RAYA INFORMATION TECHNOLOGY

Techforward Summit 2025

AI & Cybersecurity: Managing Opportunities Against Threats

Dr. Tim Nedyalkov
tim@cyberkeynote.com
www.linkedin.com/in/tim4ned/

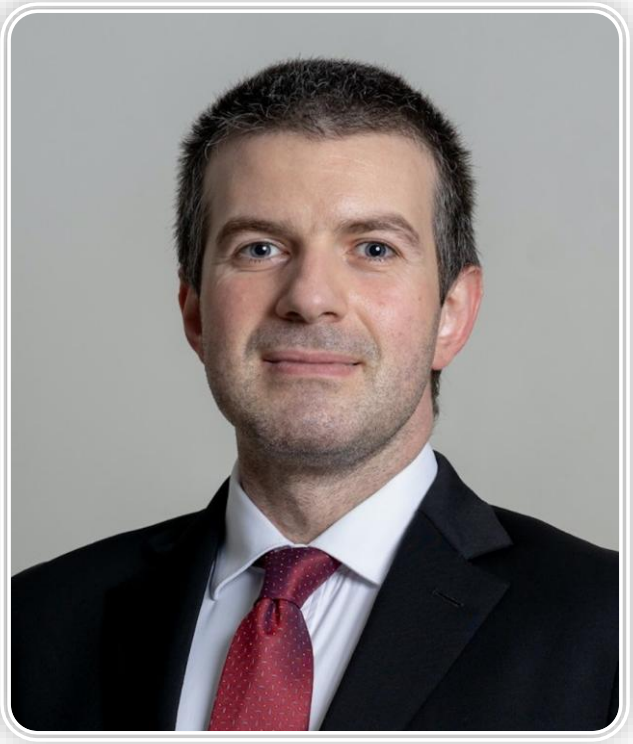
Four Seasons Sharm El Sheikh
From 15-17 May-2025

RAYA INFORMATION TECHNOLOGY

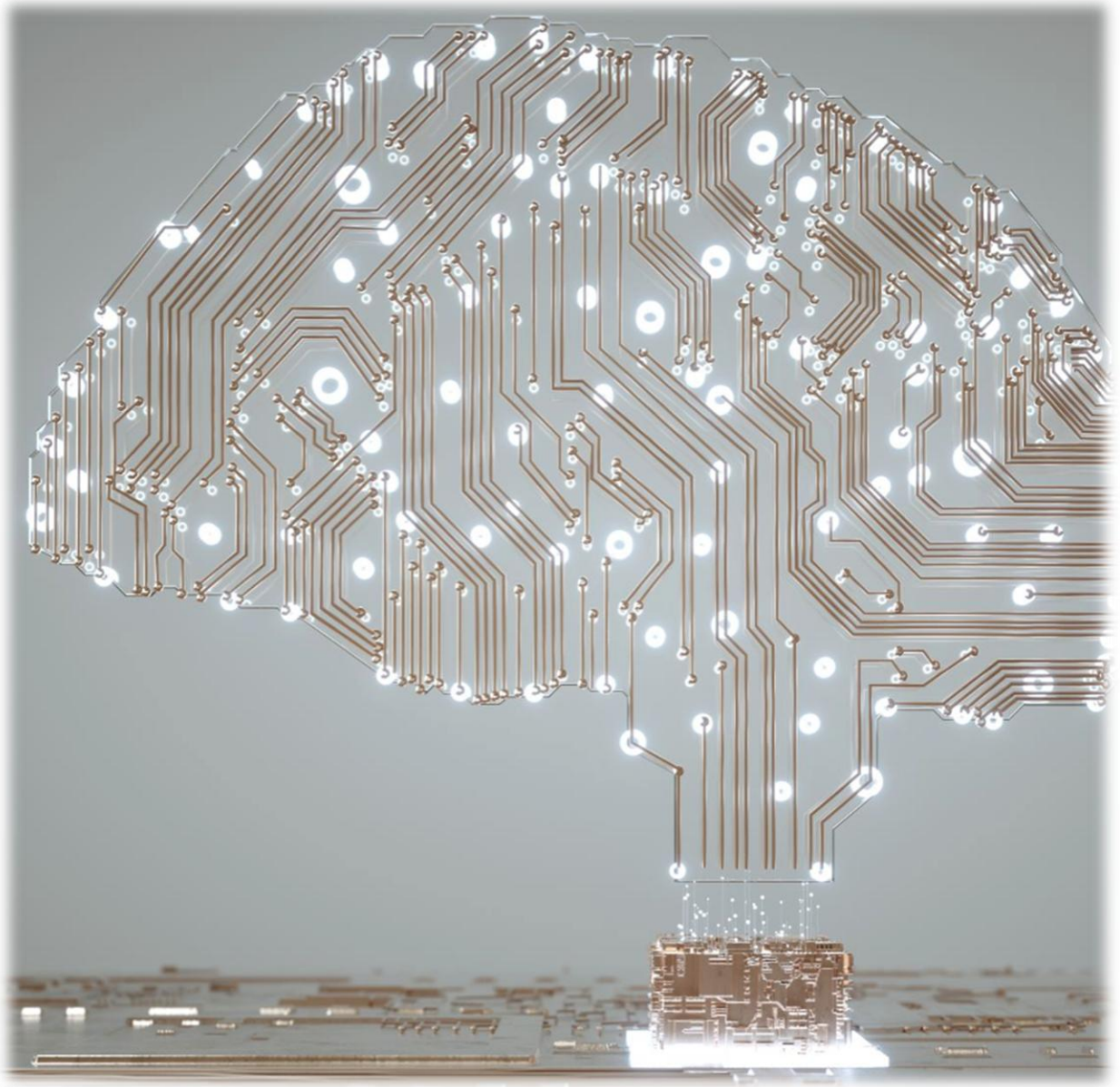
AI in the Boardroom

Speaker Introduction

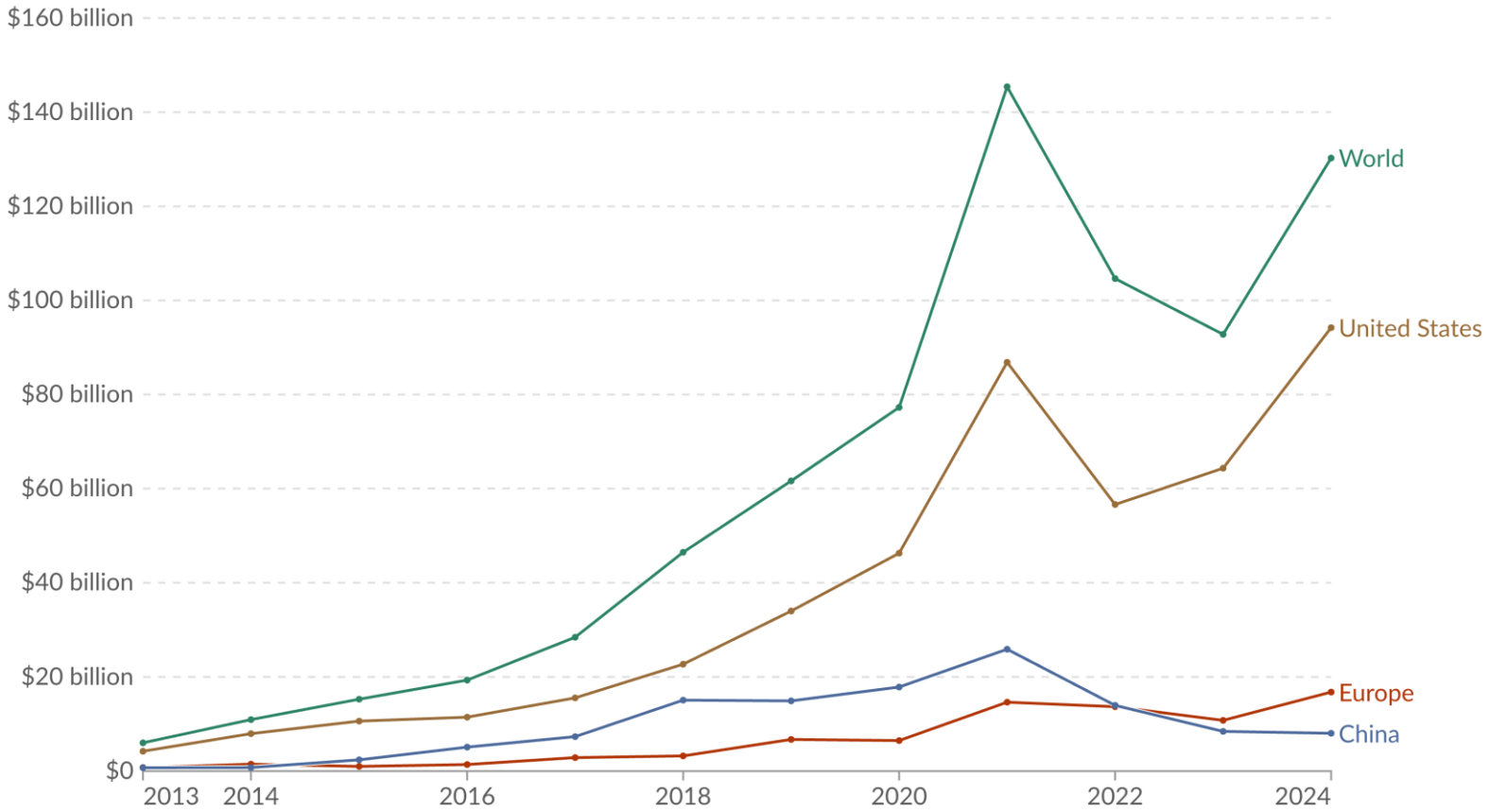
- 20+ years of multi-industry IT and Cybersecurity experience across Europe, the USA, Australia and the Middle East
- Technology Information Security Officer - Commonwealth Bank of Australia
- Cyber Security Manager – Riyadh Metro Network
- Information & Cyber Security Manager – Australian Broadcasting Corporation
- C|CISO, CISSP, CCSP, CISA, CISM, CRISC, CGEIT, ISO 27001 LA, CDPSE; Doctorate in Cyber Security/Information Assurance
- 2024 iTnews Benchmark Awards Security in Finance; 2022 Top Cyber News Magazine Global 40 under 40 in Cybersecurity; 2022 Onalytica Global Who's Who in Cybersecurity Influential Voices



The AI Transformation



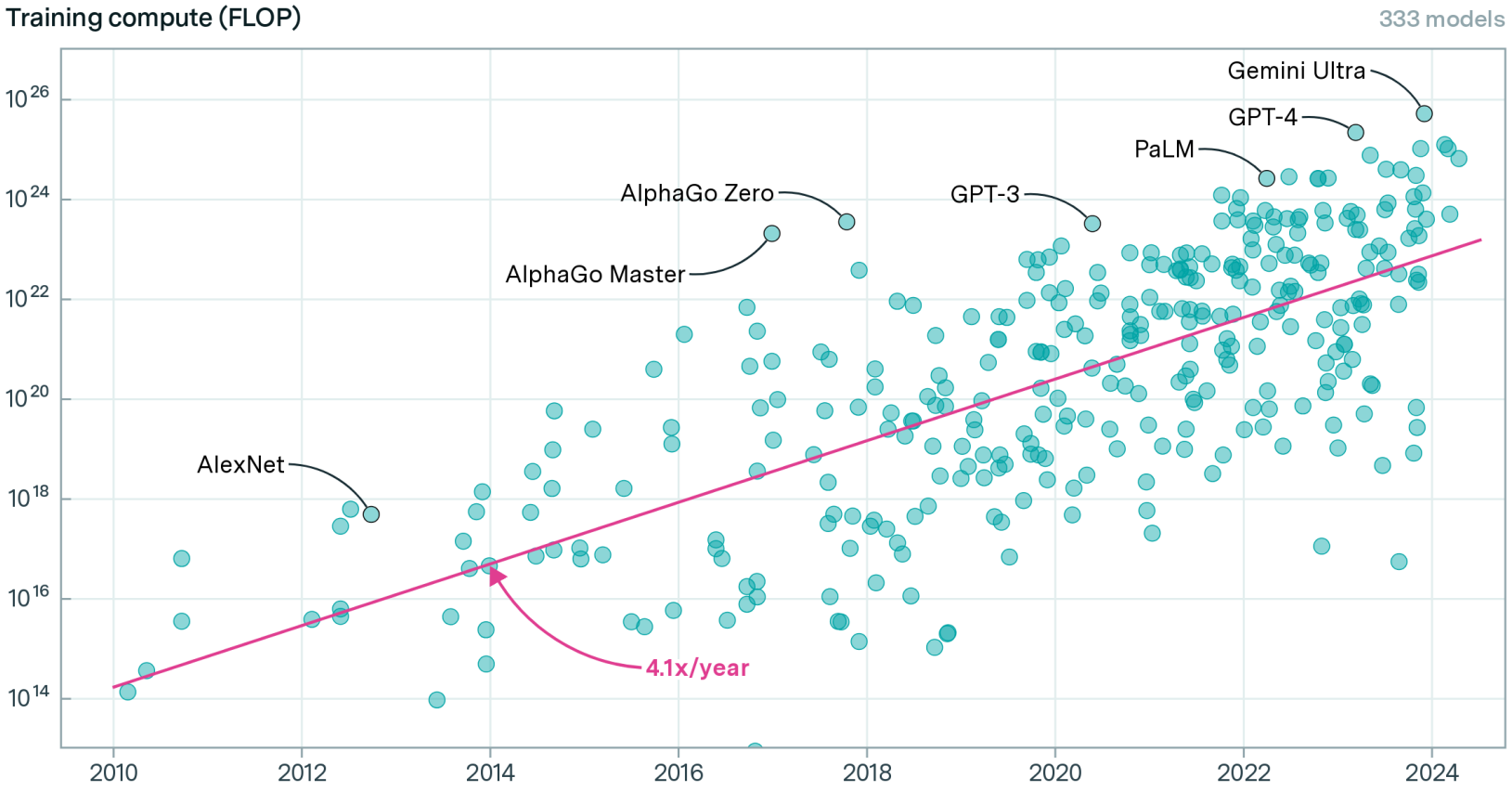
The AI Transformation



$130 billion invested in AI globally in 2024

Source: Stanford HAI, 2025 AI Index Report

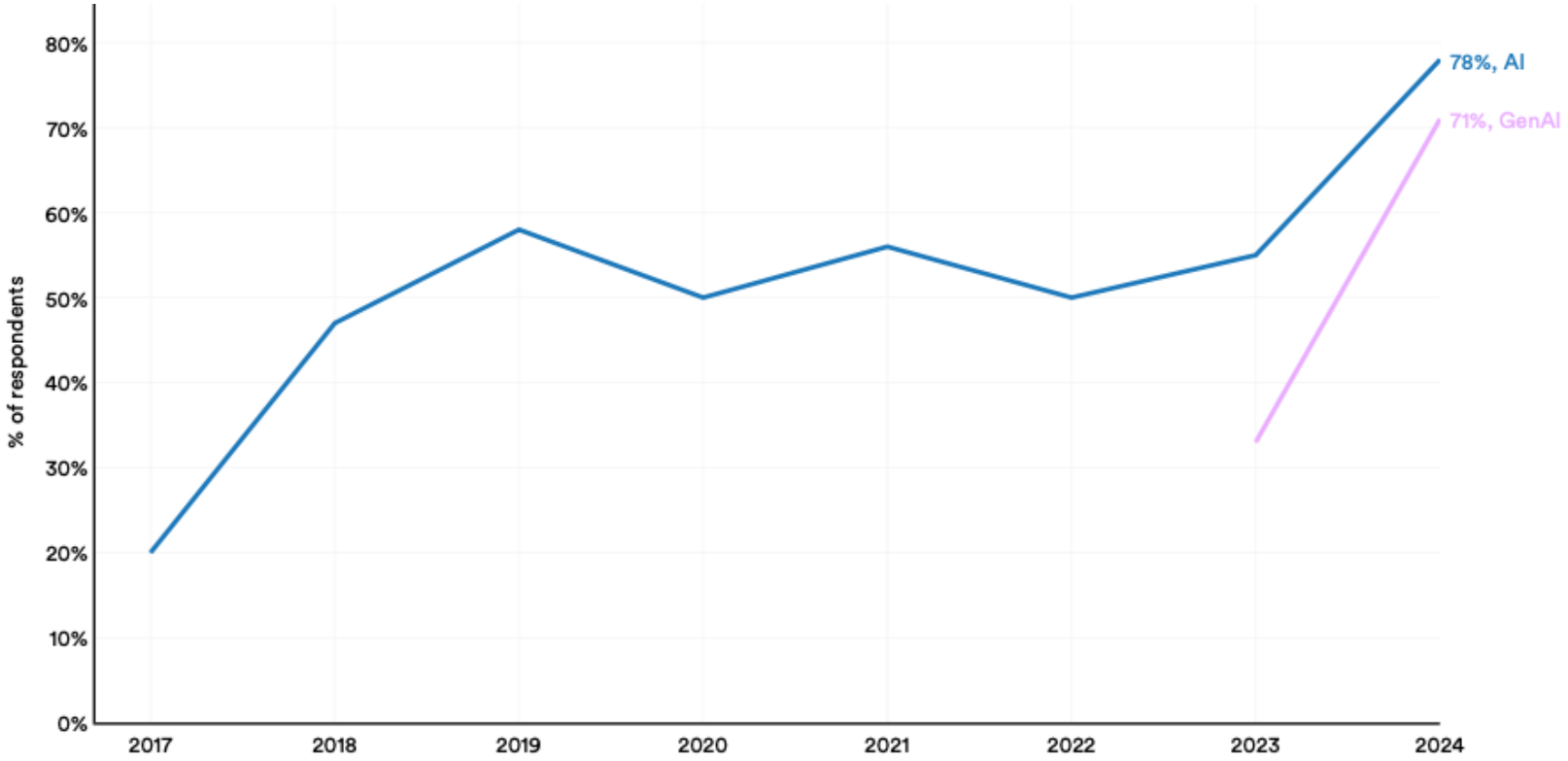
The AI Transformation



Training compute of AI models is **doubling roughly every 5 months!**

Source: Epoch AI

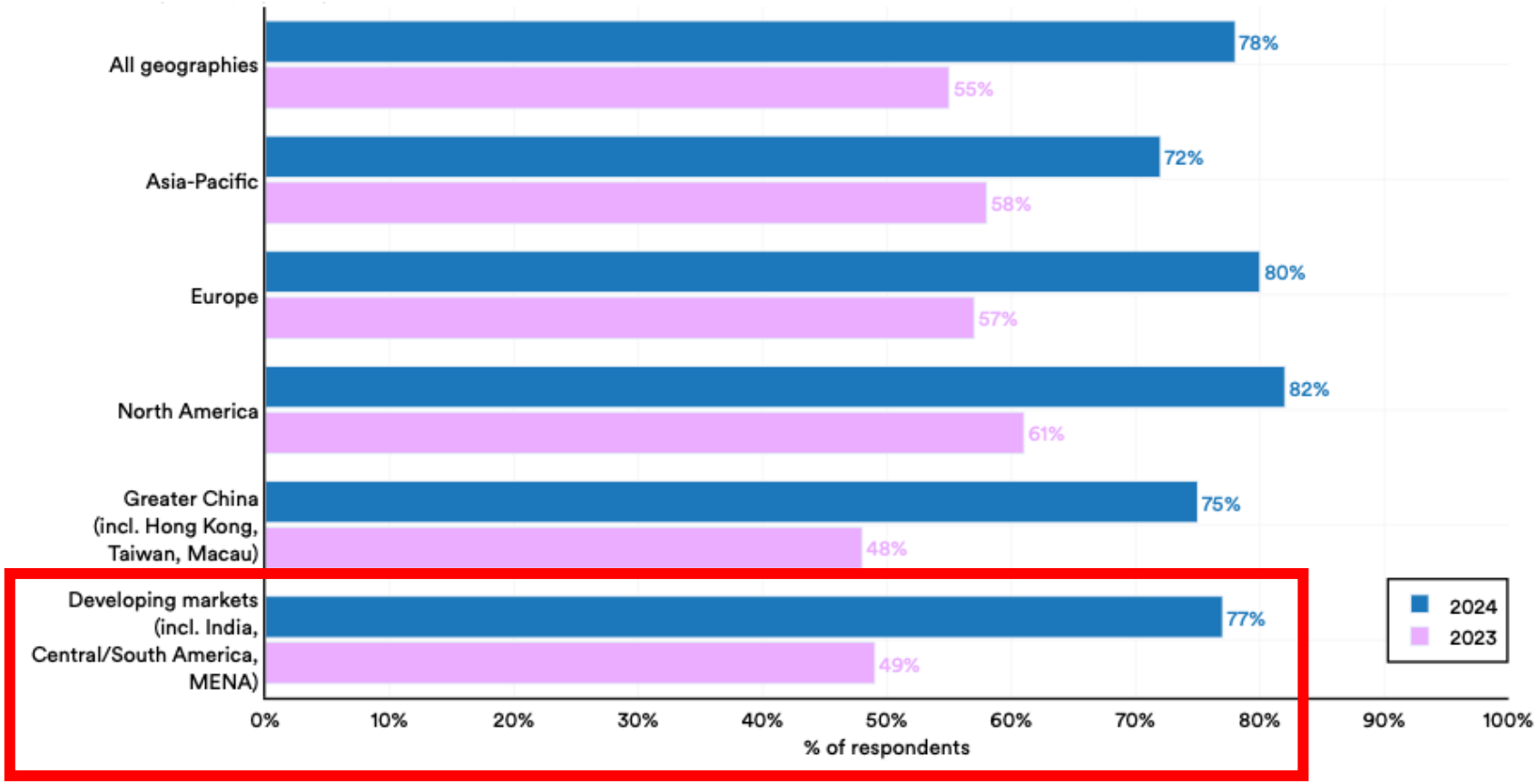
The AI Transformation



78% of organizations have begun to use AI in at least one business function

Source: Stanford HAI, 2025 AI Index Report

The AI Transformation



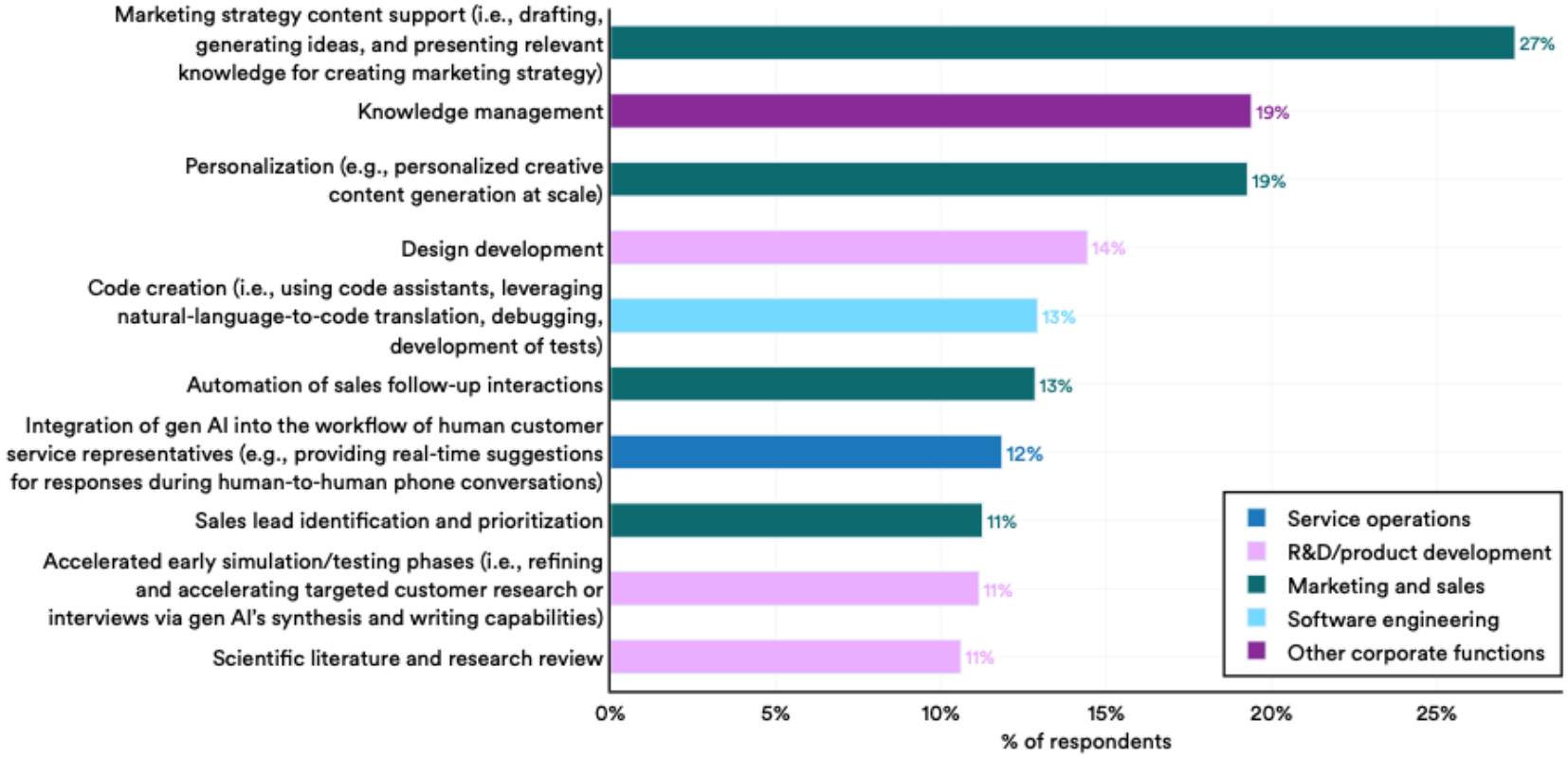
Use of AI: **+38% YoY** in developing markets such as MENA

Source: Stanford HAI, 2025 AI Index Report

9 MENA Countries with AI Strategy

Source: Oxford Insights, 2024

The AI Transformation



The most common applications:

27% Marketing content support

19% Knowledge management

19% Personalization of content generation

Source: Stanford HAI, 2025 AI Index Report

The AI Transformation

Wells Fargo

To help employees locate information they need to assist customers, Wells Fargo built an AI agent for **35,000** bankers across **4,000** branches.

75% of searches happen through the agent, cutting query response times from **10 minutes to just 30 seconds.**

Source: Wells Fargo

The AI Transformation

Dow

The global materials science company is deploying AI agents to uncover hidden losses and streamline shipping operations.

Once the AI system is fully scaled, Dow expects increased logistics rates and billing accuracy, which will **save millions in the first year.**

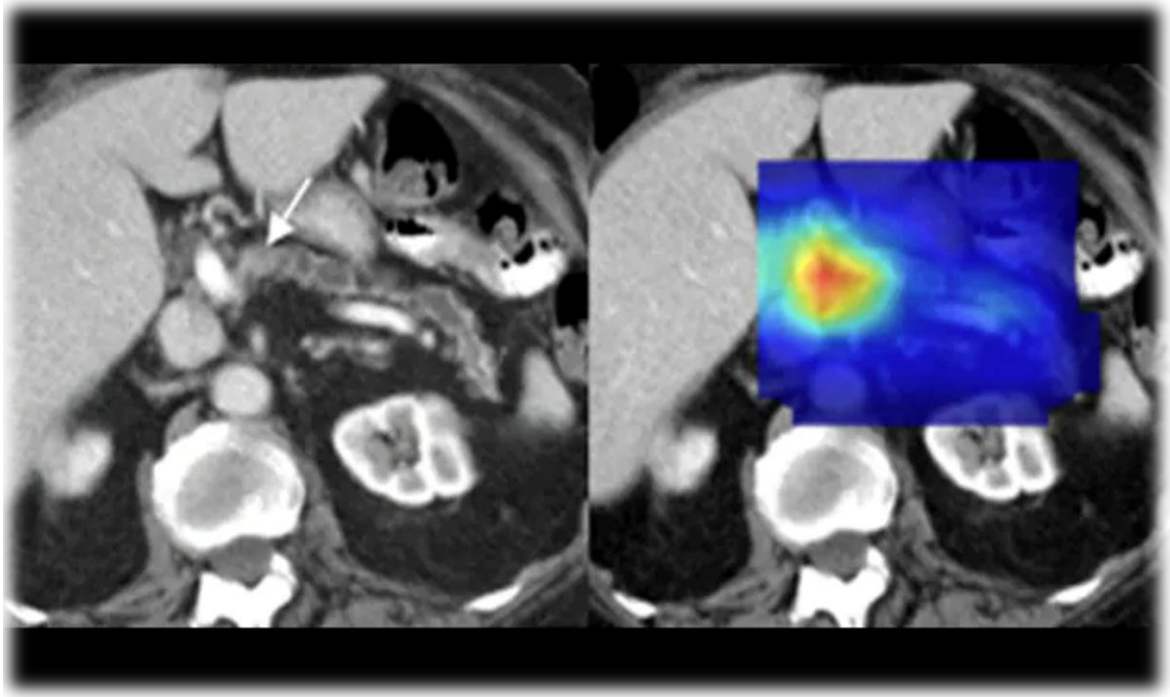
Source: Dow Inc

The AI Transformation

Mayo Clinic

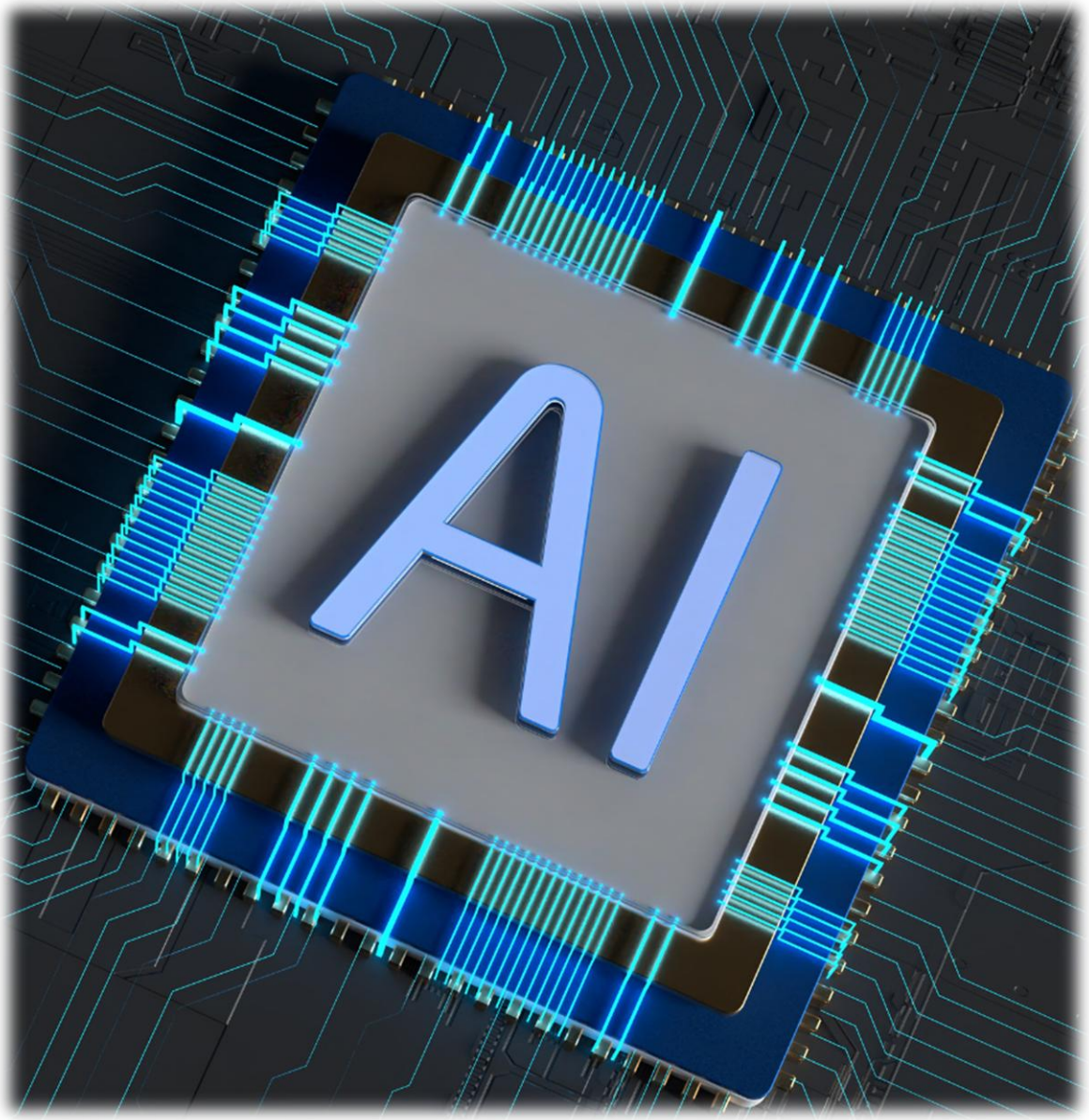
An AI system analyzes medical images and detects diseases at earlier stages. The deep learning algorithm for detecting diabetic retinopathy achieved 87% sensitivity and 91% specificity.

The AI system now screens over **30,000 patients** annually, identifying cases requiring intervention 28% earlier on average, with a projected **annual savings of $4.2 million** in treatment costs.



Source: Mayo Clinic

The Security Imperative of AI



The Security Imperative of AI

75% of senior IT leaders reported concerns that AI-based technologies pose a potential security risk.

73% are concerned about biased outcomes.

Source: HRB

The Security Imperative of AI

281

Fortune 500 companies cite AI as a potential risk factor

108

Fortune 500 companies specifically mention generative AI in annual financial reports

+250.2%

Increase in the number of mentions of AI in Fortune 500 companies' annual reports since 2022

+473.5%

Increase in Fortune 500 companies citing AI as a risk factor in annual reports since 2022

Source: Arize AI

The Security Imperative of AI

The Walt Disney Company

"Rules governing new technological developments, such as developments in generative artificial intelligence (AI), remain unsettled, and these developments may affect aspects of our existing business model, including revenue streams for the use of our IP and how we create our entertainment products."

Source: Walt Disney Company

The Security Imperative of AI

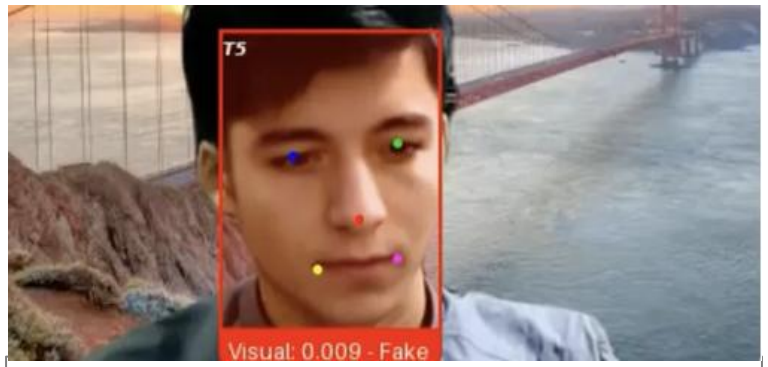
Vertex Pharmaceuticals

“Risks relating to inappropriate disclosure of sensitive information or inaccurate information appearing in the public domain may also apply from our employees engaging with and use of new artificial intelligence tools, such as ChatGPT."

Source: Vertex Pharmaceuticals Incorporated

The Security Imperative of AI

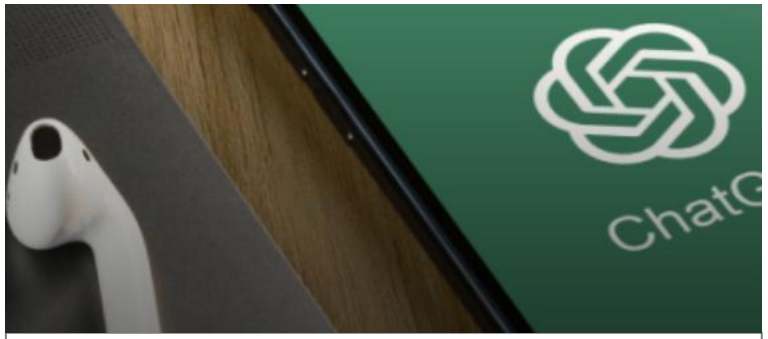
How verified accounts helped make fake images of a Pentagon explosion go viral
(ABC News)



Deepfake job applicant allegedly used AI tools to apply for remote role at US security startup
(Incidentdatabase.ai)

OpenAI's 4o model allegedly used to generate fake receipts and prescriptions
(Incidentdatabase.ai)

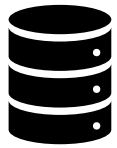
Microsoft blocks 1.6M bot signup attempts per hour amid global AI-driven fraud surge
(Microsoft)



Over 100,000 ChatGPT accounts stolen via info-stealing malware
(Infosec Magazine)

12,000+ live API keys in LLM training data reportedly poses security risks
(Incidentdatabase.ai)

1600+ AI Risks in 7 Domains



MIT AI Risk Database Domains

1. Discrimination & Toxicity
2. Privacy & security
3. Misinformation
4. Malicious actors & misuse
5. Human-computer interaction
6. Socioeconomic & environmental harms
7. AI system safety, failures and limitations

Source: MIT AI Risk Repository

New Cybersecurity Threats

Data Confidentiality & Privacy Protection

Unwanted & Inaccurate Outputs

Unwanted & Inaccurate Outputs

AI in Automotive Services

Data Confidentiality & Privacy Protection

AI in Healthcare

Data Confidentiality & Privacy Protection

AI in Healthcare

New Cybersecurity Threats

An attacker attempts to *indirectly* prompt LLMs integrated in applications

INJECTION METHOD

- Passive methods *(by retrieval)*
- Active methods *(e.g. emails)*
- User-driven injections
- Hidden injections

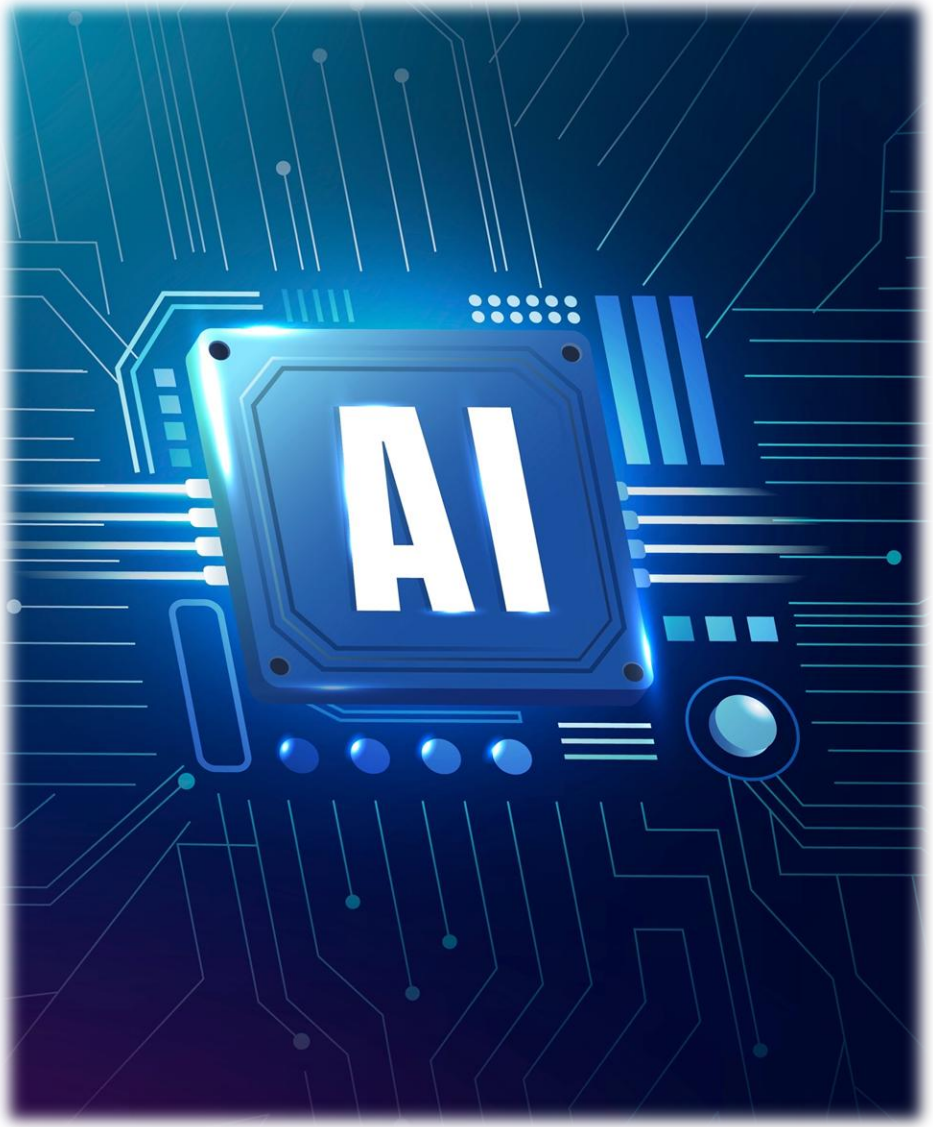
AFFECTED PARTIES

- End users
- Developers
- Automated systems
- LLMs

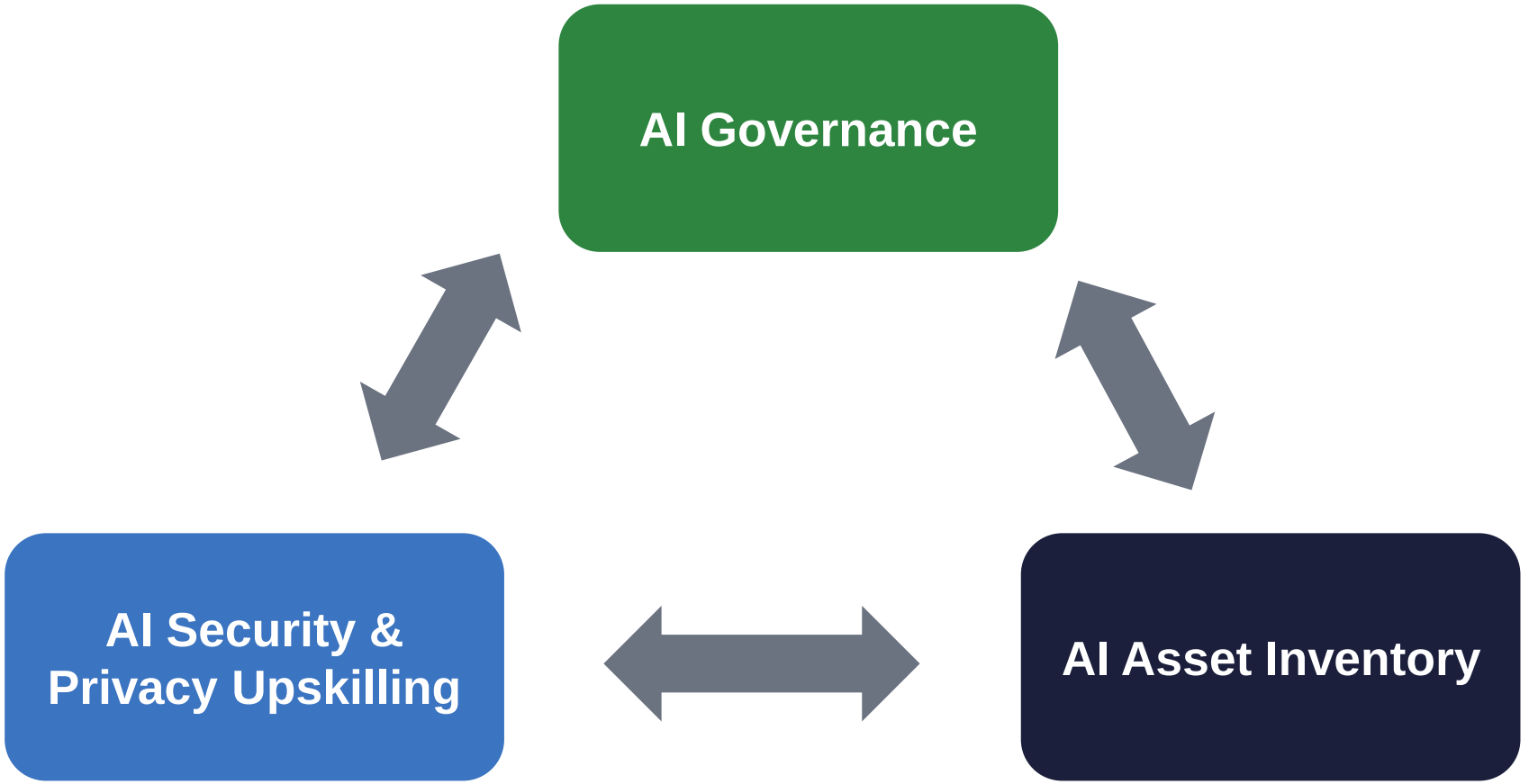
Existing Threats

Information Gathering	Fraud	Intrusion	Malware	Manipulated Content	Availability
- Personal data - Credentials - Chat leakage	- Phishing - Scams - Masquerading	- Persistence - Remote Control - API Calls	- Spreading Injections *(prompts as worms)* - Spreading Malware	- Wrong Summary - Disinformation - Propaganda/bias - Data hiding - Ads/Promotion	- DoS - Increased Computation

Value Protection in AI-Enabled Organizations



Value Protection in AI-Enabled Organizations



AI Governance Priorities

Define Acceptable AI Use Policies

Implement Data Classification & Access Management

Establish Systems to Approve User Applications & Attestations

Ongoing Governance, Monitoring & Compliance

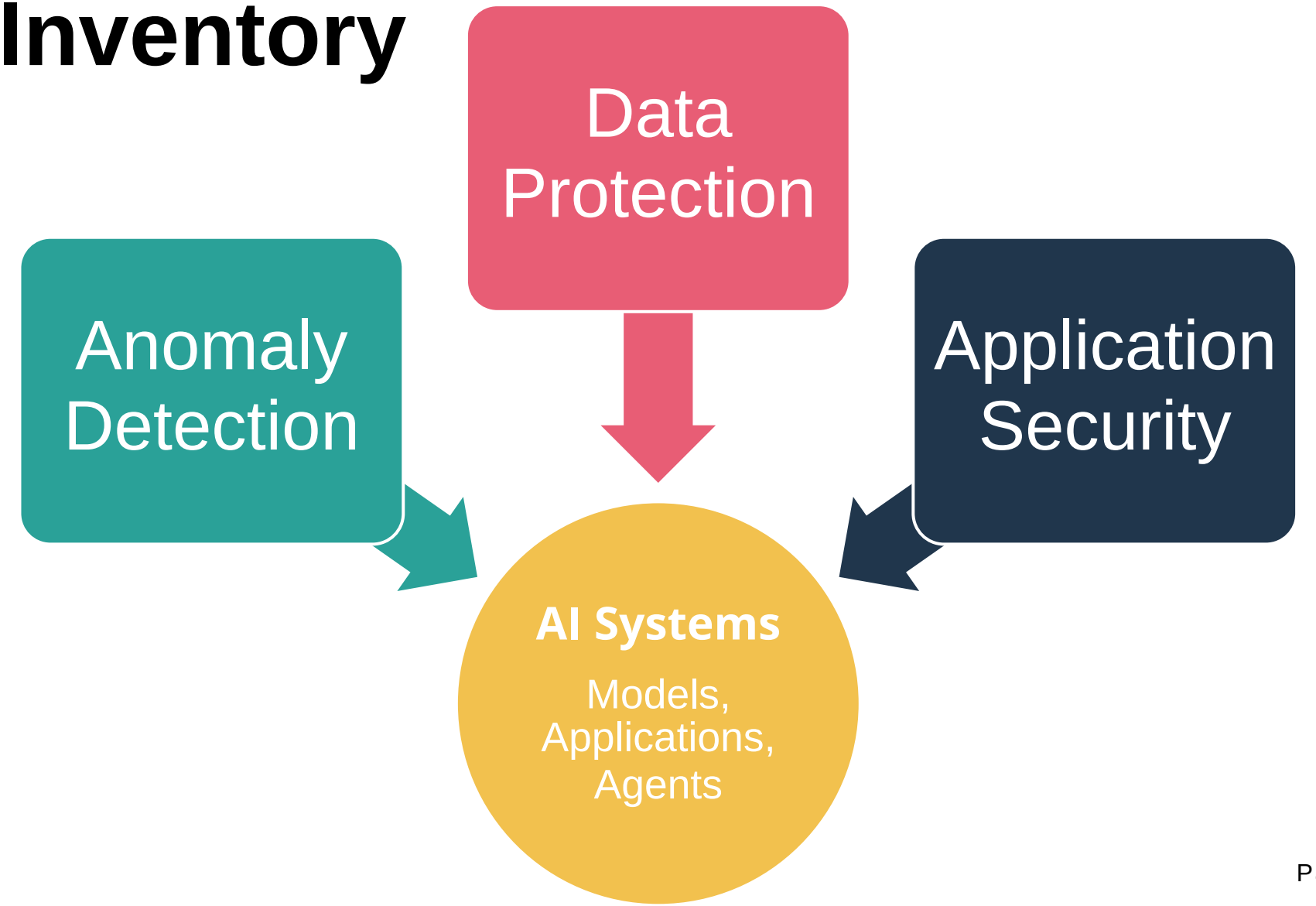
Organizational AI Governance

Privacy, Fairness, Bias Control

Measurement, Workflows, Policies

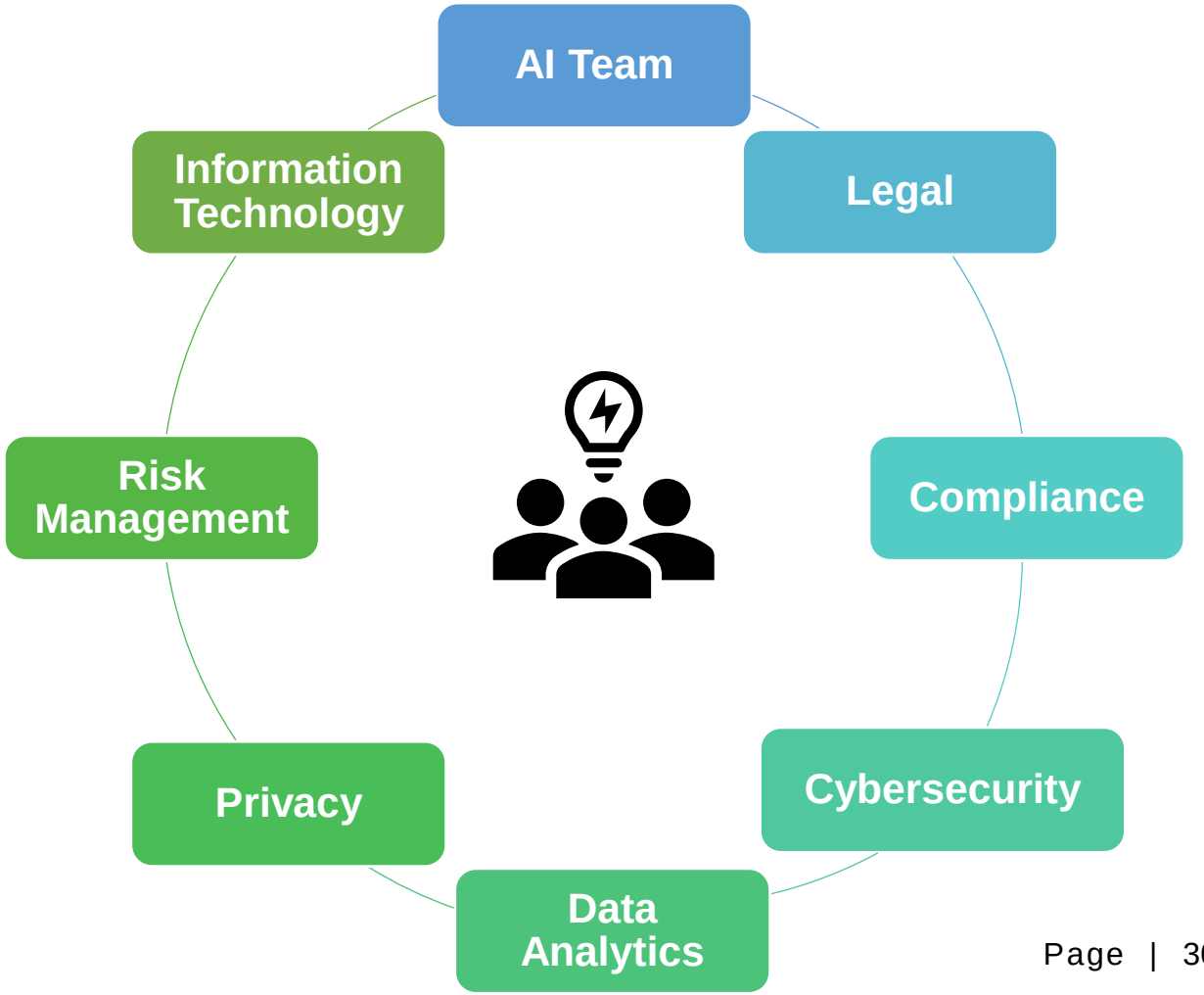
Transparency & Explainability

AI Asset Inventory



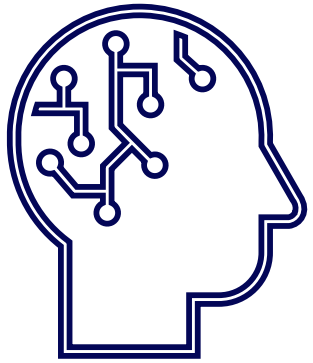
Working Groups for AI Risk, Privacy & Cybersecurity

- ✓ Better informed decisions
- ✓ Faster time to market
- ✓ Improved ownership
- ✓ Enhanced customer trust

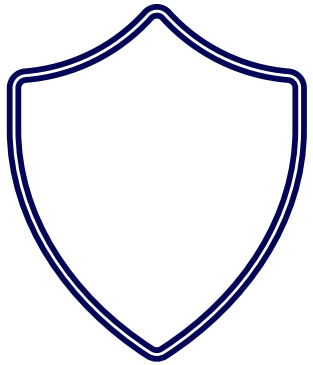


RAYA INFORMATION TECHNOLOGY

The Path Forward

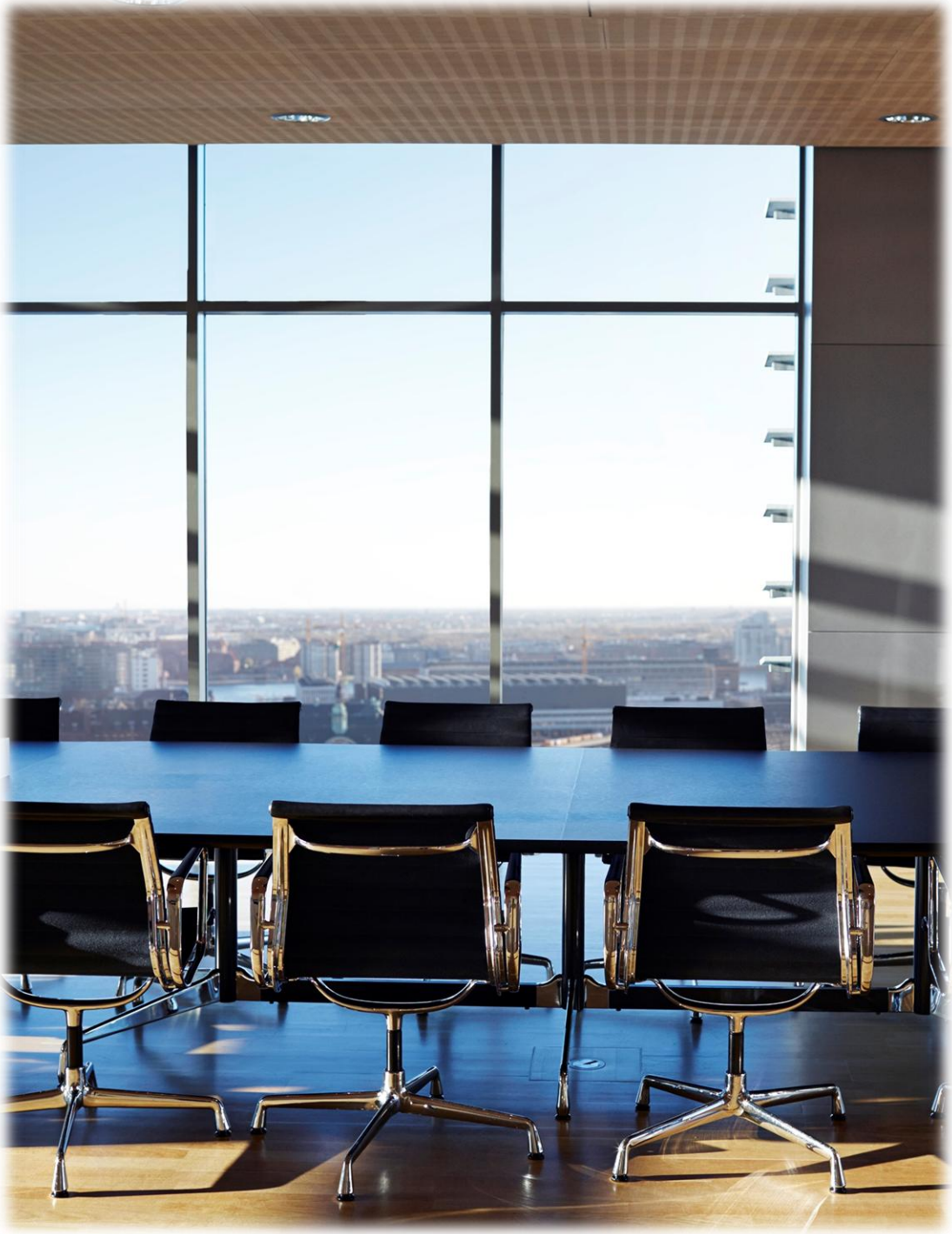


Understand AI Systems



Build-up Security

Evaluate Responses



Q & A

RAYA INFORMATION TECHNOLOGY

Techforward Summit 2025

AI & Cybersecurity: Managing Opportunities Against Threats

Dr. Tim Nedyalkov
tim@cyberkeynote.com
www.linkedin.com/in/tim4ned/

Four Seasons Sharm El Sheikh
From 15-17 May-2025