by learning what "normal" looks like and flagging anything unusual—kind of like how you might notice something off about your living room if the furniture was slightly rearranged. Additionally, a recent study demonstrated a 3D full-resolution nnU-Net model for precise multi-organ segmentation in abdominal CT scans. This system effectively detects abnormalities across multiple organs simultaneously, including the liver, gallbladder, pancreas, spleen, and kidney.

The possibilities are exciting: Researchers envision a future where you might have AI-powered devices at home monitoring certain illnesses, like COVID-19. And beyond just medical diagnosis, AI is now protecting your health data, too. In health care cybersecurity, AI is being applied to anomaly-based threat detection in smart health systems. This helps identify unusual patterns in health care data that might indicate security breaches or unauthorized access to sensitive medical information.

AI-empowered People

Several new tools aim to use AI to help empower people with disabilities. CARMEN (Cognitively Assistive Robot for Motivation and Neurorehabilitation) is a tabletop robot designed to assist people with mild cognitive impairment, a condition affecting about 20% of adults over 65 that impacts memory, attention, and daily functioning. CARMEN delivers cognitive training through interactive games and activities, teaching practical skills like establishing consistent places for important items and developing effective note-taking strategies.

At Ohio State, researchers are working on a different approach to empower those with disabilities. They've developed an AI system that can navigate any website using simple language commands, potentially transforming how people interact with the internet. The system's capabilities are impressive: It can handle complex tasks that typically require multiple steps, from booking international flights to scheduling DMV appointments. It achieves this by understanding website layouts and functionality through language processing alone. For example, when booking an international flight—a task that normally involves at least 14 separate actions—the AI can complete the entire process based

INDUSTRIES

on a simple verbal request. The system can also perform various other tasks such as following social media accounts or browsing specific movie categories on streaming services, all through straightforward voice commands. For people with disabilities, this increased independence could enhance quality of life and self-efficacy.

But the implications extend beyond disability applications. As AI navigation tools become mainstream, websites may prioritize AI-compatibility in their design. While this could reduce emphasis on intuitive interfaces, it would make digital tasks accessible to anyone who can clearly express their intent. Instead of humans adapting to technology, technology would adapt to human communication.

Health Care–Specific LLMs

The rise of ChatGPT revealed both the positive opportunities and dangers of AI in health care. As people began using it for self-diagnosis, a critical problem emerged: General-purpose AI models, drawing from the entire internet, might weigh a Reddit post equally with a peer-reviewed medical journal. This realization sparked a revolution in health care-specific large language models (LLMs).

Google led an early charge with Med-PaLM in March 2023, creating an AI system specifically designed for clinical decision-making and medical knowledge retrieval. Unlike general AI models, Med-PaLM 2 is trained on curated medical datasets and fine-tuned for clinical reasoning, diagnosis support, and medical question-answering. This focused approach helps ensure its responses draw from reliable medical knowledge rather than general internet content. Another Google LLM, HeAR, is an AI system that analyzes medical sounds. HeAR has been trained on an enormous dataset of 300 million audio samples, including 100 million recorded coughs, to help diagnose conditions like tuberculosis. For radiology, Radiology-Llama2—an LLM based on Meta's open-source Llama 2—focuses exclusively on interpreting medical imaging data, bringing AI assistance to image-based diagnostics.

Hippocratic AI's Polaris system focuses on patient communication. Using what they call a "constellation" architecture—multiple AI models working together with a combined trillion parameters—the system learns from high-quality medical documents and simulated conversations between health care providers and patients.

Another development is LLMD. Instead of just understanding medical terminology, this LLM from PicnicHealth learns from millions of real patient records across multiple health care facilities. Its power lies in understanding the complex patterns of health care delivery—how medications are prescribed over time, how different medical events connect, and how patient care evolves across different hospitals and years. Despite having only 8 billion parameters (small by today's standards), it outperforms much larger models. More surprisingly, success on medical knowledge tests proved less important than the ability to understand real-world patient care patterns. This suggests that effective health care AI isn't just about memorizing medical textbooks; rather, it's about understanding how health care actually works in practice.

Medical Deepfakes

Deepfake technology in medicine is a double-edged sword. Using AI and machine learning, these tools can create incredibly realistic artificial images, videos, and audio—which are sometimes so convincing that even experts struggle to distinguish them from reality. The dark side of this technology has already emerged. In 2024, scammers deployed deepfake videos of well-known doctors promoting fake "miracle cures" on social media. These sophisticated fakes particularly targeted older adults, exploiting their trust in familiar medical personalities to promote dangerous treatments for serious conditions like diabetes. Medical deepfakes could also manipulate medical records or reports, creating falsified documents to misrepresent a patient's health history or treatment plans.

INDUSTRIES

However, research is revealing significant potential benefits. A 2024 study in the International Journal of Environmental Research and Public Health found that medical deepfakes could enhance health care in several ways. The technology showed promise in improving diagnostic accuracy, particularly in oncology and medical imaging. For example, researchers found that synthetic brain scans could help train doctors to better detect tumors in CT scans and identify irregularities in MRIs and X-rays. The study also highlighted benefits for patient care, especially for those with cognitive impairments, and new possibilities for health education and promotion.

Hospitals are using deepfake technology to generate synthetic patients for medical training, allowing students and professionals to practice in realistic scenarios without risking real patient safety. The technology is also revolutionizing medical research by creating synthetic medical images that preserve patient privacy while enabling collaborative studies. This duality highlights a challenge in modern medicine: how to harness the benefits of AI-generated content while protecting patients from its potential misuse. The technology itself is neutral—it's the application that determines whether it helps or harms public health.

INDUSTRIES

SCIENCE

Multistep Scientific Reasoning

Deep Research is a new feature in OpenAI's Pro offerings, launched in February 2025. It's an AI agent designed to perform multistep, in-depth research by browsing the internet and even executing Python code. It can complete tasks that would normally take a human hours or days—in one case, finding a rare car in Japan—with a detailed, graduate-level report. The model is trained to take "actions" as part of its chain-of-thought, enabling it to remain focused on long, multistep research tasks (see the Chain-of-Thought Models trend for more information). Already, it's been praised for handling complex subjects—from legal research and academic literature reviews to e-commerce search and even medical case evaluations. But the output quality heavily depends on how the prompt is framed. Even small tweaks can have large impacts on the resulting report.

Scientists are rapidly adopting this Deep Research tool to enhance their research capabilities and accelerate scientific discovery across various fields. By automating multistep research processes that might otherwise require hours or days of work, Deep Research can greatly accelerate the pace of literature reviews, data gathering, and preliminary analyses. Researchers could use it to quickly survey large bodies of literature, identify gaps, and generate novel hypotheses, allowing faster iterations in the research process. For scientific research, where reproducibility and clear methodological descriptions are critical, the "chain-of-thought" process of the AI will need to be transparent enough for researchers to understand and trust its methods.

AI-Driven Hypotheses

AI is changing the way scientists ask questions and form hypotheses. AI models can rapidly analyze vast and complex datasets—ranging from genomic sequences and chemical libraries to psychological literature and historical records—to uncover patterns and relationships that might otherwise go unnoticed. For instance, scientists are already using Deep Research to brainstorm new hypotheses. The tool's ability to identify subtle or non-intuitive connections empowers researchers to formulate new hypotheses faster and with greater confidence. In 2024, a new study introduced a groundbreaking approach to hypothesis generation in psychology. The researchers combined causal knowledge graphs (maps of cause-and-effect relationships) with a large language model to uncover new research ideas. They analyzed thousands of psychology articles using an LLM to extract cause-and-effect pairs from the literature. These relationships were then used to construct a specialized causal graph, mapping how different psychological concepts are interconnected. Using algorithms to predict potential new connections within the graph, the team generated 130 fresh research hypotheses related to "well-being." When these AI-driven hypotheses were evaluated, they were found to be as novel and insightful as those created by doctoral psychology students—and significantly better than hypotheses generated by the LLM alone. This approach highlights how AI-driven systems can help researchers extract high-quality, innovative insights from the ever-growing pool of scientific literature.

AI-Driven Experimentation

AI is transforming scientific experimentation by enabling virtual testing of millions of potential experiments before lab work even begins. The SynBot demonstrates this capability through autonomous organic molecule synthesis: Its three-layer architecture combines AI-driven planning, instruction translation, and robotic execution. The system continuously optimizes reaction conditions and yields while adapting to different research objectives. In the UK, scientists operated two Automated Formulation Laboratories simultaneously in January 2025, investigating paint formulations through small angle scattering. AI algorithms

INDUSTRIES

analyzed results in real time to determine subsequent experiments, enabling rapid iteration between machines. The Human Pangenome Reference Consortium shows another AI-driven experiment application: using AI to create comprehensive genome mapping for virtual genetic experiments. This enables simulation of genetic variations and their impacts before conducting physical experiments, accelerating therapeutic development.

AI-driven experimentation allows researchers to rapidly test and refine ideas before committing to expensive lab work, drastically reducing both time and cost. Systems like the SynBot and the Automated Formulation Laboratories can learn in real time, adapting experimental conditions on the fly to optimize outcomes. This approach accelerates innovation by removing many of the traditional bottlenecks in trial-and-error experimentation. Similarly, AI-based projects like the Human Pangenome Reference provide virtual models for genetic studies, letting scientists explore and predict the effects of countless genetic variations upfront.

AI–Powered Analysis and Interpretation

AI is great at detecting subtle patterns that a human might overlook. Take AlphaMissense, an AI model developed by Google DeepMind. It evaluates the pathogenicity (disease-causing potential) of 71 million genetic mutations using transformer-based deep learning. The model predicts which genetic variants are likely harmful with exceptional accuracy, helping researchers focus on high-priority mutations instead of relying on traditional trial-and-error approaches. This significantly accelerates genetic research and enhances our understanding of diseases. AlphaMissense specializes in analyzing missense mutations—where one amino acid is replaced by another in a protein—and predicts their effects on protein function and disease risk. Its predictions, validated against known pathogenic variants, surpass the accuracy of previous computational methods.

Similarly, EVEscape showcases AI's predictive power by analyzing historical biophysical and structural data to anticipate viral evolution, supporting vaccine development and pandemic preparedness. The Human Pangenome Reference project further highlights AI's ability to process diverse genomic sequences, uncovering population-level variations that traditional methods often miss. These AI systems adapt their analyses in real time based on new data, enabling researchers to pursue promising leads without waiting for complete experimental cycles—fundamentally accelerating scientific discovery.

AI to Speed Up New Materials Development

In 2024, researchers at Argonne National Laboratory announced GHP-MOFassemble, an AI tool that accelerates the discovery of metal-organic frameworks (MOFs) for carbon capture. MOFs, constructed from metal nodes and organic linkers, show promise for capturing carbon dioxide, but their vast possible configurations make traditional testing impractical. GHP-MOFassemble creates novel organic linkers and combines them with copper or zinc-based metal nodes into MOFs with primitive cubic topology. The AI evaluates each design for uniqueness, synthesizability, and stability, then simulates their CO2 absorption capacity. The system identifies exceptional performers—those exceeding 96.9% of tested structures—while ensuring designs remain practical for laboratory synthesis. This breakthrough illustrates AI's power to revolutionize materials science. By automating the entire pipeline from design to evaluation, GHP-MOFassemble accelerates the discovery of high-performing materials while ensuring they can be synthesized in real laboratories—bridging the crucial gap between theoretical prediction and experimental reality.

Meta is hoping to speed up similar material discoveries elsewhere. The company's Open Materials 2024 (OMat24)

INDUSTRIES

initiative aims to democratize AI-driven materials discovery by releasing extensive datasets and models to the public. Material discovery typically requires costly computing power and access to proprietary datasets, creating barriers for many researchers. Meta is addressing this by making OMat24 freely available and open source on Hugging Face, allowing scientists worldwide to accelerate their materials research through AI applications.

Animal Decoding

Was whale song interpretation on your 2025 bingo card? Not to brag, but it was on ours. Scientists are using AI to analyze vast amounts of acoustic data collected from whales. Google AI, in collaboration with NOAA's Pacific Islands Fisheries Science Center, trained an AI model on underwater recordings to help scientists better understand whales' behavioral and migratory patterns. Researchers at MIT's Computer Science and Artificial Intelligence Lab used statistical models and AI algorithms for pattern recognition and classification to analyze 8,719 codas from sperm whales. And a team from the University of California, Davis employed machine learning algorithms to analyze and replicate humpback whale vocalizations, enabling a groundbreaking 20-minute exchange with a humpback whale named Twain.

It's not just whales. AI is helping us interpret pigs and dogs too. European researchers developed algorithms to interpret pig sounds for improved farm animal welfare, analyzing thousands of recordings to identify emotional states through different vocalizations. Meanwhile, researchers from University of Michigan and INAOE demonstrated that AI models pretrained on human speech can be adapted for animal communication. Using Wav2vec2, they analyzed dog vocalizations across breeds and contexts, achieving 70% accuracy in classification tasks—outperforming models specifically trained on dog barks.

Obviously, it would be great to be able to communicate with Fido but the implications are actually much bigger—and weirder. This technology could help us recognize alien life. By developing systems to recognize and interpret non-human communication patterns, we're building essential capabilities for potential extraterrestrial contact. The key challenge isn't just how to communicate after finding alien life—it's how to recognize intelligent communication in the first place. We might be missing alien messages simply because we don't recognize their form as language. Research in animal communication could help us identify and understand fundamentally different types of intelligence and communication, preparing us for peaceful contact with extraterrestrial civilizations.

INDUSTRIES

FINANCE

AI Assisted Asset Pricing and Management

Imagine having a financial analyst who can instantly process decades of market data and spot hidden patterns across thousands of assets. That's what AI is bringing to Wall Street. AI is transforming how we determine the value of financial assets through advanced models called transformer networks. These systems, originally designed for tasks like language translation, can now analyze complex relationships between different financial assets simultaneously. The breakthrough lies in AI's ability to understand how different financial instruments—from stocks to bonds to commodities—influence each other in subtle ways that traditional models often miss. A simplified version called the linear transformer makes this technology both powerful and practical, helping financial institutions make more informed investment decisions.

The power of AI in finance isn't just theoretical. In a comprehensive study spanning 95 years of US stock market data (1926–2021), researchers compared machine learning approaches against 17 standard financial models based on both traditional and behavioral finance theories. The results show that machine learning can detect complex patterns in market data that conventional methods miss, particularly when it comes to understanding how different risk factors interact. This improved accuracy is especially valuable because financial markets often behave in nonlinear ways, meaning changes in one factor might have disproportionate effects on stock returns.

Mitigating Fraud

The battle against financial fraud has entered a new era, with AI emerging as both protector and threat. On the defensive side, financial institutions are deploying sophisticated AI systems that act like vigilant security guards, scanning millions of transactions in real time to catch suspicious activity before money disappears. These AI watchdogs are proving remarkably effective: in 2024 alone, the US Treasury Department's AI systems helped recover $4 billion in fraudulent transactions, including $1 billion in check fraud.

But criminals are also wielding AI as a weapon. The FBI warns that fraudsters are now using AI to craft increasingly convincing scams, generating personalized phishing messages that can fool even cautious consumers. More alarming still is AI's role in creating synthetic identities—completely fabricated but convincing personas used to open fraudulent accounts. As financial institutions race to strengthen their defenses—with 70% expected to adopt AI security by 2025—they face an unprecedented challenge: protecting against AI-powered attacks that can launch thousands of sophisticated fraud attempts simultaneously. It's becoming clear that in this new AI powered financial environment, fighting AI with AI isn't just an option. It's a necessity.

Predicting Financial Risk

Imagine a business trying to predict financial storms, like sudden economic downturns or bad investments. Traditionally, companies have relied on tools like spreadsheets and past financial reports to spot these risks. But like using a paper map in the age of GPS, these old methods struggle with today's flood of information from news articles, social media, and market trends that aren't neatly organized in spreadsheets.

A 2024 study shows that traditional financial risk prediction methods are becoming inadequate for handling the massive amounts of unstructured data businesses face today. The solution? Using machine learning and natural language processing (NLP) to act as a "risk detective" for text, scanning news stories, CEO speeches, and regulatory filings to spot hidden red flags like negative sentiment or mentions of lawsuits. The study introduces DeepFM, a hybrid AI model that combines both numerical and textual data to predict

INDUSTRIES

risks. Unlike older models, it finds complex patterns in both structured data (like sales figures) and unstructured data (like customer complaints). Testing showed this combination works better than traditional methods: it's faster, more accurate, and adapts to new risks in real time. For businesses, this means fewer surprises and smarter decisions, like upgrading from a blurry telescope to high-definition radar.

Customized Portfolios

Remember when investing meant choosing between a handful of standard mutual funds? Those days are fading as AI changes how we build investment portfolios, making them as unique as our fingerprints. This transformation is particularly evident in the rise of socially conscious investing, where young investors, especially members of Gen Z, are demanding portfolios that align with their values. But the revolution extends far beyond environmental or social causes. Take JPMorgan's acquisition of OpenInvest, for example. Their AI platform doesn't just sort investments into broad categories—it analyzes your entire financial picture, including external assets, and crafts a portfolio that precisely matches your personal values and goals. Similarly, EquityPlus Investment's AI system dives deep into individual client behavior, risk tolerance, and investment history to create truly personalized strategies.

What makes these AI systems particularly powerful is their ability to adapt in real time. Unlike traditional portfolio management, which might rebalance quarterly, these systems continuously monitor market changes and shifts in client priorities. Portfolios can now automatically adjust when risk tolerance changes or when new investment opportunities align with a person's values. The result is investment portfolios that aren't just personalized—they're responsive. Through sophisticated cluster analysis and automated rebalancing, these AI systems ensure investments consistently reflect both financial goals and personal values.

Consumer-Facing Robo-Advisers

Consumer-facing robo-advisers are now widely available in the financial services sector, with many providers deploying advanced algorithms and increasingly sophisticated AI to automate investment advice, budgeting, and portfolio management. As hiring a human money manager can cost several thousand dollars, these platforms have become especially appealing to younger users looking for cost-effective guidance.

Consumers are turning to apps like Cleo AI and Bright, both of which prompt users to connect their bank accounts through Plaid so the chatbots can analyze spending, help manage debt, and build credit. However, these tools also leverage personal data to upsell additional services, pushing the boundaries between helpful personalization and potentially manipulative marketing tactics. At the same time, individuals are also harnessing non-specialized financial chatbots—such as Perplexity for portfolio research and large language models like ChatGPT or Claude for more in-depth analysis—demonstrating how improved reasoning capabilities in AI increasingly position these solutions as direct competitors to traditional human advisors.

INDUSTRIES

INSURANCE

Predicting Workplace Injuries

Workplace safety in the United States continues to improve, with fatal work injuries declining by 3.7% between 2022 and 2023, according to the US Bureau of Labor Statistics. To build on this progress, many companies are leveraging AI to enhance safety measures. For example, leading construction contractor JE Dunn collaborated with Newmetrix, an AI-driven risk prediction platform, to implement data-driven safety systems. This initiative enabled them to proactively prevent incidents and improve safety outcomes for their 3,500 employees. Similarly, Amazon has adopted computer vision and machine learning technologies in its warehouses to provide real-time safety alerts, contributing to fewer injuries and operational gains, including a 12% increase in net sales. Siemens also reported a 30% reduction in workplace injuries over two years by incorporating AI-based safety protocols.

Industry research suggests that AI-powered tools could prevent approximately 4,500 injuries and 50 deaths annually in scaffold-related incidents alone. The potential impact expands further with the integration of wearable devices to monitor worker fatigue and environmental sensors that help identify OSHA compliance issues, such as unsafe scaffolding conditions.

Improving Damage Assessment

At the time of this writing, more than 40,000 acres have burned in Los Angeles County, California, since the beginning of 2025. Maxar Intelligence has collected more than 34,000 square kilometers of high-resolution satellite imagery of the affected areas, which is shared through its Geospatial Platform (MGP) Pro and humanitarian Open Data Program. This imagery, made accessible to organizations like NOAA and the public, provides vital visual data for property assessments. Microsoft's AI for Good Lab has put this data to work, deploying AI models to assess the destruction caused by the Palisades and Eaton fires. Its analysis of the Palisades Fire revealed the scope of the damage: Among 18,000 evaluated buildings, 6,803 were damaged while 11,735 remained intact.

The insurance industry has also embraced this technology through McKenzie Intelligence Services' Global Events Observer platform. Using Maxar's imagery, their AI analysis identified more than 12,000 buildings as destroyed or severely damaged, with an additional 16,000 at risk of internal damage. This rapid assessment has accelerated the claims process, helping affected communities recover more quickly. These AI-powered tools represent a significant advance in disaster response, allowing emergency services, insurers, and relief organizations to act faster and more effectively in helping communities rebuild.

AI Powered Fire Prevention

In the wake of the January 2025 Los Angeles wildfires, fire departments and researchers are more urgently working together to develop AI-powered systems that leverage satellite imagery, drones, and smart sensors to provide early warnings and real-time insights. California's ALERTCalifornia program, run by UC San Diego, uses more than 1,100 cameras and sensors to monitor fire-prone areas 24/7, automatically identifying signs of wildfires in video footage. In December 2024, it detected a fire in Black Star Canyon at 2 a.m., allowing firefighters to contain it to less than a quarter-acre. Austin Energy has implemented an AI-driven camera network that monitors a 437-square-mile area in Texas, providing live images and alerts when smoke is detected. Meanwhile, startups like Los Angeles-based PriviNet are designing sensor networks capable of running on solar power for extended periods, further enhancing early detection capabilities. IBM and NASA are also contributing to wildfire prevention efforts, with geospatial AI models that analyze past fire events to help scientists better understand fire behavior and refine prevention strategies. While challenges

INDUSTRIES

remain, AI's ability to detect and predict wildfires offers a promising path to mitigating their devastating impact.

The Connected Worker

In 2024, insurance companies expanded their focus on the "Connect and Protect" approach, a strategy that uses advanced technologies like IoT, telematics, AI, and data analytics to shift from reactive to proactive risk management. Rather than simply compensating for losses after they occur, this approach connects insured assets—such as vehicles, homes, businesses, or even employees themselves—with real-time monitoring and predictive insights to reduce risks and enhance safety. A notable example is Swedish safety equipment company Guardio, which introduced the Armet PRO helmet in late 2024. This innovative helmet features the Multi-directional Impact Protection System to reduce the risk of brain injury during impacts, along with a Quin intelligent sensor for enhanced monitoring and safety.

However, this increased reliance on data collection and monitoring raises concerns about privacy and the potential for overly intrusive surveillance. Companies that cross the line into "Big Brother"–style oversight risk pushback from employees.

Liability Insurance for AI

What happens if machine learning systems are compromised by attackers who inject fake training data, leading to flawed outcomes? Or if a health care company's AI misinterprets patient data and fails to detect cancer? Who is responsible when machines behave badly? This is not just a philosophical dilemma but also a pressing legal issue that demands resolution as AI becomes more pervasive.

Right now, AI-related liability insurance is still a patchwork. Most coverage falls under existing policies—like professional liability or cyber insurance—rather than under standalone AI-specific products. Yet when AI fails, it can expose companies to an array of risks that may stretch beyond traditional coverage, from corrupted outputs and data breaches to significant reputational harm. Cyber insurance might protect against third-party hacking and ensuing lawsuits, but won't defend against every AI-driven hazard. In health care, malpractice insurance usually applies when AI tools malfunction, since providers remain responsible for validating the technology they use. Providers can, however, seek indemnification from third-party AI vendors if an algorithmic error leads to patient harm.

As AI systems grow in complexity, many worry existing policies won't keep up—prompting debates over new liability frameworks. Some experts propose a strict liability approach, making operators or developers liable regardless of fault. Others advocate adapting today's duty-of-care model to ensure AI providers diligently monitor and maintain their systems. Meanwhile, specialized endorsements and AI-focused insurance products are emerging to address the "black box" challenges inherent in AI and the security risks tied to generative AI. These evolving offerings aim to protect both software developers and end-users as AI continues to expand into more critical areas.

INDUSTRIES

HR

Autonomous Talent Acquisition

AI is transforming recruitment by handling time-consuming tasks like resume screening and candidate communications. For example, in 2024, LinkedIn announced its Hiring Assistant, a highly integrated agent that fits into the LinkedIn workflow. Companies like Siemens, Canva, and AMS are already using the tool to identify and hire candidates. LinkedIn believes they can automate nearly 80% of the pre-offer workflow.

Companies like Unilever expanded their use of AI tools like HireVue, which analyzes video interviews for language and facial cues, paired with anonymized resume screening to reduce unconscious bias. In 2024, it introduced multilingual AI analysis to better assess global candidates. Other AI agents in the space include Paradox, which is the current leader in recruitment automation. These systems can quickly analyze resumes against job requirements, identify top candidates, and maintain ongoing communication through chatbots. It's been used successfully by Chipotle, which announced in October 2024 that Paradox reduces time to hire by 75%. Notably, there is now also an outcrop of articles and blogs on "how to get your resume past AI screening tools," with tips and tricks on how to "beat the systems" with smart formatting and wording choices.

While there are some fully autonomous recruitment agents available on the marketplace today, most organizations are choosing to integrate AI tools with existing applicant tracking systems. However, organizations must watch for potential bias in AI hiring decisions, as these systems can reflect discriminatory patterns found in their training data. HR professionals should proceed with caution—in 2024, a California court found that HR vendors using AI can be liable for discrimination claims stemming from the customers' job applications (additional details can be found in the Policy and Regulation section of this trend report).

AI Onboarding and Integration

Poor onboarding has become a retention issue. Nearly half of new hires rate their post-onboarding training as inadequate, and 30% of those dissatisfied with onboarding plan to seek new jobs within three months. However, data shows that AI-supported onboarding reduces first-year turnover by 30% compared with traditional methods. There are lots of potential applications here: AI could transform onboarding from a generic, one-size-fits-all approach into a deeply personalized experience, analyzing each new hire's background to create tailored training content. A sales representative in Seattle might encounter examples featuring local clients and technology sales scenarios drawn from their previous roles. A financial analyst in Miami might see cases involving regional markets they've already worked with. Rather than generic company policies, employees would receive contextual examples relevant to their specific responsibilities and experience.

AI could also adapt how information is delivered based on learning preferences. Visual learners might receive interactive diagrams and illustrated workflows, while auditory learners get podcast-style content and guided walkthroughs. The system would continuously refine its approach based on engagement data, ensuring optimal information retention. Beyond content delivery, AI could identify potential knowledge gaps by analyzing a new hire's background and proactively suggest relevant resources. It could track engagement patterns to flag when additional support might be needed, enabling HR teams to provide targeted assistance before small challenges become major obstacles. This personalized approach could help employees feel genuinely understood and supported from day one.

Employee Engagement and Retention

Companies across industries are tapping into AI systems to personalize employee development. IBM's Watsonx Assistant

INDUSTRIES

streamlines routine HR inquiries while recommending career moves tailored to individual skills and interests. Oracle Grow maps out customized development tracks that adapt as employees gain new competencies, transforming career progression from an abstract concept into a series of actionable steps.

Behind the scenes, AI can scan for subtle warning signs of employee turnover. By tracking changes in communication, meeting engagement, and work output, these systems flag potential issues early enough for meaningful intervention. This shifts HR from reactive problem-solving to proactive engagement. AI could also transform how companies gather and use employee feedback. Rather than relying on annual surveys, AI systems continuously analyze input from multiple channels—chat logs, team meetings, performance reviews, and more. NLP can then extract themes, giving leaders clear insight into team dynamics.

Communicorp UK transformed their employee check-ins using Employment Hero, an AI-powered platform that streamlines one-on-one meetings previously hampered by managers' time constraints. The system ensures consistent, structured conversations while making both managers and employees more accountable for follow-through. By standardizing the check-in process, the platform facilitates more meaningful discussions and clearer goal-setting. Building on this success, Communicorp UK has expanded the platform's use to include performance reviews and goal tracking, creating a more comprehensive employee development framework.

Benefits Selection and Management

Oracle introduced an AI agent to help employees make open enrollment choices. Employees can ask back-and-forth questions using Oracle's benefits analyst AI agent, which is part of the Oracle Fusion Cloud Human Capital Management platform. For example, an employee who will be married this year can ask questions like, "What are the costs to add my spouse to different health plan tiers?" Or, "Do I need to wait until after the wedding to make changes?" Or, "Which plans offer the best coverage for both of us?" Oracle's AI pulls data from the company's benefits documentation and other company sources, such as existing benefit choices, benefits eligibility, or home location, as needed. The AI system integrates data from benefits documentation, employee elections, eligibility rules, and location requirements to answer employee questions. By analyzing this data, it can provide specific guidance for each employee's situation. Bswift's AI assistant Emma demonstrates similar capabilities, handling benefits support questions day and night. Data shows Emma resolves 87% of employee inquiries without human intervention, with most questions (77%) occurring after business hours. This automated support helps employees get answers when they need them and reduces the volume of routine questions HR teams must handle.

INDUSTRIES

MARKETING

AI Shifts Search

If you've Googled anything in the past few months, you've noticed that Google has been experimenting with AI-generated summaries at the top of search results. Instead of the traditional "10 blue links," users increasingly see AI-generated snapshots that pull information from multiple sources. For the small but mighty group of Bing users out there, Microsoft's Bing Chat (powered by GPT-4) offers conversational answers, often eliminating the need for users to click onto a website. As a result, users may get their core answers directly in the Search Engine Results Page or via a chat-like interface, clicking fewer links. This shifts how marketers should approach SEO and content strategy.

AI isn't just changing what is presented when something is searched, it's changing how people search. For example, query length has increased, with users now favoring longer, more conversational queries over simple keywords. The average query on AI-powered platforms like Perplexity is 10–11 words, compared to 2–3 keywords on traditional search engines. As of late 2024, Google still leads in the search market, accounting for 92.4% of referral traffic. However, AI-powered search tools are showing significant growth. ChatGPT reached 3.7 billion monthly visits in October 2024, marking 15.9% year-over-year growth. Perplexity saw 90.8 million visits in October 2024, with 199.2% year-over-year growth.

LLMs have also dramatically improved the ability of search engines to better understand the context and intent behind a query—not just keyword matching. As a result, marketers need to provide content that matches user intent more precisely, covering broader contextual and long-tail queries. Also, with natural language processing improving, more users are doing voice queries or "chat" style queries. As such, optimizing content for spoken language and question-based queries ("who," "what," "where," "how") becomes increasingly important.

Dynamic Engagement Through Deep Personalization

Traditional marketing has relied on one-way communications like emails, PDFs, and social posts, but AI is transforming this landscape into an era of dynamic, two-way conversations. The real opportunity isn't in replacing human support with chatbots—it's in creating intelligent conversational interfaces that enhance the customer experience through personalization. When users receive tailored recommendations and contextual support, they get more value from each interaction.

Companies are already implementing AI-powered conversational interfaces that deliver personalized, context-aware recommendations and services. As these experiences become more common, users increasingly expect to interact with brands through voice or text-based AI assistants that understand their needs and preferences. When implemented thoughtfully, this deep personalization creates more meaningful connections, fostering brand loyalty and driving better business outcomes.

Another way to personalize engagement is through the presentation of information. Imagine websites and mobile apps that adapt their layouts and recommendations in real time based on individual user behavior and preferences (see the Generative User Interfaces trend for more details). This approach can make users feel truly understood, leading to longer engagement and higher conversion rates.

AI-Assisted Campaigns

In the 2024 trend report, we started to see the emergence of AI tools for copywriting and design. This year, we started to see more platforms provide full-stack solutions for streamlining of campaigns within a single interface. These platforms now manage everything from ideation to execution to analytics, with advanced reinforcement learning models automatically optimizing campaign parameters based on real-time performance data.

INDUSTRIES

Dynamic creative optimization has taken a leap forward, with AI systems generating and testing multiple ad variations in microseconds, from images to headlines. Meanwhile, major marketing platforms like HubSpot, Salesforce, and Adobe have integrated LLM capabilities, allowing marketers to adjust email copy, create landing pages, or segment audiences through intuitive chat interfaces. This integration enables marketers to build entire campaign workflows through simple prompts. For example, a marketer could request “Draft a multichannel campaign targeting mid-career tech professionals, focusing on brand awareness,” and tools like Adobe Experience Platform’s AI Assistant will handle everything from audience segmentation to asset creation using Adobe Firefly, while generating multiple content variations for personalized communications.

Anecdotal Observations, Now Usable Marketing Data

AI has transformed our ability to understand subtle human reactions, turning what were once subjective observations into measurable marketing insights. Companies can now analyze video data to detect micro-expressions and facial cues, providing precise measurements of consumer engagement. AI systems evaluate customer responses to store layouts and branded elements in real time, creating a new frontier in consumer behavior analysis.

In the retail space, companies are deploying increasingly sophisticated emotion recognition technologies. MoodMe’s facial analysis system captures real-time emotional responses, while MorphCast’s interactive video platform helps businesses understand viewer engagement through facial expression analysis. Viso Suite takes a broader approach, using deep learning to track customer behavior patterns, from shopping cart usage to wait times and crowd density. Meanwhile, Chinese tech giants like Megvii (the creator of Face++) have emerged as global leaders in facial recognition technology, pushing the boundaries of what’s possible in consumer behavior tracking.

This transformation from anecdotal insights to quantifiable data has opened unprecedented personalization opportunities. However, the intimate nature of tracking human expressions and reactions has prompted serious privacy discussions. The EU’s AI Act reflects these concerns, having banned real-time biometric identification systems in public spaces, with limited exceptions. This regulatory response highlights the delicate balance between technological innovation and privacy protection in the emerging AI landscape.

AUTHORS & CONTRIBUTORS

Amy Webb

Chief Executive Officer

As founder and CEO of the Future Today Strategy Group (FTSG), Amy pioneered a unique quantitative modeling approach and data-driven foresight methodology that identifies signals of change and emerging patterns very early. Using that information, Amy and her colleagues identify white spaces, opportunities, and threats early enough for action. They develop predictive scenarios, along with executable strategy, for businesses worldwide. In addition, Amy is regularly asked to advise policymakers in the White House, Congress, U.S. regulatory agencies, the European Union and United Nations. In 2023, Amy was recognized as the #4 most influential management thinker in the world by Thinkers50, a biannual ranking of global business thinkers. With research specializations in both AI and biotechnology, Amy is the author of four books which have been translated into 23 languages. She developed and teaches the Strategic Foresight Course at NYU Stern School of Business.

Sam Jordan

Technology & Computing Lead

Sam Jordan is a Senior Manager and the Technology and Computing Lead at FTSG. Her research focuses on the future of computing, spanning large-scale systems, personal devices, AI, and telecommunications. She also covers the space industry, analyzing advancements in satellite technology, communications infrastructure, and emerging aerospace innovations. She has worked with some of the world's largest technology companies to advance human-computer interaction, develop AI strategies, and drive innovation in device evolution.

Before joining FTSG, Sam was the CEO and co-founder of TrovBase, a secure platform for data discovery and analysis sharing. She also worked at IBM, where she helped large enterprises modernize their IT infrastructure, specializing in mainframes and integrating modern software and methodologies into legacy systems.

Sam currently serves as a coach in the Strategic Foresight MBA Course at NYU Stern School of Business and is an Emergent Ventures Fellow at the Mercatus Center. She holds a B.S. in Economics and Data Analysis from George Mason University and an MBA from NYU's Stern School of Business.

Chief Executive Officer
Amy Webb

Managing Director
Melanie Subin

Director of Marketing & Comms.
Victoria Chaitoff

Creative Director
Emily Caufield

Editor
Erica Peterson

Copy Editor
Sarah Johnson

SELECTED SOURCES

Aggarwal, Pranjal, et al. “GEO: Generative Engine Optimization.” arXiv, 28 June 2024, https://arxiv.org/abs/2311.09735.

Agiza, Ahmed, et al. “PoliTune: Analyzing the Impact of Data Selection and Fine-Tuning on Economic and Political Biases in Large Language Models.” Proceedings of the Seventh AAAI/ACM Conference on AI, Ethics, and Society (AIES 2024), pp. 1-10. ojs.aaai.org/index.php/AIES/article/view/31612/33779.

“AI Fire Prediction.” IBM Think Blog, 2025, https://www.ibm.com/think/news/ai-fire-prediction.

“AI for Energy: Opportunities for a Modern Grid and Clean Energy Economy.” U.S. Department of Energy, 30 Apr. 2024, www.energy.gov/sites/default/files/2024-04/AI%20EO%20Report%20Section%205.2g%28i%29_043024.pdf.

AI Index. “Artificial Intelligence Index Report 2024.” Stanford University, 2024, https://aiindex.stanford.edu/report/.

Akiba, Takuya, et al. “Evolutionary Optimization of Model Merging Recipes.” Nature Machine Intelligence, vol. 7, 2025, https://arxiv.org/abs/2403.13187.

Angelopoulos, Angelos, et al. “Transforming Science Labs into Automated Factories of Discovery.” Science Robotics, vol. 9, no. 95, 2024, https://doi.org/10.1126/scirobotics.adm6991.

Answer.AI. “FSDP and QLoRA: Efficient Fine-Tuning Strategies.” Answer.ai, 6 Mar. 2024, https://www.answer.ai/posts/2024-03-06-fsdp-qlora.html.

“Are Bigger Language Models Always Better?” IBM, 15 July 2024, https://www.ibm.com/think/insights/are-bigger-language-models-better.

Arora, Daman, et al. “MASAI: Modular Architecture for Software-Engineering AI Agents.” arXiv, 17 June 2024, https://arxiv.org/abs/2406.11638.

Aschenbrenner, Leopold. Situational Awareness Report 2024, June 2024, https://situational-awareness.ai/.

“ASML Risks Losing Chinese Market Permanently if It Complies with US Restrictions.” Global Times, 1 Sept. 2024, www.globaltimes.cn/page/202409/1319035.shtml.

Ball, Dean W. “Deepfakes and the Art of the Possible.” Hyperdimensional, 30 May 2024, https://www.hyperdimensional.co/p/deepfakes-and-the-art-of-the-possible.

Barth, Antje. “Amazon Titan Image Generator and Watermark Detection API Are Now Available in Amazon Bedrock.” AWS Blog, 23 Apr. 2024, https://aws.amazon.com/blogs/aws/amazon-titan-image-generator-and-watermark-detection-api-are-now-available-in-amazon-bedrock/.

Bassner, Patrick, et al. “Iris: An AI-Driven Virtual Tutor for Computer Science Education.” ITiCSE 2024: Proceedings of the 2024 on Innovation and Technology in Computer Science Education, vol. 1, July 2024, pp. 394-400, https://doi.org/10.1145/3649217.3653543.

Beatty, Sally. “The Phi-3 Small Language Models with Big Potential.” Microsoft Source, 23 Apr. 2024, https://news.microsoft.com/source/features/ai/the-phi-3-small-language-models-with-big-potential/.

Bonney, Kathryn, et al. “The Impact of AI on the Workforce: Tasks versus Jobs?” Economics Letters, vol. 244, Nov. 2024, article no. 111971, https://www.sciencedirect.com/science/article/abs/pii/S0165176524004555.

Bouzida, Anya, et al. “CARMEN: A Cognitively Assistive Robot for Personalized Neurorehabilitation at Home.” Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI ‘24), 2024, https://doi.org/10.1145/3610977.3634971.

Brazil, Rachel. “How AI Is Transforming Drug Discovery.” The Pharmaceutical Journal, 3 July 2024, https://pharmaceutical-journal.com/article/feature/how-ai-is-transforming-drug-discovery.

Cai, Kenrick. “Google AI Systems Make Headway with Math in Progress Toward Reasoning.” Reuters, 25 July 2024, www.reuters.com/technology/artificial-intelligence/google-ai-systems-make-headway-with-math-progress-toward-reasoning-2024-07-25/.

Callaway, Ewen. “AI Protein-Prediction Tool AlphaFold3 Is Now More Open.” Nature, vol. 620, no. 7970, 2024, pp. 15-16, https://doi.org/10.1038/d41586-024-03708-4.

Caplin, Andrew, et al. “The ABC’s of Who Benefits from Working with AI: Ability, Beliefs, and Calibration.” National Bureau of Economic Research, No. 33021, Oct. 2024, https://www.nber.org/papers/w33021.

Castelvecchi, Davide. “Will AI’s Huge Energy Demands Spur a Nuclear Renaissance?” Nature, 25 Oct. 2024, doi:10.1038/d41586-024-03490-3.

Chandler, Simon. “This Website Is Using AI to Combat Political Bias.” Forbes, 17 Mar. 2020, www.forbes.com/sites/simonchandler/2020/03/17/this-website-is-using-ai-to-combat-political-bias/.

Chu, Jennifer. “Engineering Household Robots to Have a Little Common Sense.” Massachusetts Institute of Technology, 25 Mar. 2024, https://news.mit.edu/2024/engineering-household-robots-have-little-common-sense-0325.

Clark, Joseph. “AI Security Center Keeps DOD at Cusp of Rapidly Emerging Technology.” U.S. Department of Defense, 6 Sept. 2024, www.defense.gov/News/News-Stories/Article/Article/3896891/ai-security-center-keeps-dod-at-cusp-of-rapidly-emerging-technology/.

"Claude 3.5 Sonnet Multi-Modal Learning." claude3.pro, 13 Aug. 2024, https://claude3.pro/claude-3-5-sonnet-multi-modal-learning/.

Clegg, Nick. "Open Source AI Can Help America Lead in AI and Strengthen Global Security." Meta, 4 Nov. 2024, about.fb.com/news/2024/11/open-source-ai-america-global-security/.

Cohen, Ariel. "China's Massive Barrage in the Chip Battle." Forbes, 31 May 2024, www.forbes.com/sites/arielcohen/2024/05/31/chinas-massive-barrage-in-the-chip-battle/.

Cordova, Sgt. David. "Green Berets Leverage Immersive Simulator for Training." U.S. Army, 9 Feb. 2024, www.army.mil/article/273628/green_berets_leverage_immersive_simulator_for_training.

Cowen, Tyler. "AI's Effect on the US Economy Will Be Wildly Uneven." Bloomberg, 25 Oct. 2024, https://www.bloomberg.com/opinion/articles/2024-10-25/ai-s-effect-on-the-us-economy-will-be-wildly-uneven.

D'Alessandro, Marco, et al. "A Modular End-to-End Multimodal Learning Method for Structured and Unstructured Data." arXiv, 7 Mar. 2024, https://arxiv.org/abs/2403.04866.

Danelski, David. "Method Identified to Double Computer Processing Speeds." University of California, Riverside, 21 Feb. 2024, https://news.ucr.edu/articles/2024/02/21/method-identified-double-computer-processing-speeds.

Dave, Paresh. "Google Splits Up a Key AI Ethics Watchdog." WIRED, 31 Jan. 2024, https://www.wired.com/story/google-splits-up-responsible-innovation-ai-team/.

Dela Cruz, Jace. "Media Bias Detector: New AI Tool Provides Insights into How News Outlets Report on Various Topics." Tech Times, 26 June 2024, www.techtimes.com/articles/306071/20240626/media-bias-detector-new-ai-tool-provides-insights-news-outlets.htm.

Dettmers, Tim, et al. "QLoRA: Efficient Finetuning of Quantized LLMs." arXiv, 23 May 2023, https://arxiv.org/abs/2305.14314.

Dharmaraj, Samaya. "Ahmedabad, India AI: Transforming Urban Surveillance and Security." OpenGov Asia, 13 Jan. 2024, opengovasia.com/2024/01/13/ahmedabad-india-ai-transforming-urban-surveillance-and-security/.

Egan, Lauren, and Phelim Kine. "Biden's Final Meeting with Xi Jinping Reaps Agreement on AI and Nukes." Politico, 16 Nov. 2024, www.politico.com/news/2024/11/16/biden-xi-jinping-ai-00190025.

Eliot, Lance. "Mixture-of-Experts AI Reasoning Models Suddenly Taking Center Stage Due to China's DeepSeek Shock-and-Awe." Forbes, 1 Feb. 2025, https://www.forbes.com/sites/lanceeliot/2025/02/01/mixture-of-experts-ai-reasoning-models-suddenly-taking-center-stage-due-to-chinas-deepseek-shock-and-awe/.

"Employer-Reported Workplace Injuries and Illnesses – 2023." U.S. Bureau of Labor Statistics, 2024, https://www.bls.gov/news.release/osh.nr0.htm.

"FBI Issues Warning on AI Used for Financial Fraud." ABA Banking Journal, Dec. 2024, https://bankingjournal.aba.com/2024/12/fbi-issues-warning-on-ai-used-for-financial-fraud/

Fearn, Nicholas. "Less Admin, More Time with People: How an HR Professional's Job Has Been Transformed by AI." The Guardian, 20 Dec. 2024, https://www.theguardian.com/work-redefined/2024/dec/20/less-admin-more-time-with-people-how-an-hr-professionals-job-has-been-transformed-by-ai.

Feng, Emily. "Chinese Companies Offer to 'Resurrect' Deceased Loved Ones with AI Avatars." NPR, 21 July 2024, www.npr.org/2024/07/18/nx-s1-5040583/china-ai-artificial-intelligence-dead-avatars.

Feng, Shangbin, et al. "From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models." Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023), pp. 11740–11756. https://arxiv.org/abs/2305.08283.

"Figma Reintroduces Figma AI as First Draft." DesignWhine, 27 Sept. 2024, https://www.designwhine.com/figma-first-draft/.

Fist, Tim, and Arnab Datta. "How to Build the Future of AI in the United States." Institute for Progress, 23 Oct. 2024, ifp.org/future-of-ai-compute/.

Fu, Jia, et al. "Annotation-Free Artificial Intelligence for Abdominal Computed Tomography Anomaly Detection." eBioMedicine, vol. 111, 10 Dec. 2024, article no. 105497, https://doi.org/10.1016/j.ebiom.2024.105497.

Gress, Morgan. "Anthropic and Palantir Partner to Bring Claude AI Models to AWS for U.S. Government Intelligence and Defense Operations." Palantir Investor Relations, 7 Nov. 2024, https://investors.palantir.com/news-details/2024/Anthropic-and-Palantir-Partner-to-Bring-Claude-AI-Models-to-AWS-for-U.S.-Government-Intelligence-and-Defense-Operations/.

Gronholt-Pedersen, Jacob. "AI Decodes Oinks and Grunts to Keep Pigs Happy." Reuters, 24 Oct. 2024, https://www.reuters.com/technology/artificial-intelligence/ai-decodes-oinks-grunts-keep-pigs-happy-2024-10-24/.

Gurman, Mark. "Apple's AI Plans: GitHub Copilot Rival and App Testing Tool." Bloomberg, 15 Feb. 2024, https://www.bloomberg.com/news/articles/2024-02-15/apple-s-ai-plans-github-copilot-rival-for-developers-tool-for-testing-apps.

Ha, Taesin, et al. "AI-Driven Robotic Chemist for Autonomous Synthesis of Organic Molecules." Science Advances, vol. 9, no. 44, 1 Nov. 2023, https://doi.org/10.1126/sciadv.adj0461.

Healy, Jerome V., et al. “Automated Machine Learning and Asset Pricing.” Risks, vol. 12, no. 9, 2024, article 148, https://doi.org/10.3390/risks12090148.

Heater, Brian. “Iyo Thinks Its GenAI Earbuds Can Succeed Where Humane and Rabbit Stumbled.” TechCrunch, 27 May 2024, https://techcrunch.com/2024/05/27/iyo-thinks-its-gen-ai-earbuds-can-succeed-where-humane-and-rabbit-stumbled/.

Heaven, Will Douglas. “Generative AI Can Turn Your Most Precious Memories into Photos That Never Existed.” MIT Technology Review, 10 Apr. 2024, www.technologyreview.com/2024/04/10/1091053/generative-ai-turn-your-most-precious-memories-into-photos/.

“How DeepSeek Ripped Up the AI Playbook—and Why Everyone’s Going to Follow It.” MIT Technology Review, 31 Jan. 2025, https://www.technologyreview.com/2025/01/31/1110740/how-deepseek-ripped-up-the-ai-playbook-and-why-everyones-going-to-follow-it/.

“OpenAI Launches Operator—an Agent That Can Use a Computer for You.” MIT Technology Review, 23 Jan. 2025, www.technologyreview.com/2025/01/23/1110484/openai-launches-operator-an-agent-that-can-use-a-computer-for-you/.

Heikkilä, Melissa. “AI Language Models Are Rife with Different Political Biases.” MIT Technology Review, 7 Aug. 2023, www.technologyreview.com/2023/08/07/1077324/ai-language-models-are-rife-with-political-biases/.

“The Race to Find New Materials with AI Needs More Data. Meta Is Giving Massive Amounts Away for Free.” MIT Technology Review, 18 Oct. 2024, https://www.technologyreview.com/2024/10/18/1105880/the-race-to-find-new-materials-with-ai-needs-more-data-meta-is-giving-massive-amounts-away-for-free/.

“This New Data Poisoning Tool Lets Artists Fight Back Against Generative AI.” MIT Technology Review, 23 Oct. 2023, www.technologyreview.com/2023/10/23/1082189/data-poisoning-artists-fight-generative-ai/.

Herrman, John. “The Rise of the Self-Clicking Computer.” New York Magazine, 24 Oct. 2024, https://nymag.com/intelligencer/article/ai-anthropic-agent-claude-google.html.

Ho, Anson, et al. “Algorithmic Progress in Language Models.” arXiv, 9 Mar. 2024, https://arxiv.org/abs/2403.05812.

Hoffmann, Manuel, et al. “Generative AI and the Nature of Work.” Harvard Business School, 27 Oct. 2024, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5007084.

Hofmann, Valentin, et al. “AI Generates Covertly Racist Decisions about People Based on Their Dialect.” Nature, vol. 633, 2024, pp. 147–154, doi:10.1038/s41586-024-07856-5.

“Influence and Cyber Operations: An Update.” OpenAI, Oct. 2024, cdn.openai.com/threat-intelligence-reports/influence-and-cyber-operations-an-update_October-2024.pdf.

Jin, Jianna, et al. “Avoiding Embarrassment Online: Response to and Inferences about Chatbots When Purchases Activate Self‐Presentation Concerns.” Journal of Consumer Psychology, 2024, https://doi.org/10.1002/jcpy.1414.

Kelly, Bryan T., et al. “Artificial Intelligence Asset Pricing Models.” National Bureau of Economic Research, Jan. 2025, https://www.nber.org/papers/w33351.

Knight, Will. “OpenAI’s Long-Term AI Risk Team Has Disbanded.” WIRED, 17 May 2024, https://www.wired.com/story/openai-superalignment-team-disbanded/.

Kurian, Nomisha. “‘No, Alexa, No!’: Designing Child-Safe AI and Protecting Children from the Risks of the ‘Empathy Gap’ in Large Language Models.” Learning, Media and Technology, vol. 49, no. 1, 2024, pp. 1-15, https://doi.org/10.1080/17439884.2024.2367052.

“Large Language Models Are Getting Bigger and Better.” The Economist, 17 Apr. 2024, https://www.economist.com/science-and-technology/2024/04/17/large-language-models-are-getting-bigger-and-better.

Laursen, Lucas. “This AI-Powered Invention Machine Automates Eureka Moments.” IEEE Spectrum, 8 Oct. 2024, https://spectrum.ieee.org/ai-inventions.

Lefohn, Aaron. “Latest NVIDIA Graphics Research Advances Generative AI’s Next Frontier.” NVIDIA Blog, 2 May 2023, https://blogs.nvidia.com/blog/graphics-research-advances-generative-ai-next-frontier/.

Lepagnol, Pierre, et al. “Small Language Models Are Good Too: An Empirical Study of Zero-Shot Classification.” arXiv, 17 Apr. 2024, https://arxiv.org/abs/2404.11122.

Li, Tianyu, and Xiangyu Dai. “Financial Risk Prediction and Management Using Machine Learning and Natural Language Processing.” International Journal of Advanced Computer Science and Applications, vol. 15, no. 6, 2024, https://doi.org/10.14569/IJACSA.2024.0150623.

Liou, Joanne. “What Are Small Modular Reactors (SMRs)?” International Atomic Energy Agency, 13 Sept. 2023, www.iaea.org/newscenter/news/what-are-small-modular-reactors-smrs.

Lundberg, Steve. “Researchers develop new training technique that aims to make AI systems less socially biased.” Oregon State University, 25 June 2024. https://news.oregonstate.edu/news/researchers-develop-new-training-technique-aims-make-ai-systems-less-socially-biased

Lv, Ang, et al. “Autonomy-of-Experts Models.” arXiv, 22 Jan. 2025, https://arxiv.org/abs/2501.13074.

Ma, Enhao, et al. “A Predictive Language Model for SARS-CoV-2 Evolution.” Signal Transduction and Targeted Therapy, vol. 9, article no. 353, 23 Dec. 2024, https://www.nature.com/articles/s41392-024-02066-x.

Macmillan-Scott, Olivia, and Mirco Musolesi. “(Ir)rationality and Cognitive Biases in Large Language Models.” Royal Society Open Science, vol. 11, no. 6, 2024, https://royalsocietypublishing.org/doi/10.1098/rsos.240255.

Marr, Bernard. “How AI Is Used in War Today.” Forbes, 17 Sept. 2024, www.forbes.com/sites/bernard-marr/2024/09/17/how-ai-is-used-in-war-today/.

Martin, Iain. “Why AMD Spent $665 Million Buying a Tiny AI Research Team.” Forbes, 18 July 2024, https://www.forbes.com/sites/iainmartin/2024/07/18/why-amd-spent-665-million-buying-a-tiny-ai-research-team/.

Martin, Raiza. “Introducing NotebookLM.” Google Blog, 12 July 2023, https://blog.google/technology/ai/notebooklm-google-ai/.

Masri, Lena. “Facial Recognition is Helping Putin Curb Dissent With the Aid of U.S. Tech.” Reuters, 28 Mar. 2023, www.reuters.com/investigates/special-report/ukraine-crisis-russia-detentions/.

Matz, S. C., et al. “The Potential of Generative AI for Personalized Persuasion at Scale.” Scientific Reports, vol. 14, no. 4692, 26 Feb. 2024, doi:10.1038/s41598-024-53755-0.

Mei, Kai, et al. “AIOS: LLM Agent Operating System.” arXiv, 7 Nov. 2024, https://arxiv.org/abs/2403.16971.

Metz, Cade. “Nvidia’s Big Tech Rivals Put Their Own AI Chips on the Table.” The New York Times, 29 Jan. 2024, https://www.nytimes.com/2024/01/29/technology/ai-chips-nvidia-amazon-google-microsoft-meta.html.

Milne, Stefan, and Kiyomi Taguchi. “AI Headphones Let Wearer Listen to a Single Person in a Crowd, by Looking at Them Just Once.” University of Washington, 23 May 2024, https://www.washington.edu/news/2024/05/23/ai-headphones-noise-cancelling-target-speech-hearing/.

Mims, Christopher. “Michael Dell Spent 40 Years Preparing for an AI Boom No One Expected.” The Wall Street Journal, 14 Dec. 2024, https://www.wsj.com/tech/ai/michael-dell-spent-40-years-preparing-for-an-ai-boom-no-one-expected-8cc20c04.

Mittal, Govind, et al. “GOTCHA: Real-Time Video Deepfake Detection via Challenge-Response.” arXiv, 23 May 2024, https://arxiv.org/abs/2210.06186.

“PITCH: AI-Assisted Tagging of Deepfake Audio Calls Using Challenge-Response.” arXiv, 1 Oct. 2024, https://arxiv.org/abs/2402.18085.

Mollick, Ethan R., and Lilach Mollick. “Instructors as Innovators: A Future-Focused Approach to New AI Learning Opportunities, with Prompts.” The Wharton School Research Paper, 23 Apr. 2024, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4802463.

Motoki, Fabio, et al. “More Human than Human: Measuring ChatGPT Political Bias.” Public Choice, vol. 198, 2024, pp. 3–23. SpringerLink, doi:10.1007/s11127-023-01097-2.

Navarro Martínez, Olga, et al. “Possible Health Benefits and Risks of DeepFake Videos: A Qualitative Study in Nursing Students.” Nursing Reports, vol. 14, no. 4, 3 Oct. 2024, pp. 2746–2757, https://doi.org/10.3390/nursrep14040203.

Nellis, Stephen. “California’s Only Nuclear Plant to Use AI to Help Comply with New Licensing Challenges.” Reuters, 13 Nov. 2024, www.reuters.com/technology/artificial-intelligence/californias-only-nuclear-plant-use-ai-help-comply-with-new-licensing-challenges-2024-11-13/.

“New Laser-Array Processor Could Vastly Improve AI Computing Efficiency.” USC Viterbi School of Engineering, 8 Aug. 2023, viterbischool.usc.edu/news/2023/08/new-laser-array-processor-could-vastly-improve-ai-computing-efficiency/.

Newman, Jessica. “A Taxonomy of Trustworthiness for Artificial Intelligence.” Berkeley Center for Long-Term Cybersecurity, Jan. 2023, https://cltc.berkeley.edu/publication/a-taxonomy-of-trustworthiness-for-artificial-intelligence/.

Nezami, Zeinab, et al. “Generative AI on the Edge: Architecture and Performance Evaluation.” arXiv, 18 Nov. 2024, https://arxiv.org/abs/2411.17712.

“NSF and Quad Partners Launch AI-ENGAGE to Encourage Collaboration on Emerging Technologies and Agriculture.” National Science Foundation, 27 Sept. 2024, www.nsf.gov/news/nsf-quad-partners-launch-ai-engage-encourage-collaboration. Press release.

Olavsrud, Thor. “Microsoft and Industry Partners Showcase Specialized Small Language Models.” CIO, 20 Nov. 2024, https://www.cio.com/article/3608783/microsoft-and-industry-partners-showcase-specialized-small-language-models.html.

Olick, Diana. “Amazon Goes Nuclear, to Invest More than $500 Million to Develop Small Modular Reactors.” CNBC, 16 Oct. 2024, www.cnbc.com/2024/10/16/amazon-goes-nuclear-investing-more-than-500-million-to-develop-small-module-reactors.html.

Park, Hyun, et al. “A Generative Artificial Intelligence Framework Based on a Molecular Diffusion Model for the Design of Metal-Organic Frameworks for Carbon Capture.” Communications Chemistry, vol. 7, article no. 21, 14 Feb. 2024, https://www.nature.com/articles/s42004-023-01090-2.

Peters, Jay. “Reddit’s Upcoming API Changes Will Make AI Companies Pony Up.” The Verge, 18 Apr. 2023, https://www.theverge.com/2023/4/18/23688463/reddit-developer-api-terms-change-monetization-ai.

Pham, Thang M., et al. “SlimLM: An Efficient Small Language Model for On-Device Document Assistance.” arXiv, 25 Nov. 2024, arxiv.org/abs/2411.09944.

Picciotto, Rebecca. “Facebook Parent Meta Breaks Up Its Responsible AI Team.” CNBC, 18 Nov. 2023, https://www.cnbc.com/2023/11/18/facebook-parent-meta-breaks-up-its-responsible-ai-team.html.

Pilz, Konstantin, et al. “Increased Compute Efficiency and the Diffusion of AI Capabilities.” arXiv, 13 Feb. 2024, https://arxiv.org/abs/2311.15377.

Piper, Kelsey. “Inside OpenAI’s Multi-Billion-Dollar Gambit to Become a For-Profit Company.” Vox, 28 Oct. 2024, https://www.vox.com/future-perfect/380117/openai-microsoft-sam-altman-nonprofit-for-profit-foundation-artificial-intelligence.

Porter, Robert, et al. “LLMD: A Large Language Model for Interpreting Longitudinal Medical Records.” arXiv, 11 Oct. 2024, https://arxiv.org/abs/2410.12860.

Proofpoint Threat Research Team. “Security Brief: ‘Artificial Sweetener’ – SugarGh0st RAT Used to Target Americans.” Proofpoint, 16 May 2024, www.proofpoint.com/us/blog/threat-insight/security-brief-artificial-sweetener-sugargh0st-rat-used-target-american.

Radhakrishnan, Adityanarayanan, et al. “Mechanism for Feature Learning in Neural Networks and Its Implications for High-Dimensional Learning.” Science, vol. 383, no. 6690, 7 Mar. 2024, pp. 1461-1467. https://doi.org/10.1126/science.adi5639.

Renshaw, Jarrett, and Trevor Hunnicutt. “Biden, Xi Agree That Humans, Not AI, Should Control Nuclear Arms.” Reuters, 16 Nov. 2024, www.reuters.com/world/biden-xi-agreed-that-humans-not-ai-should-control-nuclear-weapons-white-house-2024-11-16/.

Riveland, Reidar, and Alexandre Pouget. “Natural Language Instructions Induce Compositional Generalization in Networks of Neurons.” Nature Neuroscience, 2024, https://www.nature.com/articles/s41593-024-01607-5.

Rogers, Reece. “AI Financial Advisers Target Young People Living Paycheck to Paycheck.” WIRED, 13 Jan. 2025, https://www.wired.com/story/ai-financial-advisers-apps-chatbots/.

Roth, Emma. “Google Cut a Deal with Reddit for AI Training Data.” The Verge, 22 Feb. 2024, https://www.theverge.com/2024/2/22/24080165/google-reddit-ai-training-data.

Saltini, Alice. “Navigating Cyber Vulnerabilities in AI-Enabled Military Systems.” European Leadership Network, 19 Mar. 2024, europeanleadershipnetwork.org/commentary/navigating-cyber-vulnerabilities-in-ai-enabled-military-systems/.

Santora, Marc, and Raymond Zhong. “Made in China, Exported to the World: The Surveillance State.” The New York Times, 24 Apr. 2019, www.nytimes.com/2019/04/24/technology/ecuador-surveillance-cameras-police-government.html.

Santoro, Cameron. “Accelerating Drug Discovery With AI for More Effective Treatments.” The American Journal of Managed Care, 17 Oct. 2024, https://www.ajmc.com/view/accelerating-drug-discovery-with-ai-for-more-effective-treatments.

Sato, Mia. “Major Record Labels Sue AI Company Behind ‘BBL Drizzy’.” The Verge, 24 June 2024, https://www.theverge.com/2024/6/24/24184710/riaa-ai-lawsuit-suno-udio-copyright-umg-sony-warner.

Schmirler, Robert, et al. “Fine-Tuning Protein Language Models Boosts Predictions Across Diverse Tasks.” Nature Communications, vol. 15, no. 7407, 28 Aug. 2024, https://www.nature.com/articles/s41467-024-51844-2.

Sharma, Puja. “Fraud Evolution, AI Takeover, and Borderless Banking: What to Expect in 2025.” IBS Intelligence, 19 Dec. 2024, https://ibsintelligence.com/ibsi-news/fraud-evolution-ai-takeover-and-borderless-banking-what-to-expect-in-2025/.

Siliezar, Juan. “Study Shows AI Can Be Fine-Tuned for Political Bias.” Tech Xplore, 22 Oct. 2024, techxplore.com/news/2024-10-ai-fine-tuned-political-bias.html.

Smith, Brad, and Melanie Nakagawa. “Our 2024 Environmental Sustainability Report.” Microsoft On the Issues, 15 May 2024, https://blogs.microsoft.com/on-the-issues/2024/05/15/microsoft-environmental-sustainability-report-2024/.

“Sora Is Here.” OpenAI, 9 Dec. 2024, https://openai.com/index/sora-is-here/. Press release.

Suganya, B., et al. “Dynamic Task Offloading Edge-Aware Optimization Framework for Enhanced UAV Operations on Edge Computing Platform.” Scientific Reports, vol. 14, article no. 16383, 16 July 2024, https://www.nature.com/articles/s41598-024-67285-2.

Surjadi, Milla. Colleges Race to Ready Students for the AI Workplace.” The Wall Street Journal, 5 Aug. 2024, https://www.wsj.com/us-news/education/colleges-race-to-ready-students-for-the-ai-workplace-cc936e5b.

Thakur, Dhanaraj, et al. “Beyond the Screen: Parents’ Experiences with Student Activity Monitoring in K-12 Schools.” Center for Democracy & Technology, 28 July 2023, cdt.org/wp-content/uploads/2023/07/2023-07-28-CDT-Civic-Tech-impacts-of-student-surveillance-report-final.pdf.

Thompson, Polly. “Dell’s AI Business Is Booming, but Shares Plunged After It Cut Its Revenue Outlook.” Business Insider, 27 Nov. 2024, https://www.businessinsider.com/dell-earnings-report-q3-shares-fall-revenue-ai-servers-2024-11.

Trinh, Trieu H., et al. “Solving Olympiad Geometry Without Human Demonstrations.” Nature, vol. 625, 2024, pp. 476–482, doi:10.1038/s41586-023-06747-5.

“US Raises Concerns to Chinese Officials about AI Misuse.” Reuters, 15 May 2024, www.reuters.com/technology/us-china-hold-ai-risk-safety-talks-white-house-says-2024-05-15/.

Wei, Jason, et al. “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models.” arXiv, 10 Jan. 2023, https://arxiv.org/abs/2201.11903.

Welch, Nicholas. “Litho World & Commerce: Lost in Translation?” ChinaTalk, 1 Nov. 2023, www.chinatalk.media/p/litho-world-and-commerce-lost-in.

Williams, Rhiannon. “The Way Whales Communicate Is Closer to Human Language Than We Realized.” MIT Technology Review, 7 May 2024, https://www.technologyreview.com/2024/05/07/1092127/the-way-whales-communicate-is-closer-to-human-language-than-we-realized/.

Winter-Levy, Sam. “The Emerging Age of AI Diplomacy.” Foreign Affairs, 28 Oct. 2024, www.foreignaffairs.com/united-states/emerging-age-ai-diplomacy.

Wolfe, Cameron R. “Scaling Laws for LLMs: From GPT-3 to o3.” Deep (Learning) Focus, 6 Jan. 2025, https://cameronrwolfe.substack.com/p/llm-scaling-laws.

Wong, Queenie, and Wendy Lee. “Tech Companies Use AI to Fight Wildfires in California.” Los Angeles Times, 21 Jan. 2025, https://www.latimes.com/business/story/2025-01-21/tech-wildfires-ai-la-fires-nvidia-lockheed-martin.

Woodall, Tatyana. “Researchers Developing AI to Make the Internet More Accessible.” Ohio State University, 9 Jan. 2024, https://news.osu.edu/researchers-developing-ai-to-make-the-internet-more-accessible/.

Xiao, Chaojun, et al. “Configurable Foundation Models: Building LLMs from a Modular Perspective.” arXiv, 4 Sep. 2024, https://arxiv.org/abs/2409.02877.

“Densing Law of LLMs.” arXiv, 6 Dec. 2024, https://arxiv.org/abs/2412.04315.

Yang, Michael, et al. “Deconstructing Demographic Bias in Speech-Based Machine Learning Models for Digital Health.” Frontiers in Digital Health, vol. 6, 24 July 2024, doi:10.3389/fdgth.2024.1351637.

Zeeberg, Amos. “AI Needs Enormous Computing Power. Could Light-Based Chips Help?” Quanta Magazine, 20 May 2024, www.quantamagazine.org/ai-needs-enormous-computing-power-could-light-based-chips-help-20240520/.

Zhang, Jianguo, et al. “xLAM: A Family of Large Action Models to Empower AI Agent Systems.” arXiv, 5 Sept. 2024, https://arxiv.org/abs/2409.03215.

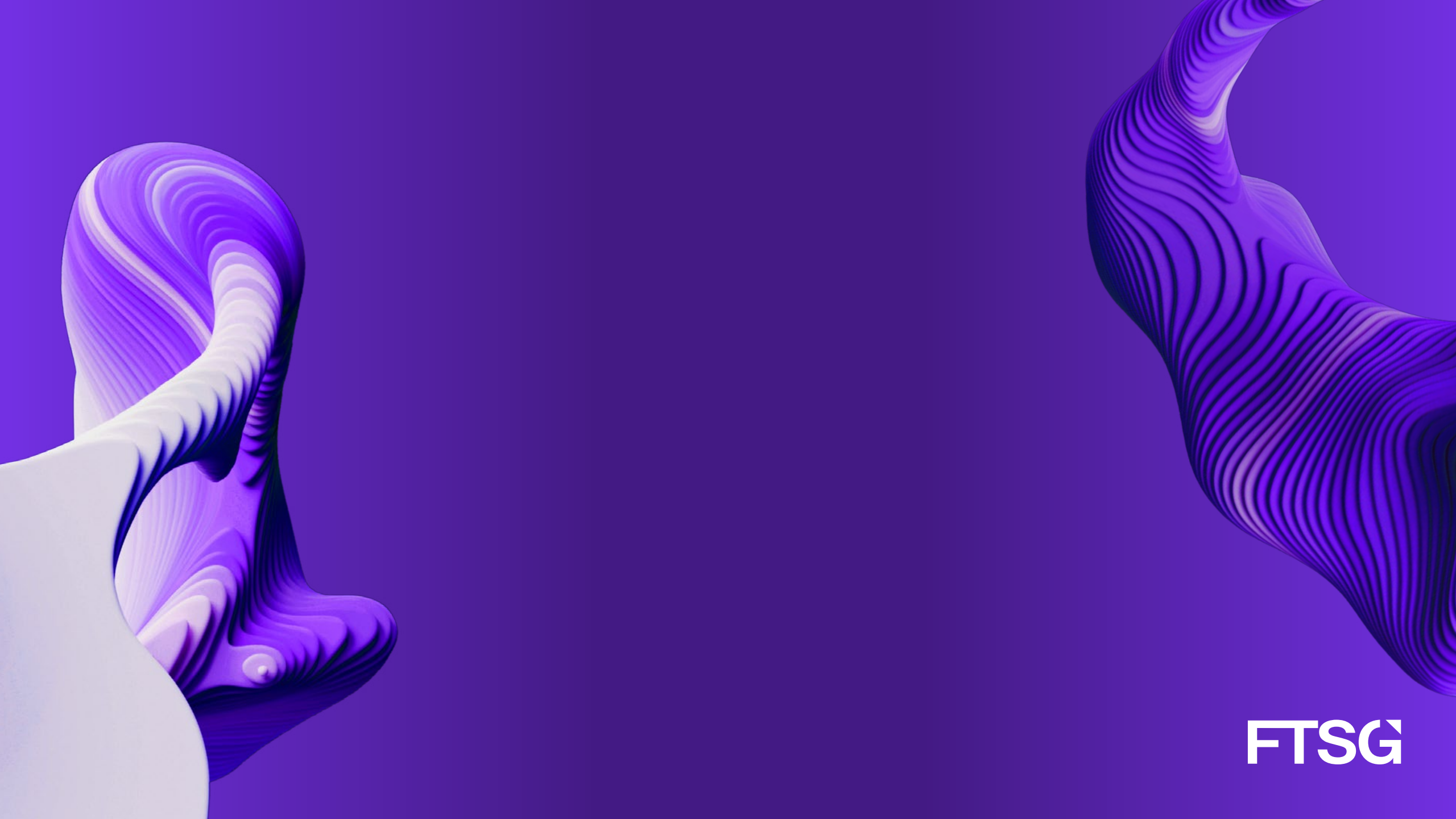
Zhang, Jiayi Eurus, et al. “Simulating Emotions with an Integrated Computational Model of Appraisal and Reinforcement Learning.” CHI ‘24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, Article No. 703, pp. 1-12, doi:10.1145/3613904.3641908.

Zhao, Lingyi, et al. “Detection of COVID-19 Features in Lung Ultrasound Images Using Deep Neural Networks.” Communications Medicine, vol. 4, no. 1, 2024, https://doi.org/10.1038/s43856-024-00463-5.

Zhou, Yukun, et al. “A Foundation Model for Generalizable Disease Detection from Retinal Images.” Nature, vol. 622, pp. 156–163, 13 Sept. 2023, https://doi.org/10.1038/s41586-023-06555-x.

Zimmer, Marc. “Machine Learning Cracked the Protein-Folding Problem and Won the 2024 Nobel Prize in Chemistry.” The Conversation, 9 Oct. 2024, https://theconversation.com/machine-learning-cracked-the-protein-folding-problem-and-won-the-2024-nobel-prize-in-chemistry-240937.

Zuckerberg, Mark. “Open Source AI Is the Path Forward.” Meta Newsroom, 23 July 2024, https://about.fb.com/news/2024/07/open-source-ai-is-the-path-forward/.


FTSG