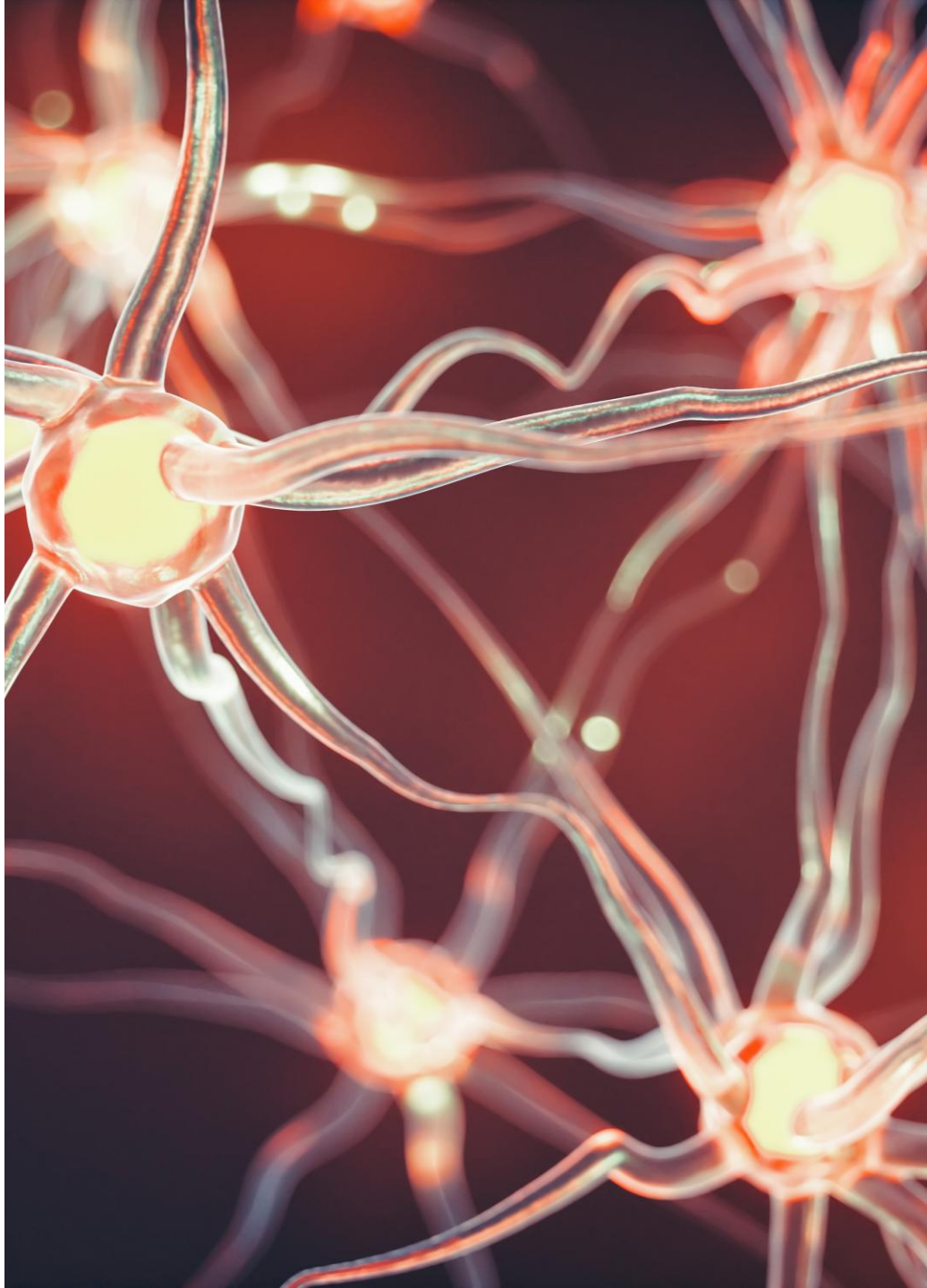
UNDERSTANDING AI

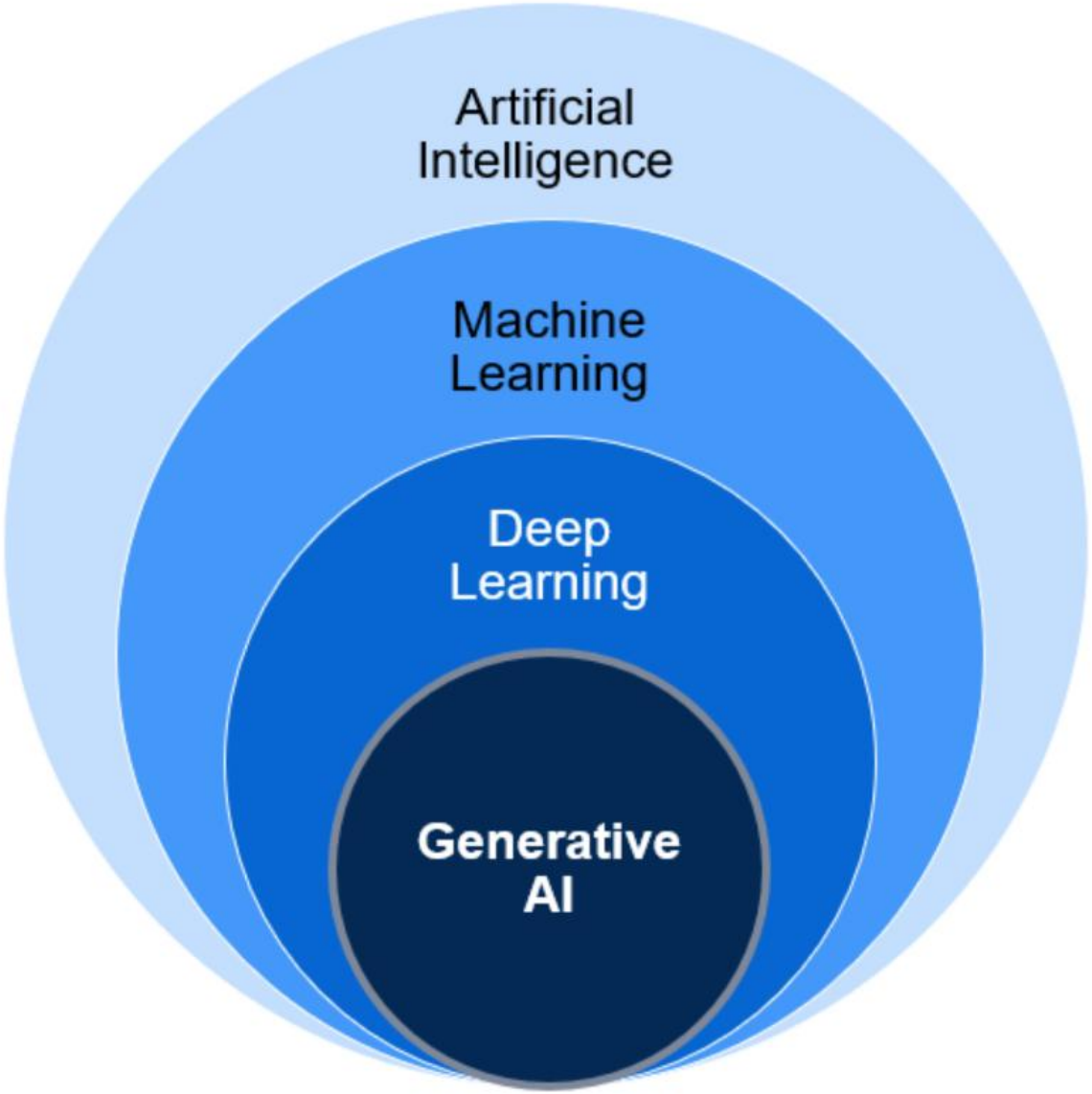
August 7,2025

Lisa O'Connor

CEO, Decoded Futures



WHAT IS AI



AI is the broad field of building machines that can mimic human intelligence.

Machine Learning (ML) is a subset where systems learn from data.

Deep Learning (DL) is a further subset using neural networks.

NLP (natural language processing) a part of ML and Deeplearning that handles human language and powers tools like ChatGPT.

Generative AI (GenAI) is a specialized area within deep learning that focuses on creating new content.

PatsyPoole, SAS Communities Library

AI MILESTONES

1950 – ALAN TURING PUBLISHES “COMPUTING MACHINERY AND INTELLIGENCE”, PROPOSING THE TURING TEST.

1951 – MARVIN MINSKY AND DEAN EDMONDS BUILD SNARC, THE FIRST ARTIFICIAL NEURAL NETWORK.

1956 – DARTMOUTH CONFERENCE: JOHN MCCARTHY COINS THE TERM "ARTIFICIAL INTELLIGENCE".

1958 – FRANK ROSENBLATT DEVELOPS THE PERCEPTRON; JOHN MCCARTHY INVENTS LISP.

1961 – UNIMATE, THE FIRST INDUSTRIAL ROBOT, BEGINS WORK AT GM.

1966 – ELIZA CHATBOT LAUNCHED BY JOSEPH WEIZENBAUM AT MIT.

1997 – IBM DEEP BLUE DEFEATS WORLD CHESS CHAMPION GARRY KASPAROV.

2011 – IBM WATSON WINS JEOPARDY!, DEMONSTRATING ADVANCED NLP.

2012 – UNIVERSITY OF TORONTO’S DEEP LEARNING SYSTEM ACHIEVES BREAKTHROUGH IMAGENET PERFORMANCE.

2016 – GOOGLE DEEPMIND’S ALPHAGO DEFEATS GO CHAMPION LEE SEDOL.

2020 – DEEPMIND’S ALPHAFOLD 2 SOLVES PROTEIN FOLDING CHALLENGE.

2022 – CHATGPT AND DALL·E RELEASED, CATALYZING GENERATIVE AI BOOM.

2023 – OPENAI LAUNCHES GPT-4 WITH MULTIMODAL CAPABILITIES.

2024 – ALPHAFOLD 3 EXPANDS PROTEIN PREDICTION; EU AI ACT PASSES.

2025 – RISE OF AGENTIC AI SYSTEMS (E.G., AUTOGPT) CAPABLE OF AUTONOMOUS TASK EXECUTION.

Model	Latest Version(s)	Context Window	Multimodal	Real-Time Web Access	Standout Features	Best For	Typical Weakness	Knowledge Cutoff
ChatGPT	GPT-4.5, o3, GPT-4o	128K–1M tokens	Yes (Text, Image)	Plugin integration	Broad capabilities, plugins, code, images	General-purpose, broad toolkit	Occasional hallucinations	June 2024
Claude	3.7 Sonnet, 3.5 Sonnet	~200K tokens	Yes (Text, Image)	No	Structured, safe reasoning, large context	Coding, factual content, reasoning	Less direct web search	April 2024
Gemini	2.5 Pro, 2.5 Flash	Up to 1M tokens	Yes (Text, Image, Audio)	Yes (some versions)	Fastest, strong with images and audio	Research, long-context tasks	Less open, API availability	Dec 2023
Perplexity	Perplexity Web, Labs	100K+ tokens*	Yes (Text, Image)	Yes	Real-time web search, citations	Fact-checking, research	Less creative writing cap.	Ongoing, near real-time
Grok	v3	Up to 1M tokens	Yes (Text, Image)	Yes	Deep reasoning, humor, rapid updates	Math, reasoning, ongoing data	Sometimes irreverent, inconsistent	Dec 2023
Llama	4/3.3	Up to 10M tokens	No (Mostly Text)	Add-ons via API	Massive context, open model for devs	Document-heavy, research, custom	Not as refined UI	Dec 2023
DeepSeek	V3, Janus Pro	64K–128K tokens	Yes (Text, Image)	Some	High price efficiency	Devs, cost-effective inference	Smaller ecosystem	Jan 2024
Qwen	Qwen-2	128K tokens	Yes**	N/A	Multilingual, open license	Multilingual tasks, cost sensitive	Limited doc ecosystem	2024

Model Name	Strengths	Token Pricing (Prompt / Completion)	Modalities (In/Out)	Ideal Use Cases
GPT-3.5 Turbo	Fast, reliable for simple tasks; low cost.	~$0.0005 / $0.0015 per 1K tokenshelp.openai.com	Text input / Text output (voice via app)	High-volume chatbots, FAQs, straightforward queries.
GPT-4 (orig.)	Very high accuracy and creativity; strong reasoning (vs. GPT-3.5).	$0.03 / $0.06 per 1K (8K context)help.openai.com	Text & Image input / Text output	Complex Q&A, creative writing, tutoring
GPT-4 Turbo	Updated knowledge (Apr 2023); much cheaper and faster than early GPT-4.	~$0.01 / $0.02 per 1K (3×/2× cheaper vs GPT-4)help.openai.com	Text & Image input / Text output	General-purpose use with long documents; up-to-date answers (for 2022–2023 info).
GPT-4o ("omni")	Multimodal (accepts text, images, audio, video); very fast responses; great non-English and audio/visual understandingopenai.com.	~$0.005 / $0.01 per 1K (50% cheaper vs Turbo) openai.comtechtarget.com	Text, Image, Audio in / Text, Image, Audio out	Rich multimedia assistants, real-time voice/chat agents, cross-language tasks.
GPT-4o mini	Small & efficient; good general intelligence for its size; cheapest model per token.	**$0.00015 / $0.00060** per 1Kopenai.com	Text, Image in / Text, *(Image out via DALL·E)*	Large-scale deployments (customer support, indexing long data) where low cost is crucial.
OpenAI o3	Expert **reasoning** model; excels at multi-step problem solving, math, code (fewest errors in complex tasks)help.openai.com.	~$0.01 / $0.04 per 1K artificialanalysis.ai	Text & Image in / Text out	Difficult analytic tasks (scientific research, coding challenges, financial analysis) requiring step-by-step logic.
GPT-4.5 (preview)	Extremely advanced conversational ability; broad knowledge; "feels human" in dialogue techtarget.comhelp.openai.com.	$0.075 / $0.15 per 1K (preview pricing)techtarget.com	Text & Image in / Text out	Experimental uses: e.g. as a highly natural virtual assistant, creative collaborator (limited preview only).
GPT-4.1	**State-of-the-art** general model – top coding skills, excellent instruction following, very aligned; huge contextopenai.comopenai.com.	~$0.0025 / $0.01 per 1K (estimated)techtarget.com	Text, Image, Audio, Video in / Text, Image, Audio out	Any complex task (coding, research, writing) especially where entire knowledge bases or lengthy materials are in context. Powers new AI "agent" applicationsopenai.com.
GPT-4.1 mini	Next-gen small model – approaches GPT-4o performance at a fraction of cost; very low latencyopenai.com.	**83% cheaper** than GPT-4o (planned pricing)openai.com	Text, Image in / Text out	Scalable deployment for moderately complex tasks; interactive apps needing fast responses
GPT-4.1 nano	Ultra-fast micro model; lowest cost; decent performance on basic tasksopenai.comopenai.com.	(To be announced)	Text, Image in / Text out	Real-time and embedded AI use (e.g. smart devices, low-latency services) where speed is wins over full accuracy.

CLASSICALLY CODED APPLICATIONS

VS

AI – POWERED APPLICATIONS

	Classically Coded Applications	AI-Powered Applications
How are they Built		
Development	Built using explicit instructions, rules, and logic defined by a programmer	Built by training algorithms on large datasets to recognize patterns
Instructions	Use traditional programming logic: conditional statements, loops, etc.	Use models like neural networks that learn from data and adjust parameters
Data Use	Data is static input	Data is used to train and optimize performance
How do they Function		
Behavior	Deterministic – same input gives same output	Probabilistic – outcomes can vary with same input depending on learned weights
Adaptability	Requires manual updates by programmers	Can continuously learn or be retrained on new data
Transparency	Code is traceable and explainable	Models can be opaque (“black box” issue)
Use Cases		
Best For	Routine, rule-based systems	Complex, uncertain, or high-dimensional problems
Examples	Accounting software, calculators, CRMs	Face recognition, chatbots, autonomous driving
Summary		
Instructions	Hard-coded rules	Learned patterns
Logic	Explicit	Inferred from data
Behavior	Predictable	Statistical/probabilistic

AI LANDSCAPE IN 2025

The AI landscape has shifted dramatically over the past year:

- **Multimodal AI** is now standard: Systems can seamlessly process and generate combinations of text, images, audio, and video. Examples include AI agents that can read documents, generate a summary, and present the findings visually.
- **Autonomous AI agents** are a breakthrough: These can now manage projects, send emails, perform web searches, and update spreadsheets—without human prompting. This is transforming roles in operations and admin-heavy departments.
- **Generative AI ecosystems** are rapidly growing:
  - **OpenAI's GPT** models power many enterprise tools, but other models have built followings
  - **Open-source LLMs** like Meta's LLaMA 3, Mistral, and xAI's Grok are competitive and contenders
  - **DeepSeek R1 and R1-Zero** have disrupted the marketplace with more efficient processing (lower cost) through Mixture of Expert (MoE) which uses a subset of parameters for processing
  - Many companies are fine-tuning small LLMs (3B–7B parameters) for specific tasks to reduce latency and protect IP.
- **Vertical-specific AI tools** are booming:
  - Law: AI legal research and contract analysis
  - Healthcare: AI-assisted diagnostics and patient triage
  - Finance: Autonomous reporting, real-time fraud detection
- **Race for AI Talent** across firms
- The shift from "AI pilots" to **enterprise-wide AI strategy** is now a defining characteristic of leading firms.
- Businesses are focusing on integrating AI into entire workflows, including compliance, reporting, and customer-facing tools.
- The emergence of **AI-native startups** is also accelerating disruption across sectors

RISKS AND SECURITY

The power of AI comes with **real responsibility**. Key risk areas:

- **Bias and fairness:**
  - AI learns from data. If your historical data is biased, your AI will be too.
  - HR systems, credit scoring, and law enforcement tools have shown systemic bias issues.
  - Businesses need procedures to audit models regularly.
- **Security and adversarial attacks:**
  - Prompt injection, model hijacking, data poisoning and data leakage are growing concerns.
  - Ensure all AI integrations are vetted through by Security controls, Red Teaming and tested under load and edge cases.
  - Growing ecosystem of tools/vendor for model and security monitoring, traceability and explainability
- **Responsible AI**
  - Human by Design, Fairness, Transparency, Explainability and Accuracy, Safety, Accountability, Compliance, Data Privacy, Data Confidentiality and Cyber Security, Robustness, Sustainability
- **Hallucinations:**
  - Even the most advanced LLMs may generate plausible but incorrect information.
  - Slopsquatting – adversaries taking advantage of vibe coding hallucinations

What is different about protecting LLMs

Data

LLMs use vast amounts of training data. Do we know the quality and nature of the Training Data? Is there transparency sharing what data was used to train the foundational model?

Access

Access to a model allows interaction with the model which could result in directly training the model. With malicious or unintentional intent, the model could learn “bad” things. Unintentional consequences or worse poisoning.

Authorization

Access to the model is de facto authorization to all the responses for most LLMs. How do you manage privacy of inquires to the model or from other authorized users?

Supply chain

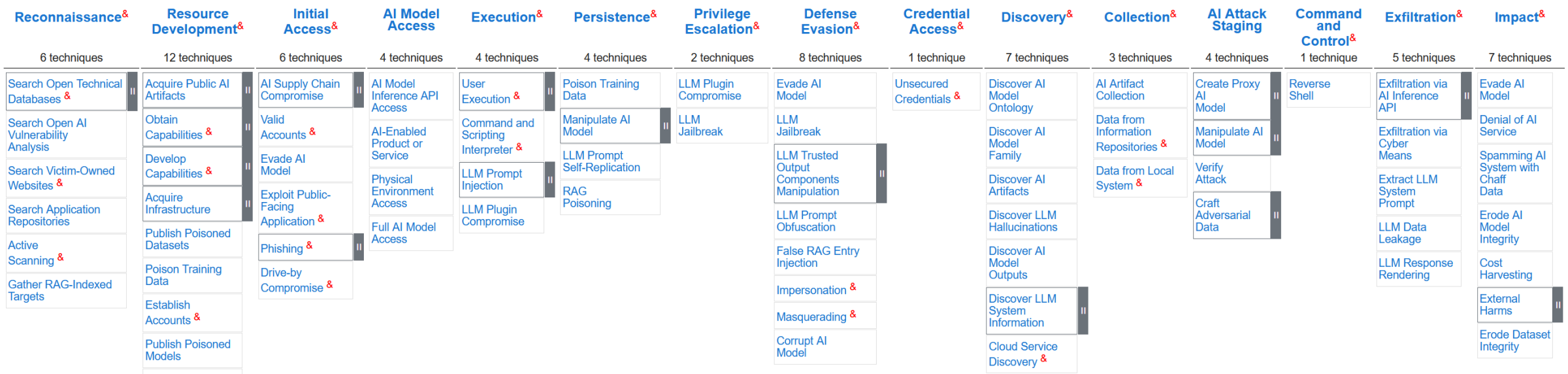
Model developers should ensure the model software development life-cycle has protections throughout the lifecycle: Trusted providers, trusted methods, trusted community participants.

Retraining Lifecycle

LLM multimodal models are “active” in creating content, code, images, video and learning. By asking questions, the model can learn. By providing it data, the model can learn. How and when is the model infused with more data for learning? Is there a fall back to a known state if it learns something it shouldn’t? Model lifecycle management, Failsafe AI, AI BOMs

ATLAS Matrix

The ATLAS Matrix below shows the progression of tactics used in attacks as columns from left to right, with ML techniques belonging to each tactic below. & indicates an adaption from ATT&CK. Click on the blue links to learn more about each item, or search and view ATLAS tactics and techniques using the links at the top navigation bar. View the ATLAS matrix highlighted alongside ATT&CK Enterprise techniques on the ATLAS Navigator.



MITRE ATLAS Framework
The ML Tactics that Adversaries Use to Attack AI Systems

https://atlas.mitre.org/matrices/ATLAS

SIGNIFICANT REGULATIONS

European Union – EU AI Act (in force 2025):

- Categorizes AI systems by risk: unacceptable, high-risk, limited-risk, and minimal-risk.
- High-risk systems (e.g., HR tools, finance scoring) must comply with rigorous transparency, testing, documentation, and audit requirements.
- Bans some AI use cases (e.g., real-time facial recognition in public spaces).

United States – Patchwork, but progressing:

- No federal AI law yet, but significant movement:
  - **Executive Order on AI (2023):** Sets directives for federal agencies to ensure safety, rights protection, and R&D. (NatLawReview)
  - **NIST AI Risk Management Framework:** Encouraged for enterprise adoption. (NIST)
  - FTC enforcement on deceptive AI use cases (e.g., deepfakes, bias).

United Kingdom – Context-based, pro-innovation:

- No standalone AI Act yet; uses existing laws (Data Protection Act, Equality Act).
- Regulators (ICO, CMA) have published joint guidance on AI risks.

China – Content control and accountability:

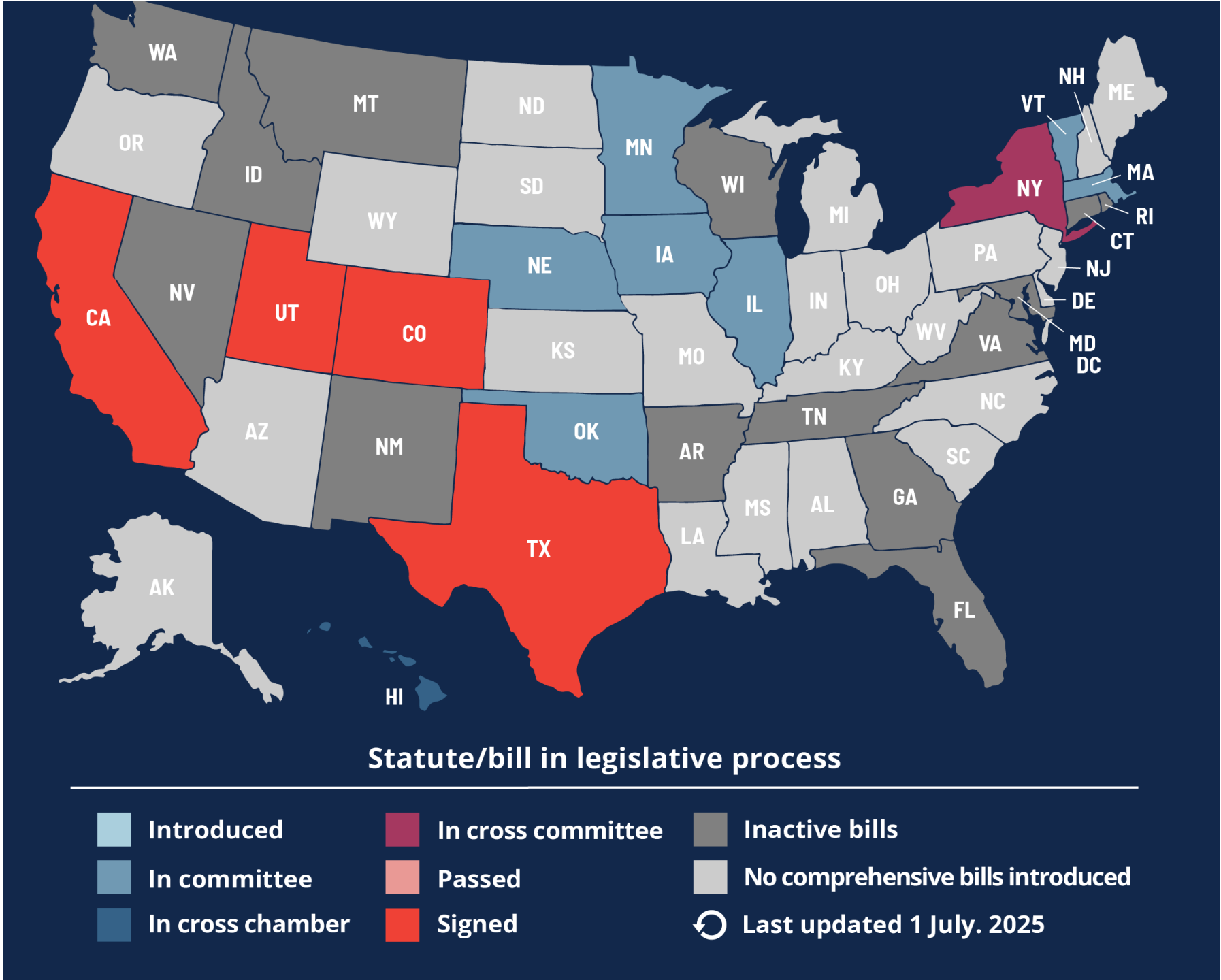
- **Generative AI Regulation (2023):** Requires labeling of AI-generated content, model registration, and value alignment.
- **Cybersecurity Law, 2016**

Canada, Australia, Singapore:

- National AI codes of conduct or voluntary frameworks.
- Australia moving toward mandatory risk assessments for high-impact AI.

iapp

US STATE AI GOVERNANCE LEGISLATION TRACKER



https://iapp.org/resources/article/us-state-ai-governance-legislation-tracker/

HIGHLIGHTS

Privacy Laws that can regulate AI

- The California Privacy Protection Act (CPPA), which regulates automated decision-making
- The Biometric Information Privacy Act in Illinois,[37] which is very broad and allows for extremely high damages for violations.

Laws that regulate AI
Colorado AI Act, May 2024

- The Act focuses on automated decision-making systems and defines a covered high-risk AI system which contribute or factor into consequential decisions, the developer and deployers must use reasonable care to avoid discrimination via AI

California

- September 2024, various bills: Defending Democracy from Deepfake Deception Act; Use of Likeness: Digital Replica Act
- California AI Transparency Act

Tennessee

- Ensuring Likeness, Voice, and Image Security Act ("ELVIS Act")

Texas

- Texas Responsible Governance of AI Act June 2025, signed

Utah, Montana, Virginia, Washington

PENDING REGULATIONS

Monitor Regulations where you have business operations, partners or customers
Align with high watermark to have a single policy and practices

U.S. Federal Level:

- **Executive Order 14110 (2023)** – Directs agencies to set AI safety standards. (WhiteHouse)
- **Trump AI Action Plan (2025)** – Focus on international competitiveness and innovation. (WhiteHouse)

International Developments:

- **India:** Proposing ethical regulatory sandbox.
- **Brazil:** AI law modeled on EU AI Act.
- **Japan & South Korea:** Developing shared cross-border standards.

California (pending): Revived 2025 AI Bill of Rights + whistleblower protections (Politico)

Texas HB 149 AI Responsible Governance Act (signed, but poised to redact) July 2025

NEXT STEPS

Be a fearless learner

Get Hands-on, play

Follow the News in AI

Things are changing rapidly-for everyone

Trust your knowledge

Verify the insights you receive

Build new ways of working with AI

AI is Strategic to your business

Understand what your business's roadmap is for enabling people and processes with AI

Understand the short and long-term costs and ROI

Ask questions

INDUSTRY RESOURCES AND NEWS

https://mitsloanedtech.mit.edu/ai/

https://hai.stanford.edu/ai-index

https://www.technologyreview.com/topic/artificial-intelligence/

https://news.mit.edu/topic/artificial-intelligence2

https://www.wired.com/category/artificial-intelligence/

https://fortune.com/section/artificial-intelligence/

https://fortune.com/author/sharon-goldman/

https://techcrunch.com/category/artificial-intelligence/

https://aimagazine.com/

https://iapp.org/resources/article/global-ai-legislation-tracker/

https://www.nist.gov/artificial-intelligence

https://www.aisafetybook.com/textbook/artificial-intelligence-and-machine-learning