Cybersecurity considerations for AI systems

Arnaud Wiehe

Author of “The Book On Cybersecurity (2023)”

29 June 2007


UBER


iPhone


Airbnb


Spotify


coinbase


TikTok


Instagram


Touch to Shazam

"The future is not some place we are going to, but one we are creating"

John Schaar - American political theorist (1928-2011)

Think Like a Futurist

Plausible/Probable

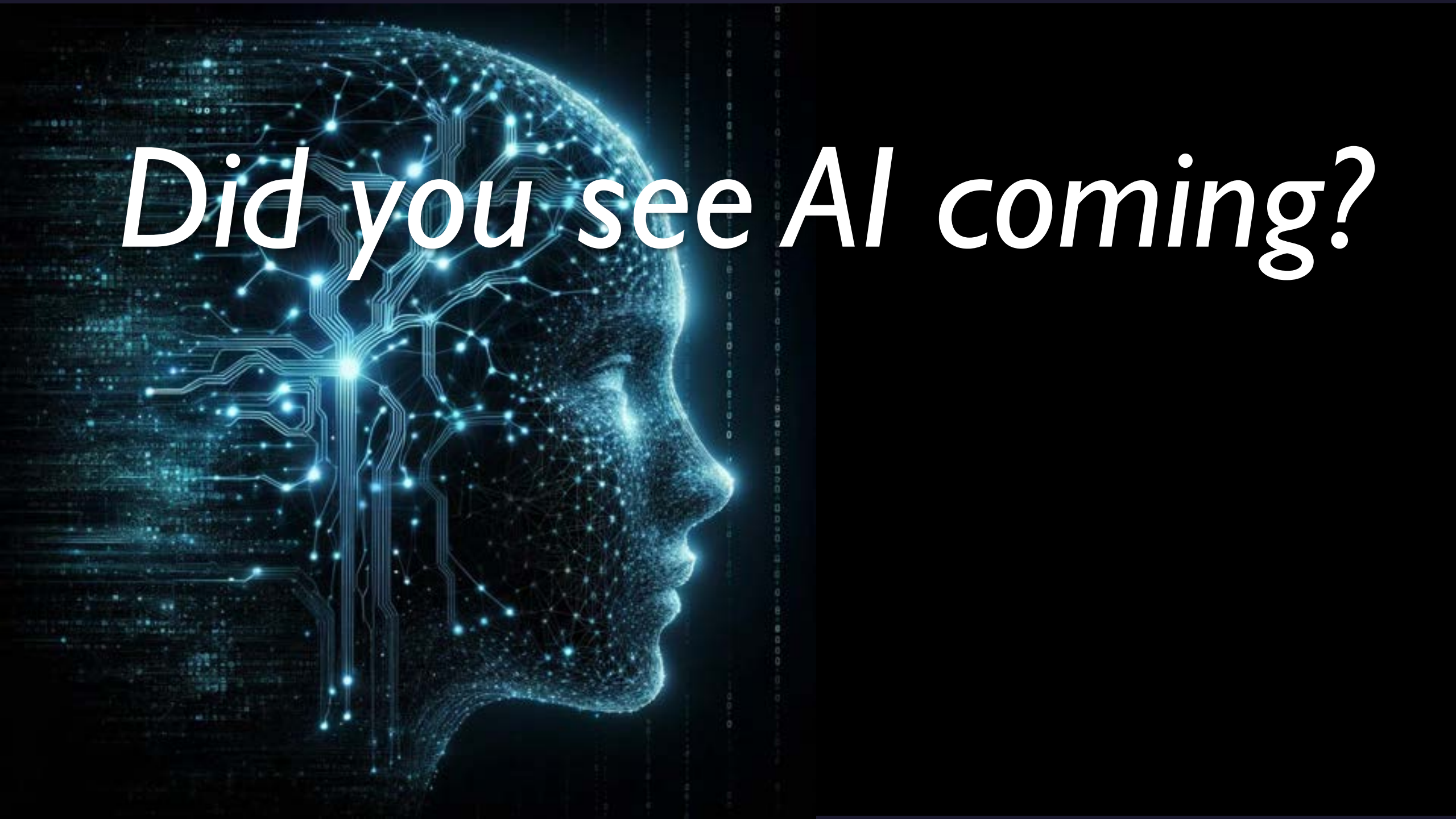
Likely to happen based on current trends. (The likely)

Possible

Everything that could happen, regardless of how likely it is. (The good, the bad and the ugly).

Preferable

What we would like to happen, encompassing our hopes and dreams for the future. (The ideal)


Did you see AI coming?

User generated image from Stable Diffusion Beta

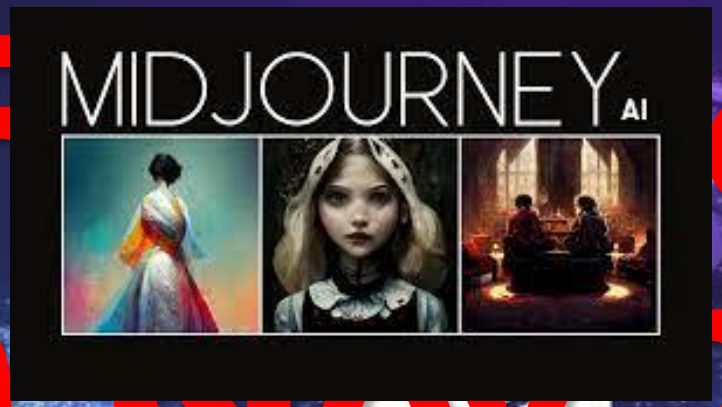
What Is Generative AI and Why Should You Be Using It?

Arnaud Wiehe
Keynote Speaker & Author of The Book On Cybersecurity. Managing Director Information Security

9 articles

November 3, 2022


Meta AI


MIDJOURNEY AI


ANTHROPIC
CLAUDE 2


Google
Bard AI


Stable Diffusion


Bing


DALL·E 3

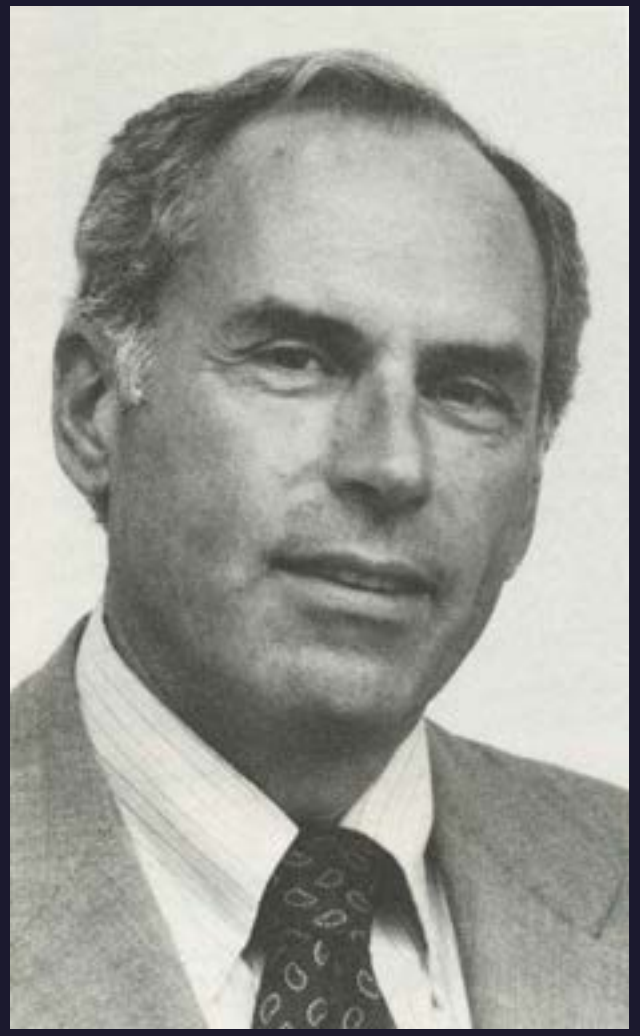

Talk to Pi,
your personal AI
Pi
pi.ai

"There are decades where nothing happens; and there are weeks where decades happen"

Vladimir Lenin

“Despite the potential dangers, it could be the greatest technology humanity has yet developed”

“What I lose the most sleep over is the hypothetical idea that we already have done something really bad by launching ChatGPT”



Roy Amara (1925 – 2007) American researcher, scientist, futurist and president of the Institute for the Future

Amara's law

We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run.

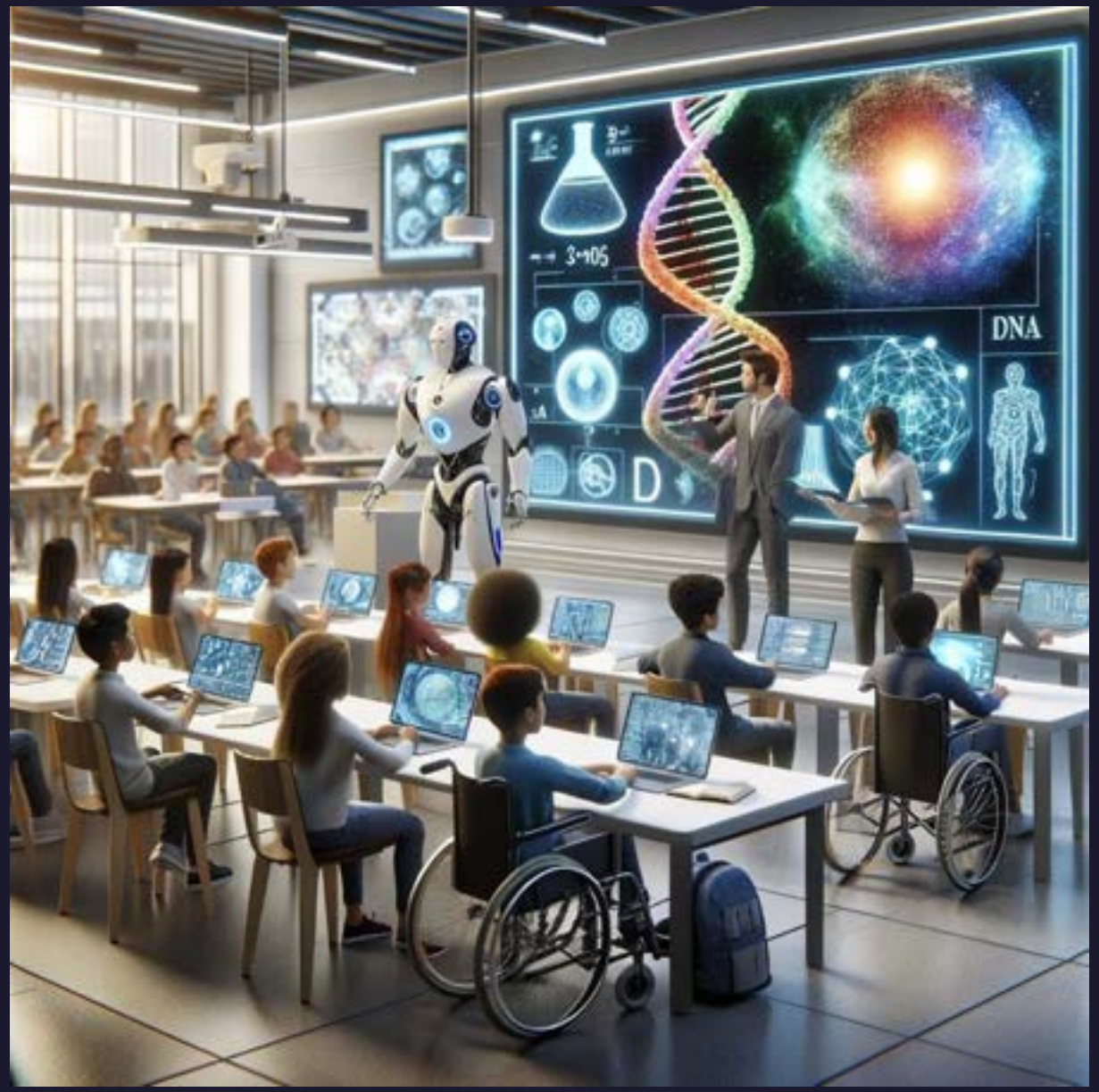
AI Benefits (The Angel)

Automation of tasks and processes

More reliable forecasting and prediction

Advancing the fields of medicine, cybersecurity, and autonomous vehicles

Improving education and learning outcomes



Smarter and faster decisions based on data analysis

Better customer service and user experience

Accurate medical diagnosis and treatment

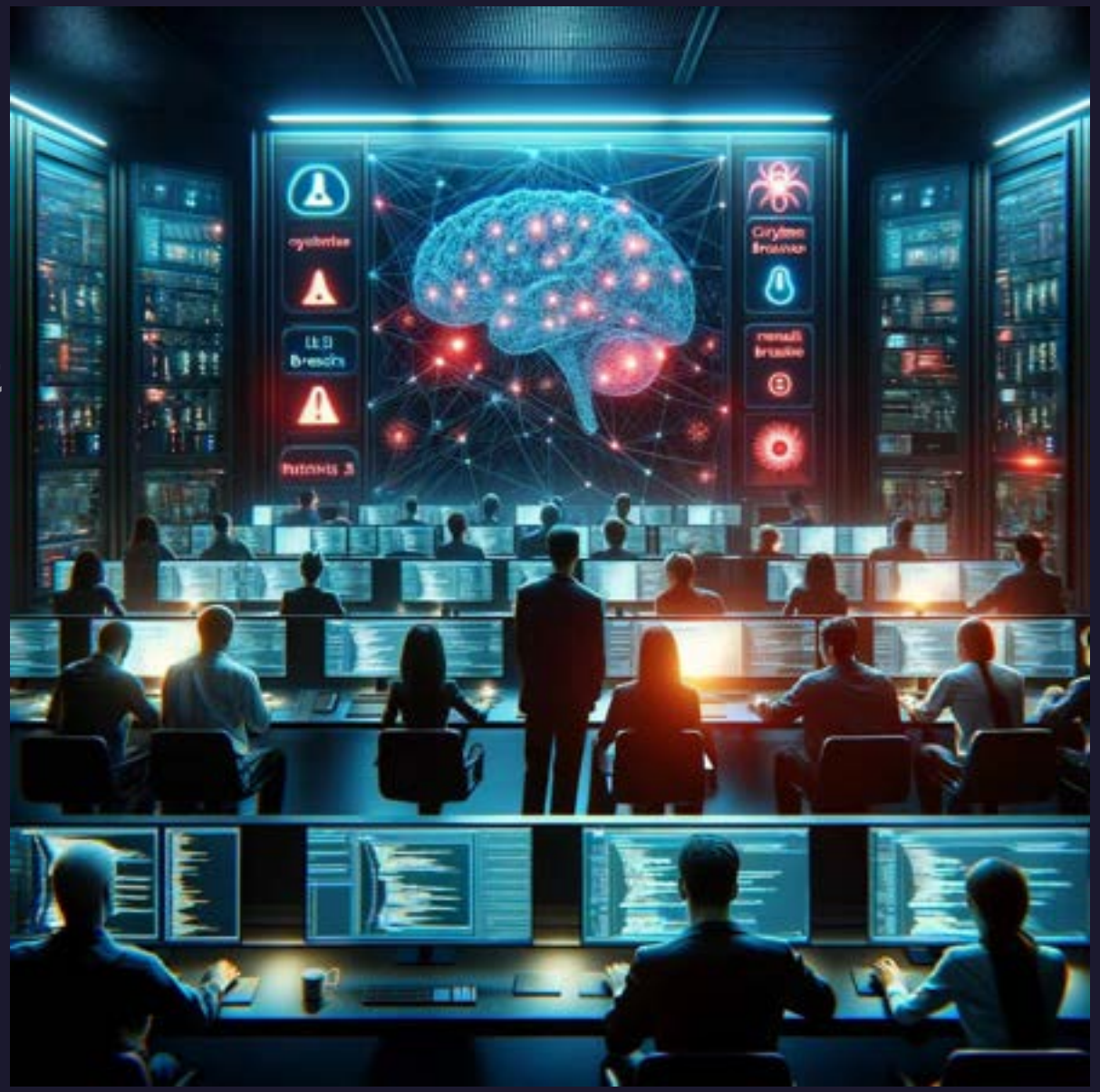
AI Risks (The Devil)

Data poisoning

Copyright infringement

Deepfakes

Job losses



AI-Powered phishing & malware

Model theft

Alignment problem

Privacy violations

Hallucinations

The Collingridge Dilemma (Double Bind Problem)

Information problem

Initially, impacts cannot be easily predicted until a technology is extensively developed and widely used.

Power problem

Later, control or change is difficult when the technology has become entrenched.

David Collingridge (Academic at University of Aston)
The Social Control of Technology (1980)

The ***Wright Flyer*** (also known as the ***Kitty Hawk***, ***Flyer* I** or the **1903 *Flyer***) – 17 December 1903

Karl Benz built his first car in 1885, called Patent Motorwagen – 1886

Microsoft

Microsoft Security

AI Security Risk Assessment

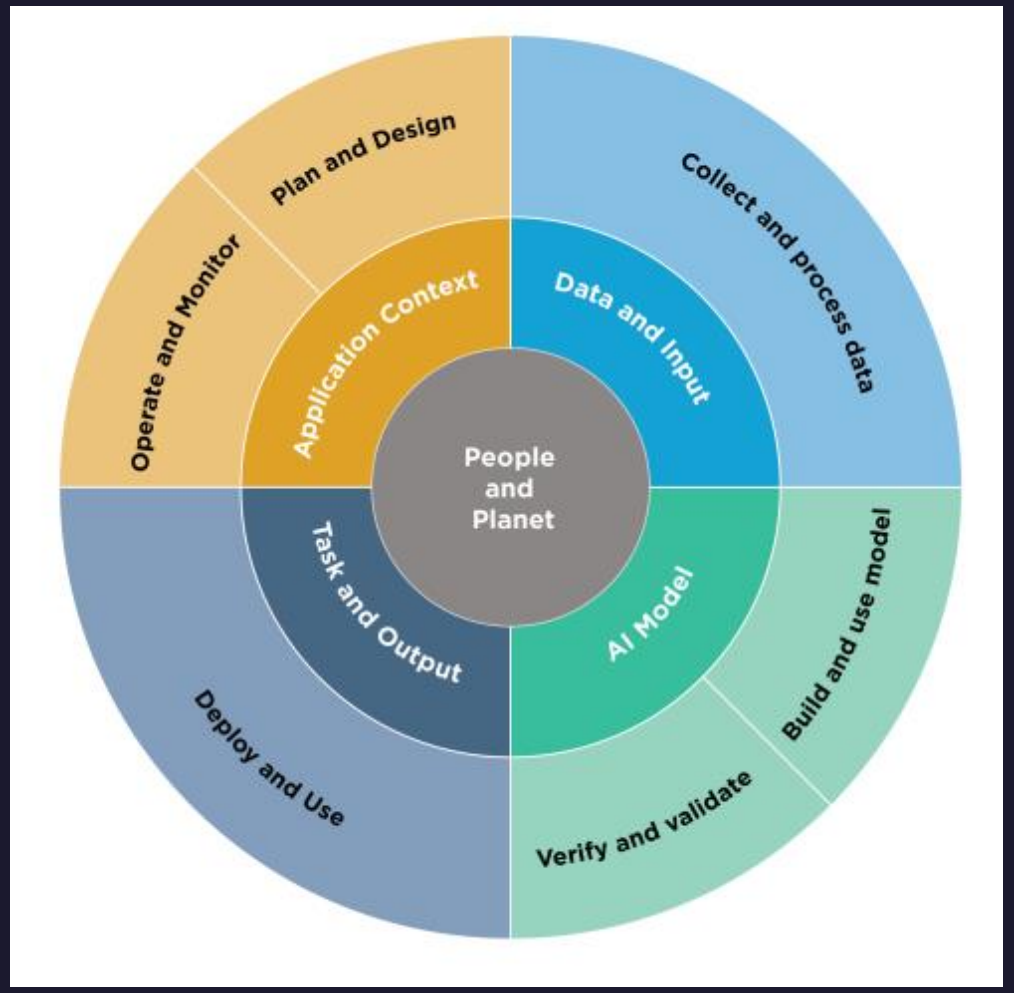
Best practices and guidance to secure AI systems

Administrative controls	
Machine learning security policies	Controls and policies relating to the documented policies that govern machine learning, artificial intelligence, and information security.

Technical controls	
Data collection	Controls and policies related to the collection, storage, and classification of data that are used for machine learning and artificial intelligence.
Data processing	Controls and policies relating to the processing and engineering of data used for machine learning and artificial intelligence.
Model training	Controls and policies relating to the design, training, and validation of models.
Model deployment	Controls and policies relating to the deployment of models and supporting infrastructure.
System monitoring	Controls and policies relating to the ongoing monitoring of machine learning systems.
Incident management	Controls and policies relating to how incidents related to AI system are handled.
Business continuity and disaster recovery	Controls and policies relating to loss of intellectual property through model stealing, degradation of service, or other AI specific vulnerabilities.

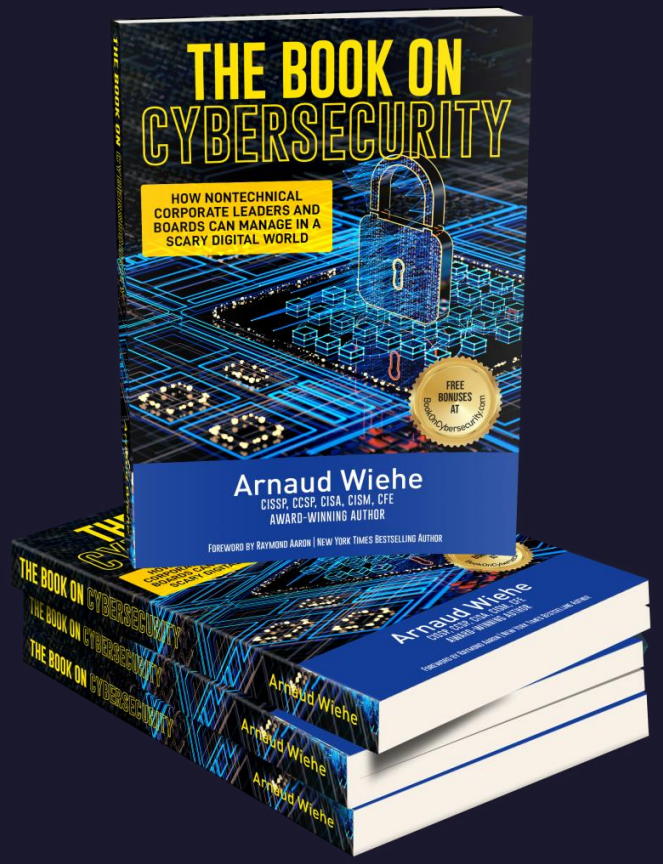
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Artificial Intelligence Risk Management Framework (AI RMF 1.0)



"I expect laws focusing on transparency of algorithms, algorithm bias, more prescription on when usage is legal or not, and guidelines to address potential abuse or misuse."

March 2023

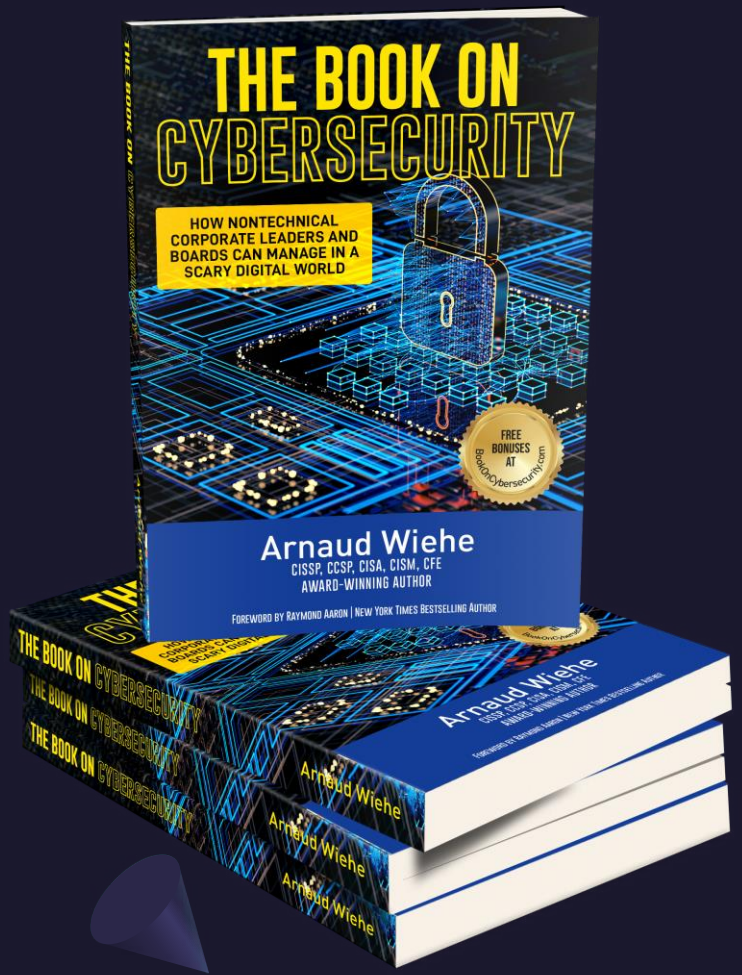


AI Regulations

- White House signed the executive order on AI (30 Oct 2023)
- UK held a two-day AI Safety Summit (1-2 Nov 2023)
- G7 release code of conduct on generative AI (2 Nov 2023)
- EU AI Act (likely to be concluded in Dec 2023)
- UN's AI Advisory Board (reporting back end 2023)

Is this the beginning of the end
or
the end of the beginning?

https://BookOnCyberSecurity.com



"Not all readers and leaders, but all leaders are readers"

Harry Truman

amazon.com

Arnaud Wiehe