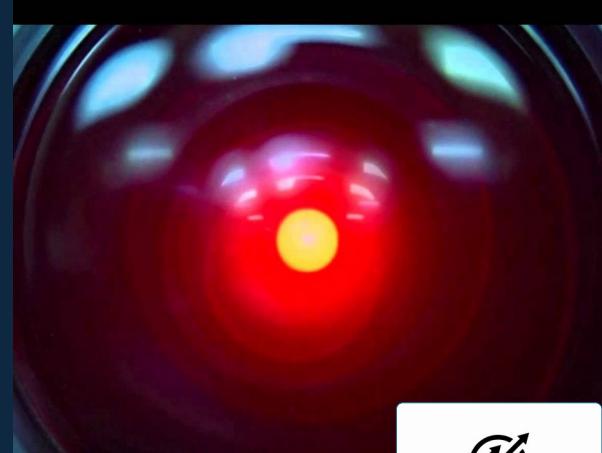
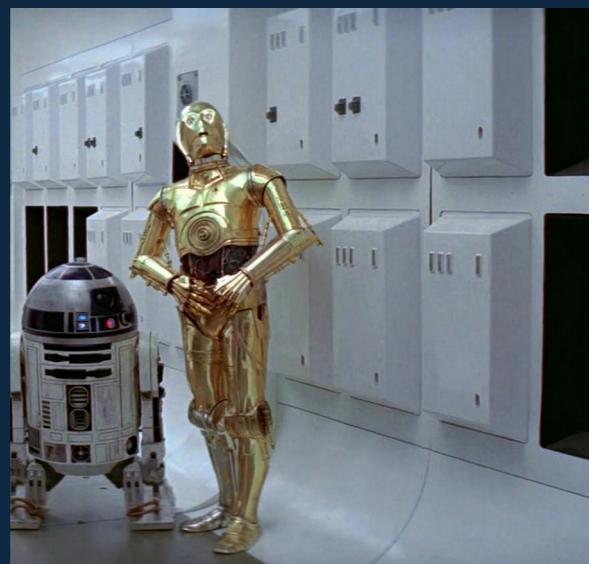


Advanced AI Risk - ISACA

It could be awesome, destroy all of humanity, or hype bigger than Y2k/dot.com; it won't be boring ...

- David Kohrell, NEbraskaCERT 18FEB2026
- <https://www.necert.org/CSF/>



KOHRELL

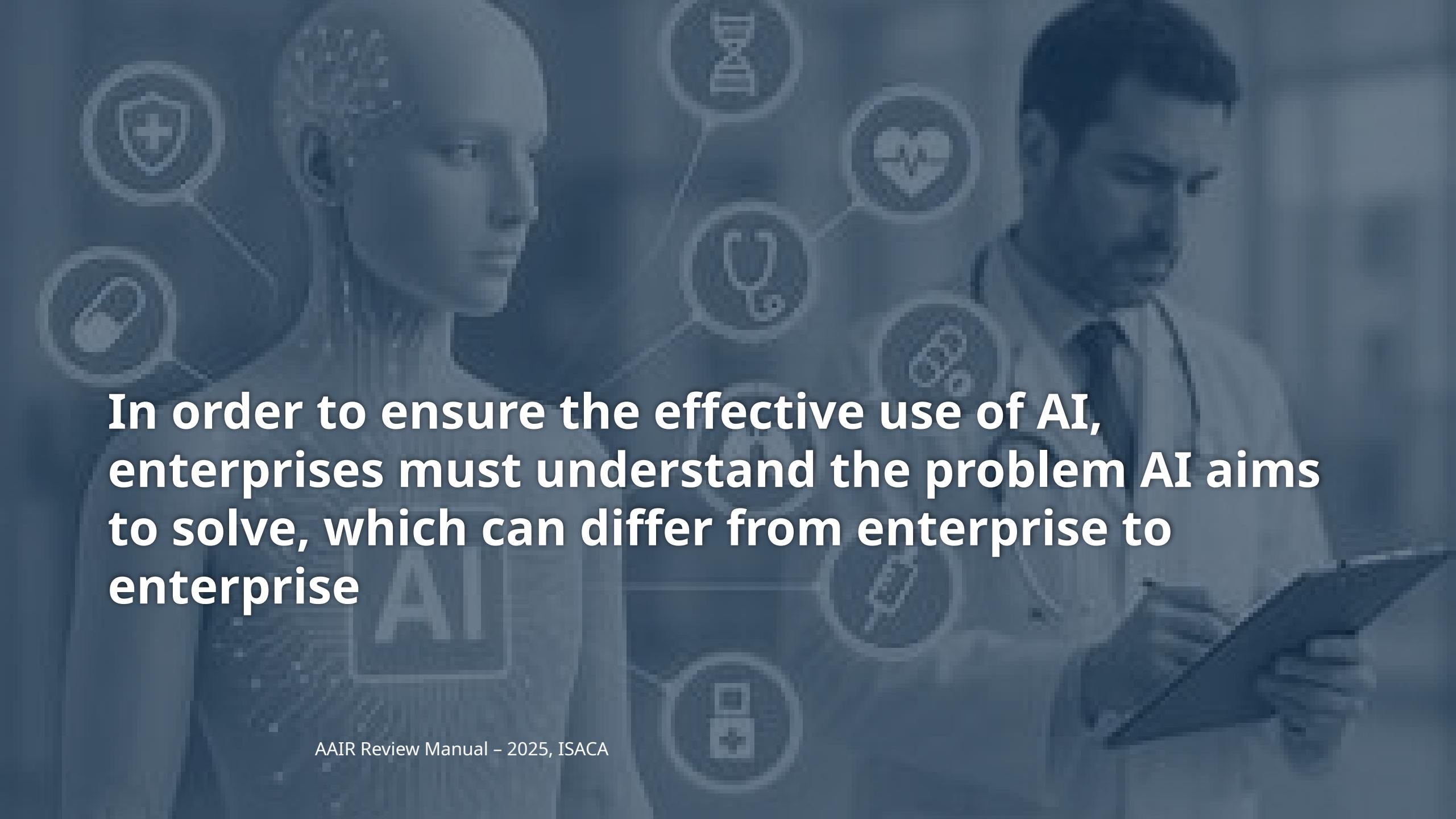
A photograph of a city street at night. The street is wet, reflecting the lights from the surrounding buildings. On the left, there's a building with large arched windows, some of which are covered in greenery. On the right, there's a building with a vertical sign that has some Japanese characters on it. A person is walking away from the camera down the street. The overall atmosphere is moody and atmospheric.

Definitions and References

Risk = Threat x Vulnerability x Asset

ISSA, Managing Risk in Information Systems, 2022
Gibson & Igonor

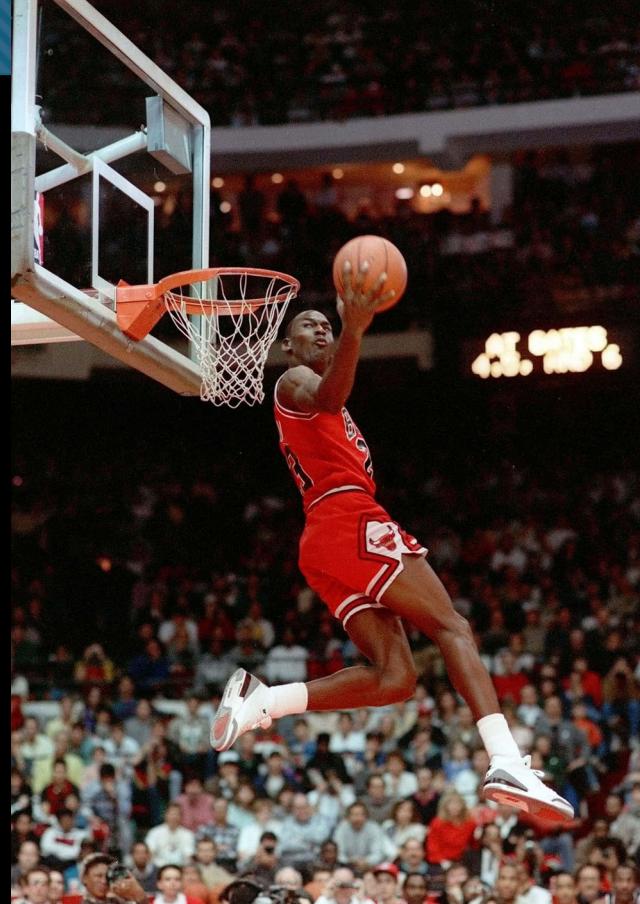
AI risk refers to the composite measure of an event's probability of occurring and the magnitude or degree of the consequences of the corresponding event. The impacts, or consequences, of AI systems can be positive, negative, or both and can result in opportunities or threats (Adapted from: ISO 31000:2018). NIST RMF, January 2023

A man in a suit is looking at a tablet. The tablet screen displays a network of medical icons, including a shield with a cross, a DNA helix, a heart, a stethoscope, a gear, and a plus sign. The background is a dark blue.

**In order to ensure the effective use of AI,
enterprises must understand the problem AI aims
to solve, which can differ from enterprise to
enterprise**

AAIR (Advanced AI Risk)

- AAIR is an AI-focused IT risk management credential crafted to enhance the expertise of certified IT risk professionals. It provides the strategic skills to guide management in addressing AI-related risks, safeguarding organizations from potential financial and reputational harm. AAIR's key practice areas include:
 - AI Risk Governance and Framework Integration
 - AI Life Cycle Risk Management
 - AI Risk Program Management



© Selftution.com

ANI vs AGI vs ASI



ANI

A Self Driving Car Stopping to Prevent Injury to the Pedestrian



AGI

A Robot Cooking Food Along with Solving a Maths Problem



ASI

A Super Intelligent Robot Leading an Army of Robots

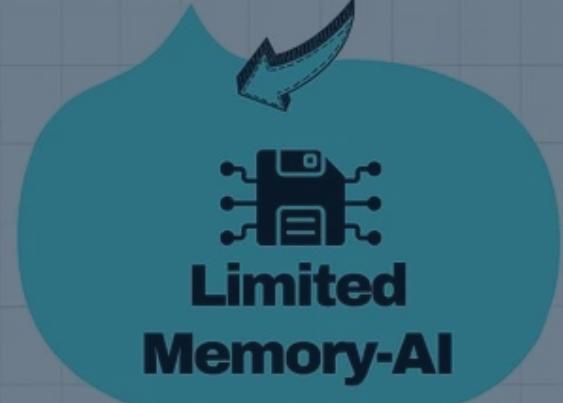
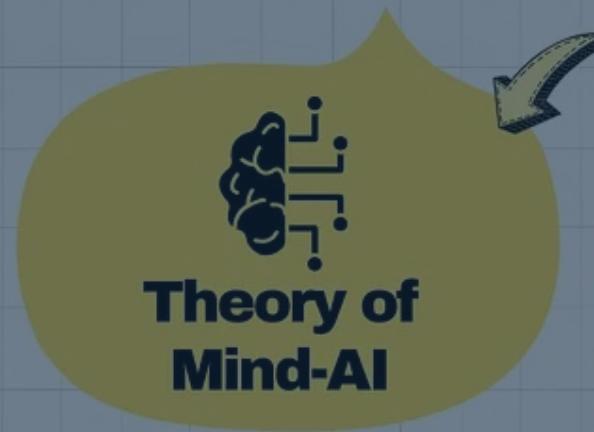
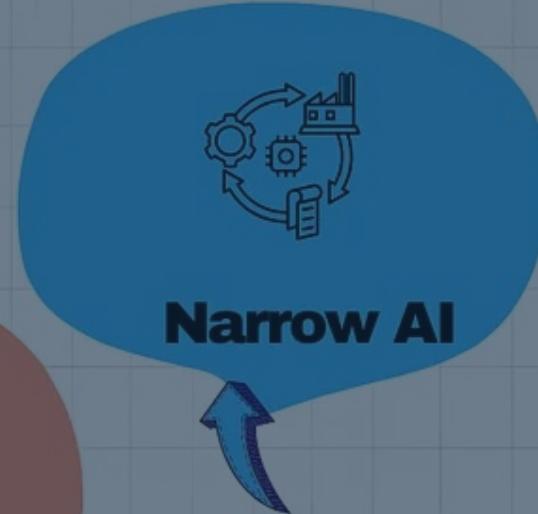
© Selftution.com

<https://selftution.com/types-artificial-intelligence-ai-examples-narrow-general-super-ani-agi-asi-machine-learning-ml/>

AI “levels”

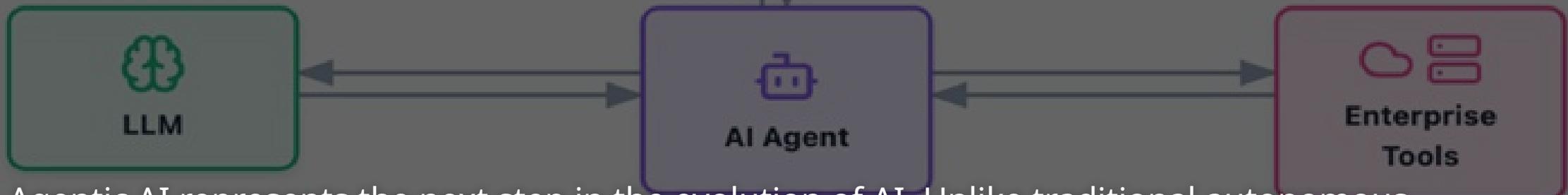
- Automated Narrow Intelligence – what we have now
- Automated General Intelligence – the hype driving investments and circular economy
- Automated Super Intelligence – Paradise or Machines turn on us. Awesome or Dystopian

4 MORE



**7 Types Of
AI**

“Agentic AI” ISACA



Agentic AI represents the next step in the evolution of AI. Unlike traditional autonomous systems that simply respond to prompts, agentic AI actively initiates actions and connects tasks with intent. Picture a digital assistant that can seamlessly execute verbal commands such as, “Create an account for a specific individual in the customer relationship management (CRM) solution, give them global read access, and send out a confirmation email when done.”





AI Risk Governance and Framework Integration

What Are the Risk Categories in the EU AI Act?

Unacceptable Risk:

- Prohibited due to significant ethical or safety concerns.
- Examples: Social scoring by governments, manipulative AI, biometric surveillance

High Risk:

- AI systems that significantly impact people's rights or safety.
- Requires conformity assessments, risk management, human oversight.
- Examples: AI used in critical infrastructure, education, employment, law enforcement, and healthcare.

General Purpose AI (GPAI):

- Broad AI systems that can be used across sectors and are subject to specific transparency and risk management requirements.
- Examples: Large AI models by companies like OpenAI. These models must comply with extra obligations when used in high-risk contexts.

Limited or Minimal Risk:

- Minimal risk to safety or fundamental rights, subject to limited transparency obligations (e.g., disclosing that users are interacting with an AI).
- Examples: Customer service chatbots, AI-based recommendations in e-commerce.



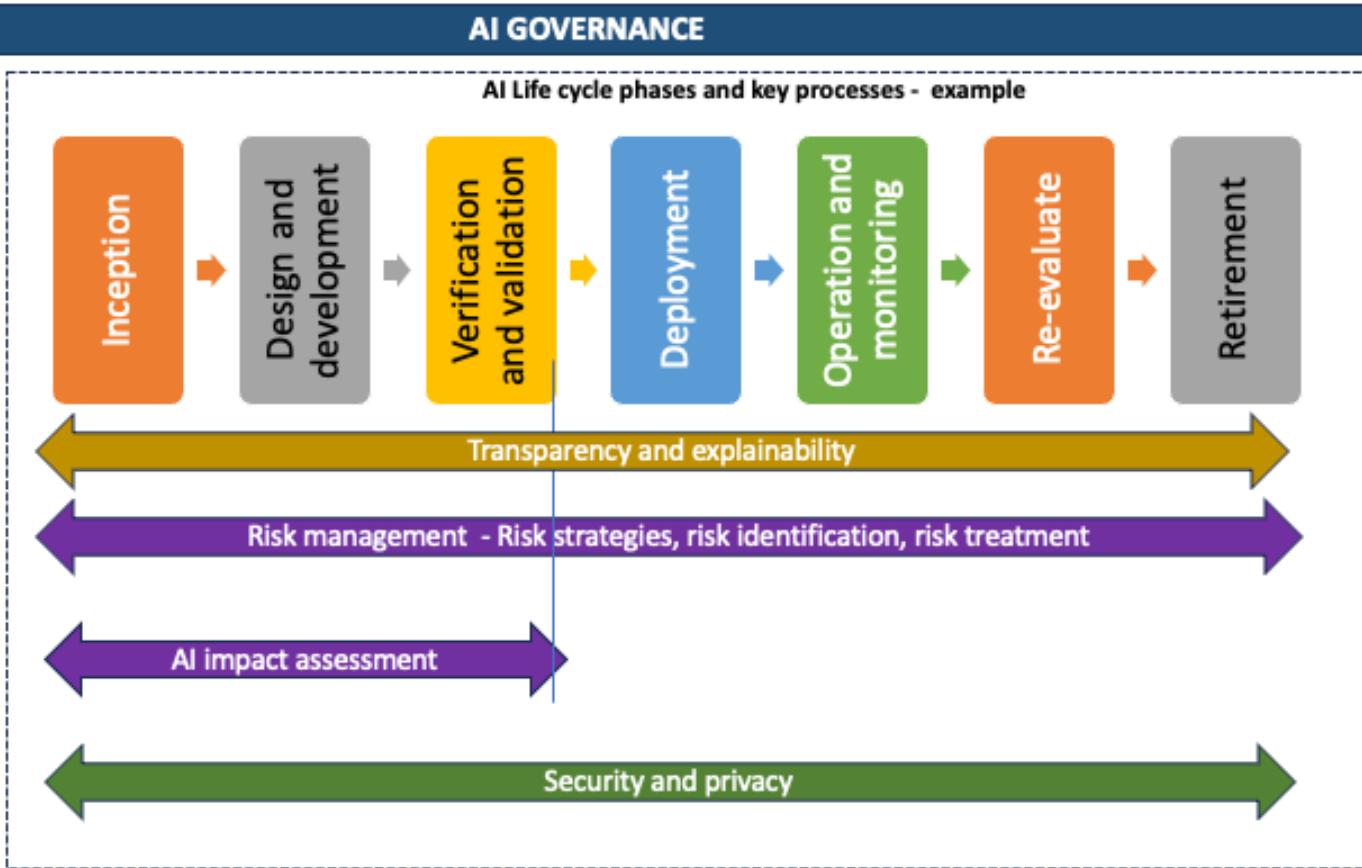
EU AI Act

[https://www.securitycompass.com/
blog/understanding-eu-ai-
act-risk-categories/](https://www.securitycompass.com/blog/understanding-eu-ai-act-risk-categories/)



NIST RMF

<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>



ISO/IEC 42001:2023 for AI

AWS Security Blog, Javid & Welch, May 2025

<https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/>

AI Life Cycle Risk Management



Four Governance Key's

Human in the Loop

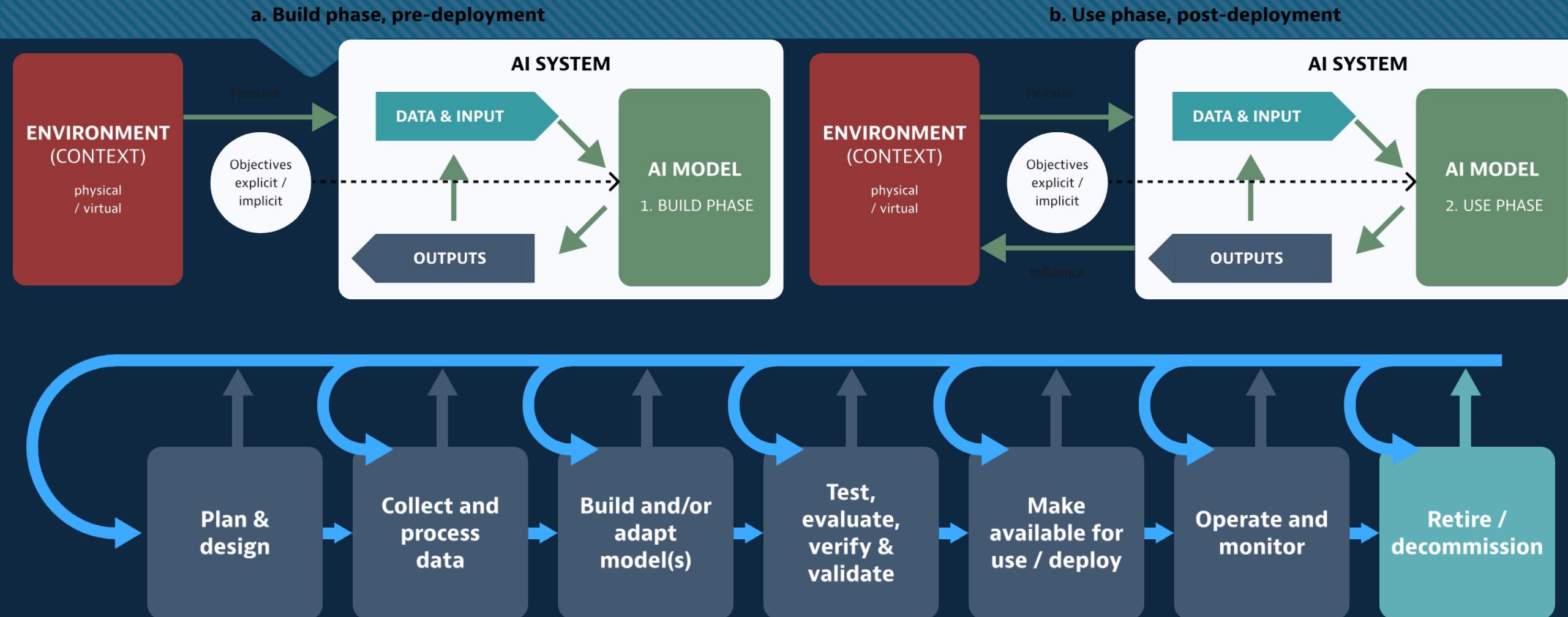
Transparency

Traceability

Explainability



OECD <https://oecd.ai/en/ai-principles> (2019 and 2024)



8 Data Types

Numeric
Categorical
Imager
Text
Time Series
Audio
Sensor

8 Data Types
That Major AI Models
Feed on to Function

AI Model Inventory





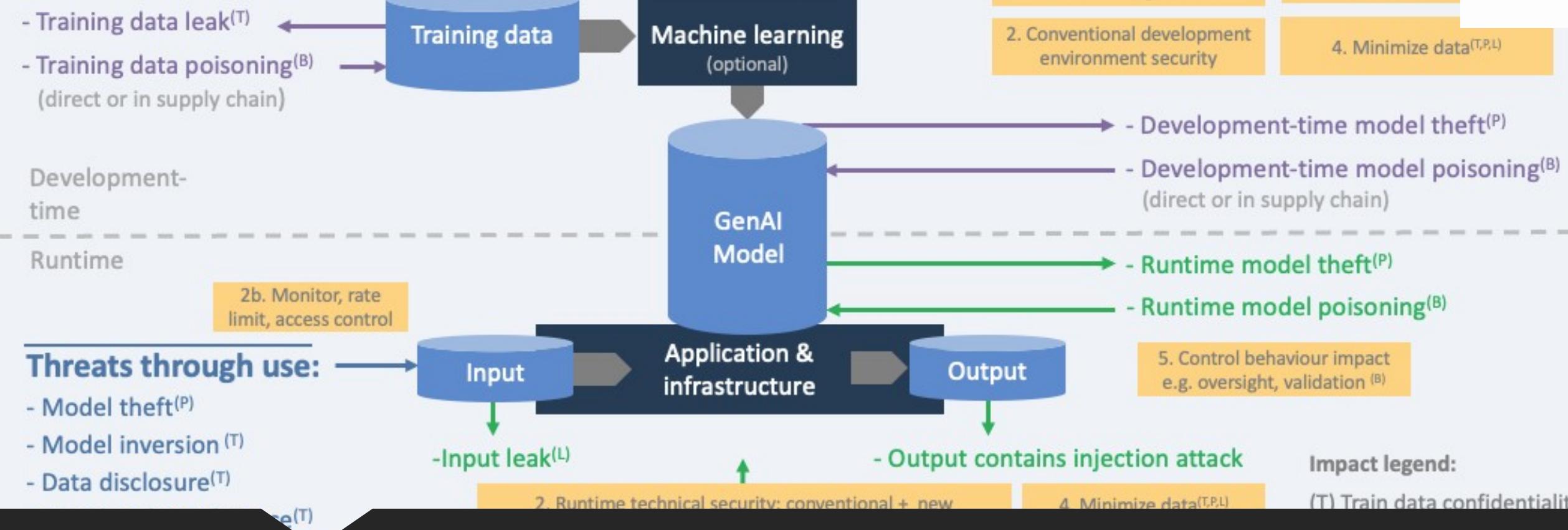
Risk Management Life Cycle

○ [https://www.alertmedia.com/
blog/risk-management-lifecycle/](https://www.alertmedia.com/blog/risk-management-lifecycle/)

A photograph of a city street at night. The street is wet, reflecting the lights from the surrounding buildings. In the background, there are large, arched windows on a building, and the words "COFFEE" and "LAB" are visible on the walls. A person is walking away from the camera down the street. The overall atmosphere is dark and moody.

AI Risk Program Management

Development-time threats



Development Time Threats – AI going bad

https://owaspai.org/docs/ai_security_overview/



Third Party Risk Management

- **Due diligence on vendor AI model** – to evaluate for vulnerabilities, training data quality and model performance with regards to biases
- **Contract review** – to define content ownership over output data and usage of input for training

Asset & Data Management

- **AI Inventory and dataflow mapping** – maintain inventory of AI systems, including components and dataflows
- **Training dataset quality** – for relevance, accuracy and bias
- **Data anonymisation** – to prevent personal information disclosure

Application Security

- **Secure coding practices** – input/output filtering, API access, securing secrets, and source code
- **Bias & fairness testing** – identify fairness metrics and test for edge case
- **Training pipeline and model protection** – access controls and segregation of environments to prevent tampering

Human Oversight

- **Human oversight design** – there are commensurate HITAL vs HOTL controls should be designed based on risks posed by AI use cases.
- **Contestability management** – where AI is involved in decision-making, there is a process for affected users to request manual review.

Identity & Access Management

- **Fine-grain permission setting and authentication** – ensure principle of least privilege and zero trust extends to AI agents. Use dynamic access controls (e.g. JIT) or two-agent check for privileged access
- **Secure AI-to-AI communication** – require authentication and encryption. Implement task segmentation

Security Awareness Training

- **AI Risk Awareness Training** – ensure users of AI systems are appropriately trained to detect for AI risks
- **AI-powered Social Engineering Attacks** – users are trained to spot for social engineering attacks that have been enhanced by AI (e.g. deepfakes, voice cloning)

Logging & Monitoring

- **Anomaly detection** – for AI attacks and unusual user/account behaviour
- **Data loss prevention** – covers AI tools as well as training pipeline
- **Model performance and traceability** – sufficient logging of model processing and output to support explainability and performance tuning

Incident Management & Response

- **AI system incident response playbooks** – update playbooks to respond to AI system incidents. Document log sources, dataflows, access, vendor support details
- **Backup management** – maintain secure history of backups in case data need to be re-trained

AI Threat Mitigation and Robustness

<https://cybercx.com.au/blog/designing-an-effective-ai-risk-mitigation-strategy/>

2025 OWASP Top 10 List for LLM and Gen AI

LLM01:25

Prompt Injection

This manipulates a large language model (LLM) through crafty inputs, causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.

LLM02:25

Sensitive Information Disclosure

Sensitive info in LLMs includes PII, financial, health, business, security, and legal data. Proprietary models face risks with unique training methods and source code, critical in closed or foundation models.

LLM03:25

Supply Chain

LLM supply chains face risks in training data, models, and platforms, causing bias, breaches, or failures. Unlike traditional software, ML risks include third-party pre-trained models and data vulnerabilities.

LLM04:25

Data and Model Poisoning

Data poisoning manipulates pre-training, fine-tuning, or embedding data, causing vulnerabilities, biases, or backdoors. Risks include degraded performance, harmful outputs, toxic content, and compromised downstream systems.

LLM05:25

Improper Output Handling

Improper Output Handling involves inadequate validation of LLM outputs before downstream use. Exploits include XSS, CSRF, SSRF, privilege escalation, or remote code execution, which differs from Overreliance.

LLM06:25

Excessive Agency

LLM systems gain agency via extensions, tools, or plugins to act on prompts. Agents dynamically choose extensions and make repeated LLM calls, using prior outputs to guide subsequent actions for dynamic task execution.

LLM07:25

System Prompt Leakage

System prompt leakage occurs when sensitive info in LLM prompts is unintentionally exposed, enabling attackers to exploit secrets. These prompts guide model behavior but can unintentionally reveal critical data.

LLM08:25

Vector and Embedding Weaknesses

Vectors and embeddings vulnerabilities in RAG with LLMs allow exploits via weak generation, storage, or retrieval. These can inject harmful content, manipulate outputs, or expose sensitive data, posing significant security risks.

LLM09:25

Misinformation

LLM misinformation occurs when false but credible outputs mislead users, risking security breaches, reputational harm, and legal liability, making it a critical vulnerability for reliant applications.

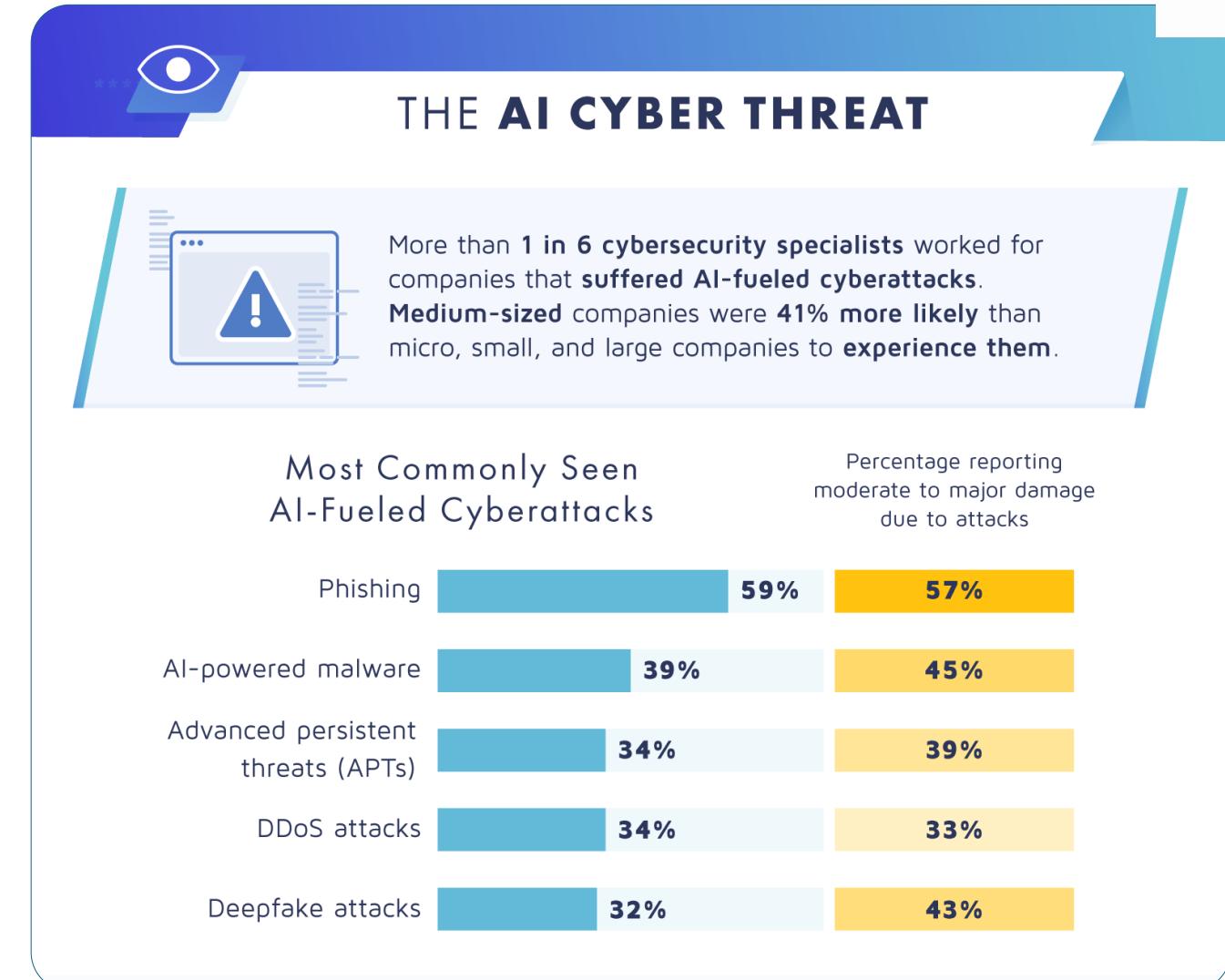
LLM10:25

Unbounded Consumption

Unbounded Consumption occurs when LLMs generate outputs from inputs, relying on inference to apply learned patterns and knowledge for relevant responses or predictions, making it a key function of LLMs.

AI doing bad things

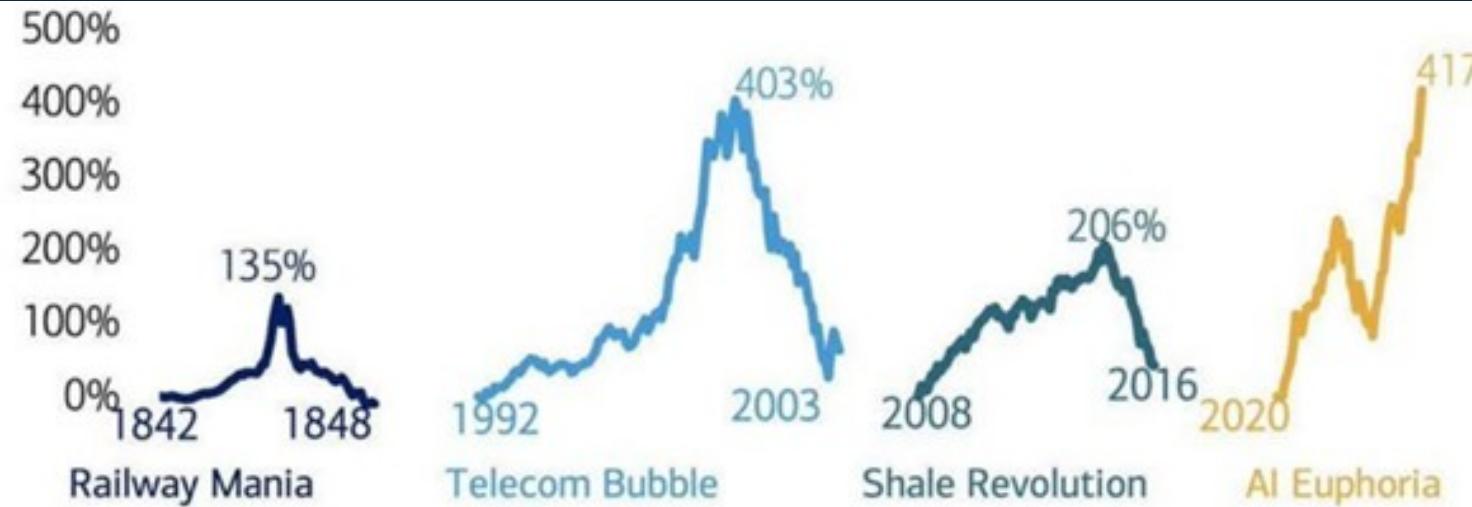
- Most Common Threats aided by AI
- <https://www.digitalinformationworld.com/2023/08/a-growing-threat-how-ai-poses-risks-to.html>



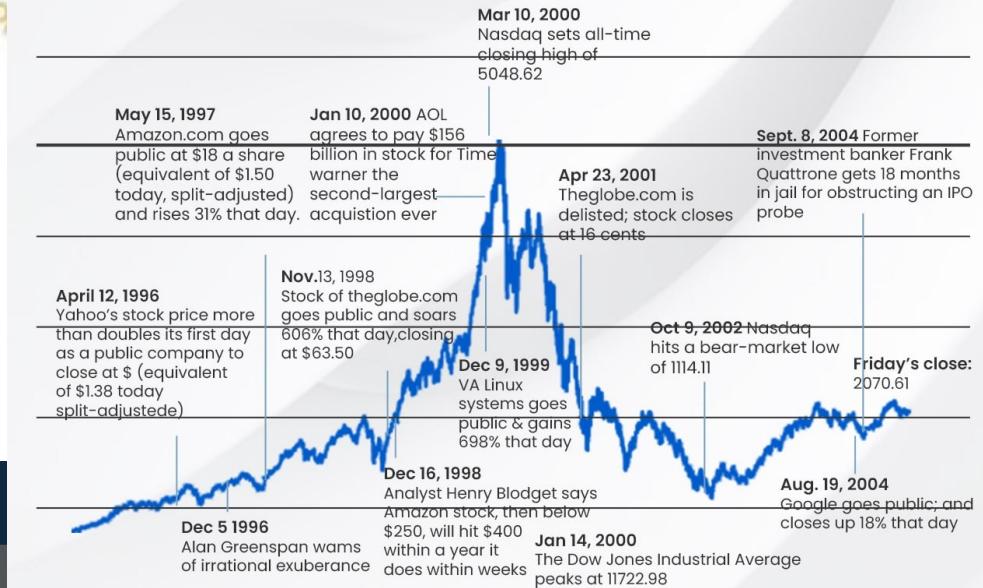
Two Final Thoughts

signature

Date



Dot-com bubble of 1999–2000



Strike

Economic Risk – Too big to fail What if?

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task[△]

Nataliya Kosmyna ¹

MIT Media Lab
Cambridge, MA

Eugene Hauptmann

MIT
Cambridge, MA

Ye Tong Yuan

Wellesley College
Wellesley, MA

Jessica Situ

MIT
Cambridge, MA

Xian-Hao Liao

Mass. College of Art
and Design (MassArt)
Boston, MA

Ashly Vivian Beresnitzky

MIT
Cambridge, MA

Iris Braunstein

MIT
Cambridge, MA

Pattie Maes

MIT Media Lab
Cambridge, MA

Impact on you

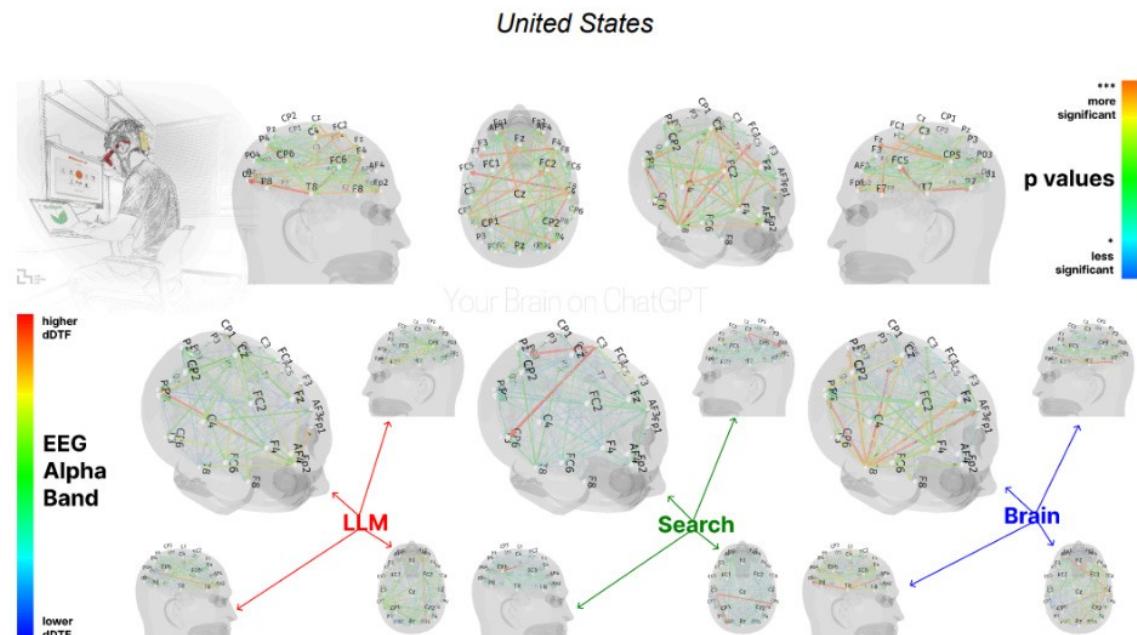
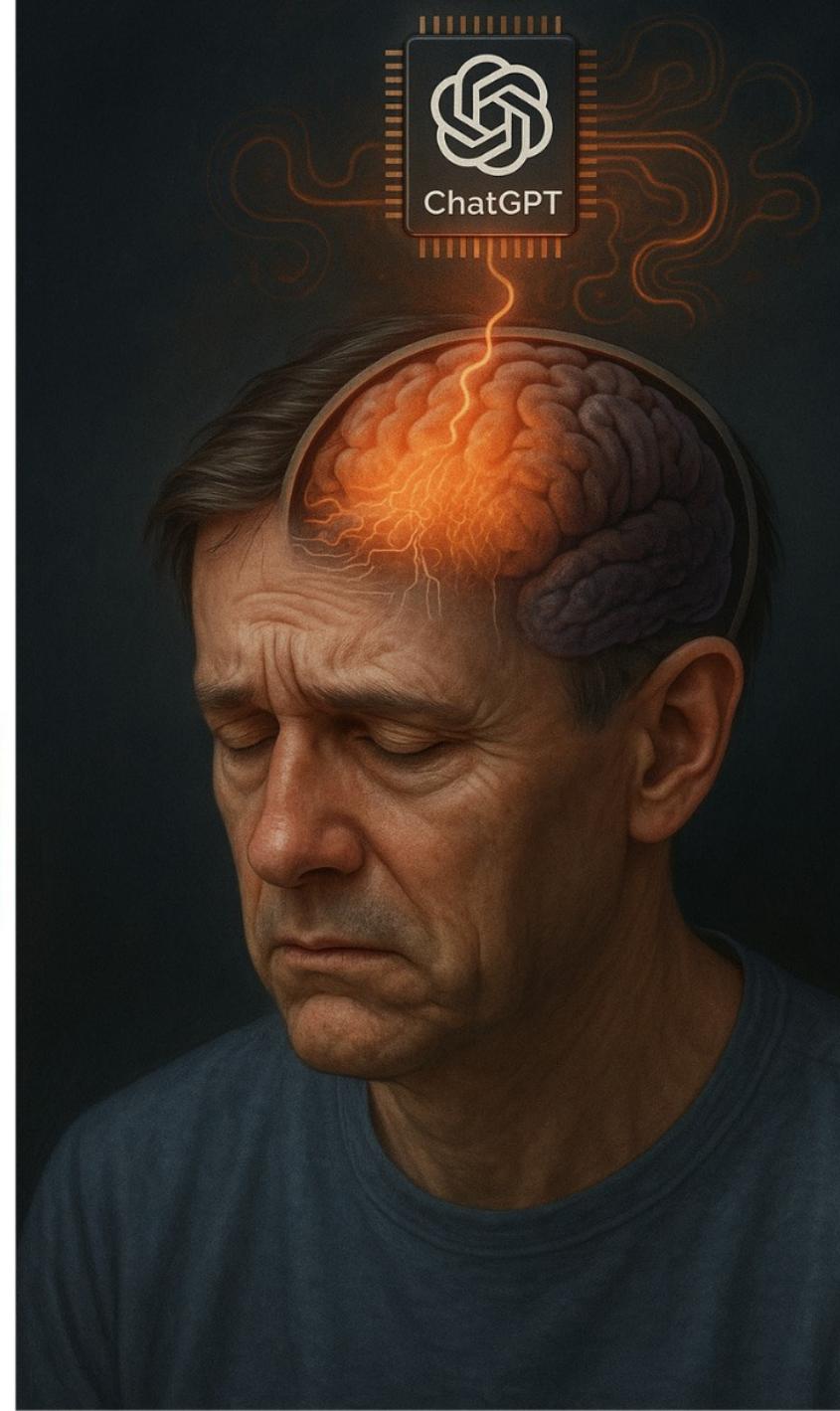


Figure 1. The dynamic Direct Transfer Function (dDTF) EEG analysis of Alpha Band for groups: LLM, Search Engine, Brain-only, including p-values to show significance from moderately significant (*) to highly significant (**).





NEbraskaCERT's Wisdom

Contact

AI is Risky Business

LINKEDIN – DKOHRELL

DAVID KOHRELL – DAVID@KOHRELL.ORG
[HTTPS://WWW.KOHRELL.ORG](https://www.kohrell.org)

