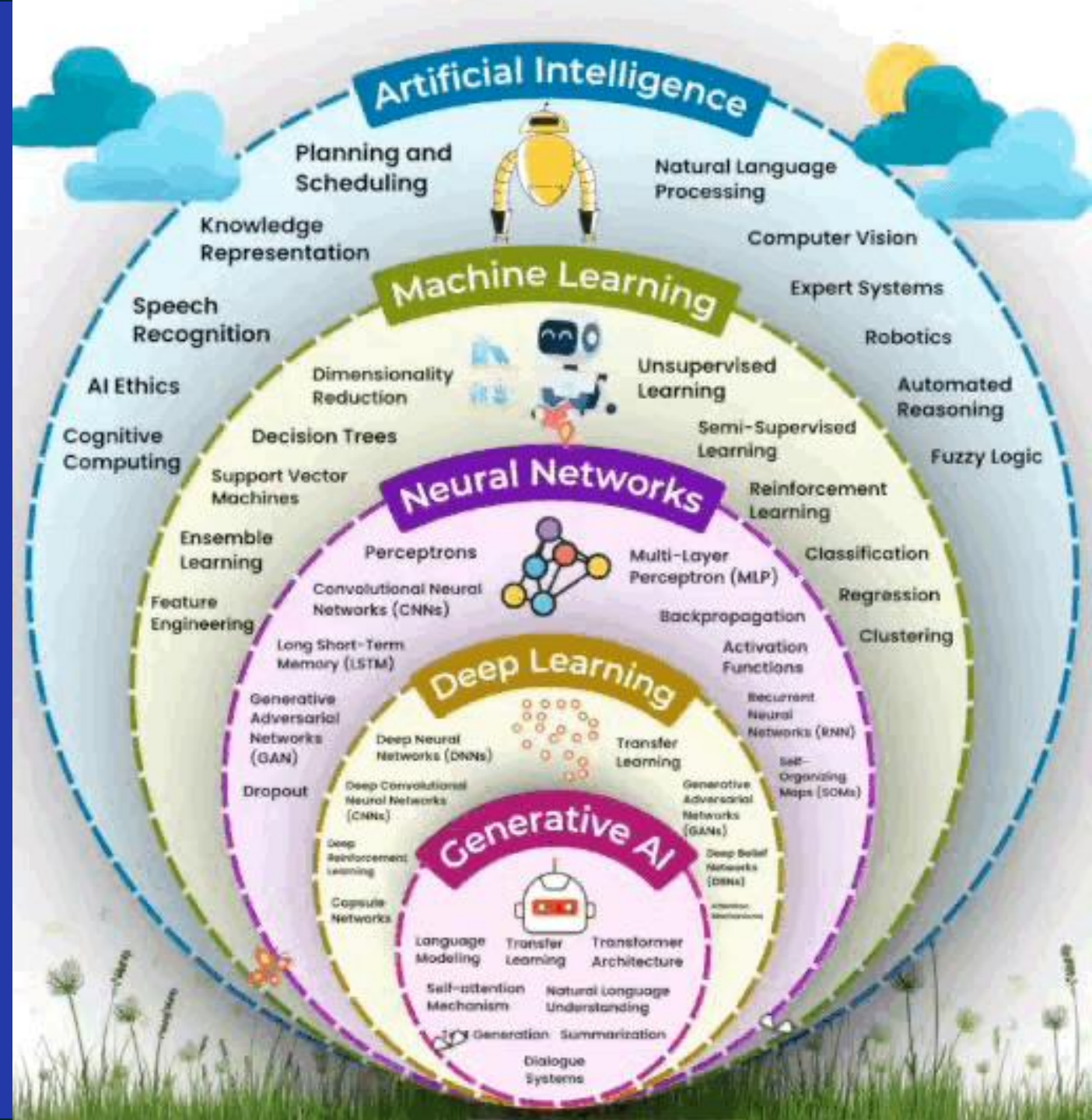# Future of Generative AI

Dr. Eng. Pujianto Yugopuspito, M.Sc.

The AI Universe

Objectives

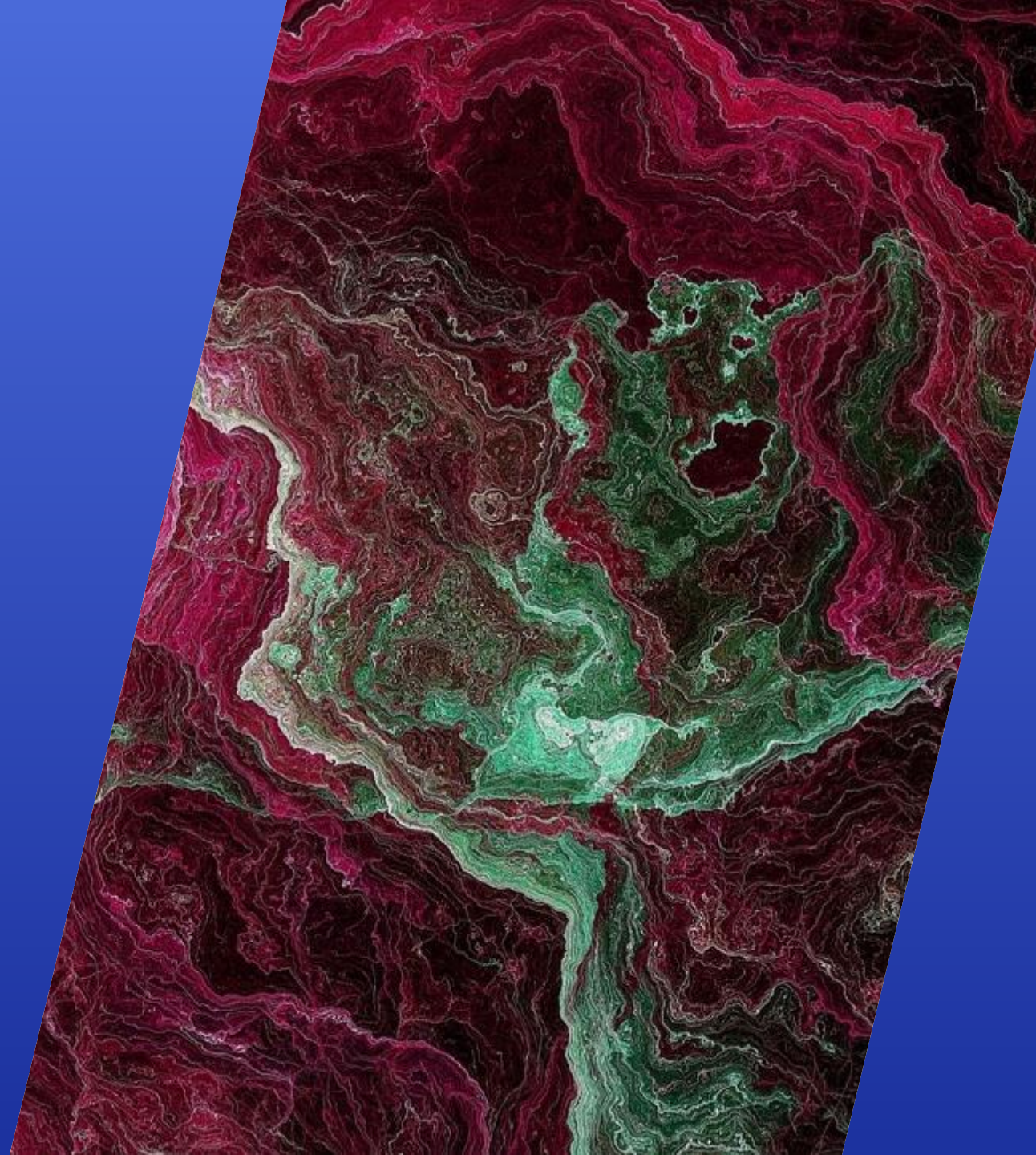**ETHICAL ASPECTS**

TECHNOLOGICAL PREDICTION

Future of Generative AI

Ethical Aspects

Dr. Eng. Pujianto Yugopuspito, M.Sc.

AI Risk Management Framework

President's Executive Order 14110:

Safe, Secure, and Trustworthy Artificial Intelligence.

Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile

Risks Unique to or Exacerbated by Generative AI

CBRN Information	Confabulation	Dangerous or Violent Recommendation	Data privacy	Environmental
Human-AI configuration	Information Integrity	Information Security	Intellectual Property	Obscene, Degrading, and/or Abusive Contents

Toxicity, Bias, and Homogenization	Value Chain and component integration
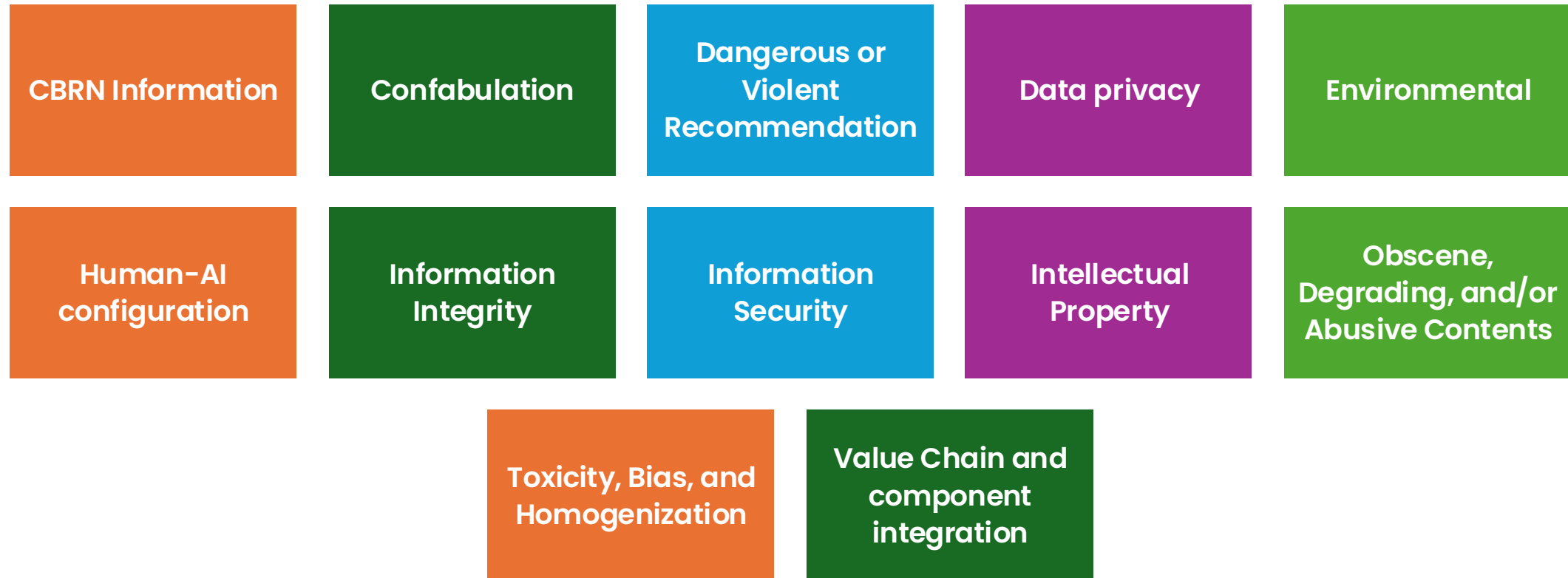
CBRN Information

Lowered barriers to entry or eased access to materially nefarious information related to chemical, biological, radiological, or nuclear (CBRN) weapons, or other dangerous biological materials.

Confabulation

The production of confidently stated but erroneous or false content (known colloquially as "hallucinations" or "fabrications").

Dangerous or Violent Recommendations

Eased production of and access to violent, inciting, radicalizing, or threatening content as well as recommendations to carry out self-harm or conduct criminal or otherwise illegal activities.

Data Privacy

Leakage and unauthorized disclosure or de-anonymization of biometric, health, location, personally identifiable, or other sensitive data.

Environmental

Impacts due to high resource utilization
in training GAI models, and related outcomes
that may result in damage to ecosystems.

Human-AI Configuration

Arrangement or interaction of humans and AI systems which can result in algorithmic aversion, automation bias or over-reliance, misalignment or mis-specification of goals and/or desired outcomes, deceptive or obfuscating behaviors by AI systems based on programming or anticipated human validation, anthropomorphization, or emotional entanglement between humans and GAI systems; or abuse, misuse, and unsafe repurposing by humans.

Information Integrity

Lowered barrier to entry to generate and support the exchange and consumption of content which may not be vetted, may not distinguish fact from opinion or acknowledge uncertainties, or could be leveraged for large-scale dis- and mis-information campaigns.

Information Security

Lowered barriers for offensive cyber capabilities, including ease of security attacks, hacking, malware, phishing, and offensive cyber operations through accelerated automated discovery and exploitation of vulnerabilities; increased available attack surface for targeted cyber attacks, which may compromise the confidentiality and integrity of model weights, code, training data, and outputs.

Intellectual Property

Eased production of alleged copyrighted, trademarked, or licensed content used without authorization and/or in an infringing manner; eased exposure to trade secrets; or plagiarism or replication with related economic or ethical impacts.

Obscene, Degrading, and/or Abusive Content

Eased production of and access to obscene, degrading, and/or abusive imagery, including synthetic Child Sexual Abuse Material (CSAM), and Nonconsensual Intimate Images (NCII) of adults.

Toxicity, Bias, and Homogenization

Difficulty controlling public exposure to toxic or hate speech, disparaging or stereotyping content; reduced performance for certain sub-groups or languages other than English due to non-representative inputs; undesired homogeneity in data inputs and outputs resulting in degraded quality of outputs.

Value Chain and Component Integration

Non-transparent or untraceable integration of upstream third-party components, including data that has been improperly obtained or not cleaned due to increased automation from GAI; improper supplier vetting across the AI lifecycle; or other issues that diminish transparency or accountability for downstream users.

Ethical Framework of Generative Artificial Intelligence

Solaki

Timeline of global AI regulation

Canada:
Bill C27, AI and Data Act
↓
Fine: $25M or 5% of company's gross global revenue

European Union:
GDPR
↓
Fine: €20M or 4% of company's annual turnover

Australia:
AI Ethics Framework

United States:
National Artificial Intelligence Initiative

United Kingdom:
AI Regulation Policy

New York City:
AI Hiring Law

2018 — **2019** — **2020** — **2021** — **2022** — **2023** — **2024**

Germany:
AI Strategy

South Korea:
AI Strategy

India:
National Strategy for AI

Colombia:
National Policy for Digital Transformation and AI

Norway:
National AI Strategy

Serbia:
Strategy for the development of AI

United States:
NIST issues an AI risk management framework

European Union:
The AI Act
↓
Fine: €30M or 6% of company's global revenue

China:
Internet Information Service Algorithmic Management (IISARM) regulations

United States:
AI Bill of Rights validation for algorithms to be explainable and protect against discrimination

Singapore:
Launches AI Verify— a testing framework and toolkit

European Union:
EU AI Act endorsed and begins rolling out in 2024

IBM Principles of Responsbile AI

01

The purpose of AI is to augment human intelligence

02

Data and insight belong to their creator

03

AI systems must be transparent and explainable

Technology Prediction

Future of Generative AI

Dr. Eng. Pujianto Yugopuspito, M.Sc.

Pros

AI triumphs...

November 2022
OpenAI publicly debuts ChatGPT-3.5, a breakthrough generative AI tool that can generate long-form human-like responses based on text inputs. ChatGPT surpasses one million users in five days.

March 2023
OpenAI releases ChatGPT-4.0, which includes significant improvements in reliability, creativity, and problem-solving.

September 2023
OpenAI announces DALL-E 3, an advanced version of its AI image generator, capable of creating more detailed images.

December 2023
Google introduces a new AI model, Gemini, which will power a broad range of products and services.

May 2024
OpenAI introduces GPT-4o, which is capable of more complex real-time interaction across text, audio, image, and video inputs.

February 2023
Google rushes to announce its generative AI tool, Bard, a day before Microsoft announces its own AI chatbot, Bing Chat.

March 2023
Salesforce debuts Einstein GPT, a new AI model for CRM clients.

October 2023
A study published in *cancers* journal suggests that AI can be used to assist in the early diagnosis of lung cancer.

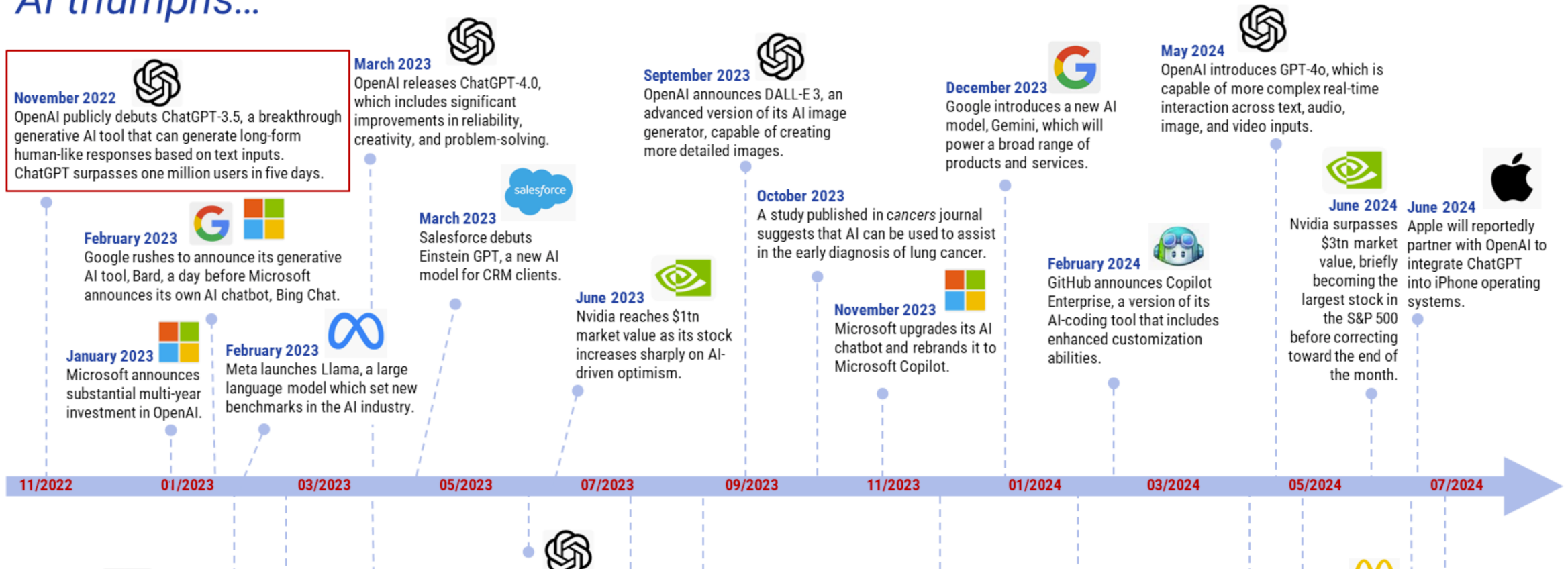**February 2024**
GitHub announces Copilot Enterprise, a version of its AI-coding tool that includes enhanced customization abilities.

June 2024
Nvidia surpasses $3tn market value, briefly becoming the largest stock in the S&P 500 before correcting toward the end of the month.

June 2024
Apple will reportedly partner with OpenAI to integrate ChatGPT into iPhone operating systems.

January 2023
Microsoft announces substantial multi-year investment in OpenAI.

February 2023
Meta launches Llama, a large language model which set new benchmarks in the AI industry.

June 2023
Nvidia reaches $1tn market value as its stock increases sharply on AI-driven optimism.

November 2023
Microsoft upgrades its AI chatbot and rebrands it to Microsoft Copilot.

11/2022   01/2023   03/2023   05/2023   07/2023   09/2023   11/2023   01/2024   03/2024   05/2024   07/2024

Doubts

11/2022 — 01/2023 — 03/2023 — 05/2023 — 07/2023 — 09/2023 — 11/2023 — 01/2024 — 03/2024 — 05/2024 — 07/2024

February 2023
Google's AI chatbot, Bard, provides an incorrect response to a question in its first public demo. Google-parent Alphabet's shares tumble in response.

February 2023
Google's senior vice president, Prabhakar Raghavan, warns that ChatGPT could provide "completely fictitious" answers to questions.

March-April 2023
ChatGPT is briefly banned in Italy due to concerns about data privacy.

June 2023
OpenAI is sued after ChatGPT hallucinates embezzlement claims.

July 2023
FTC launches an investigation into ChatGPT to determine if it has caused harm by providing users with incorrect information.

August 2023
Concerns about OpenAI's financial stability arise after a report shows that ChatGPT costs the company $700k per day to run.

December 2023
Microsoft's AI image generator receives criticism for generating violent images, including of President Biden.

February 2024
Google's chatbot Gemini is criticized after it generates racially inaccurate and insensitive historical images.

April 2024
Meta's stock falls sharply after the company announces a multi-billion-dollar AI investment plan.

May 2024
Salesforce reduces its revenue outlook in its earnings report, sparking concerns that AI isn't providing as much of a revenue tailwind as expected.

June 2024
McDonald's removes AI-assisted ordering from its drive-thrus after numerous order errors.

June 2024
Nvidia's stock slides and the company loses its status as the largest stock in the S&P 500.
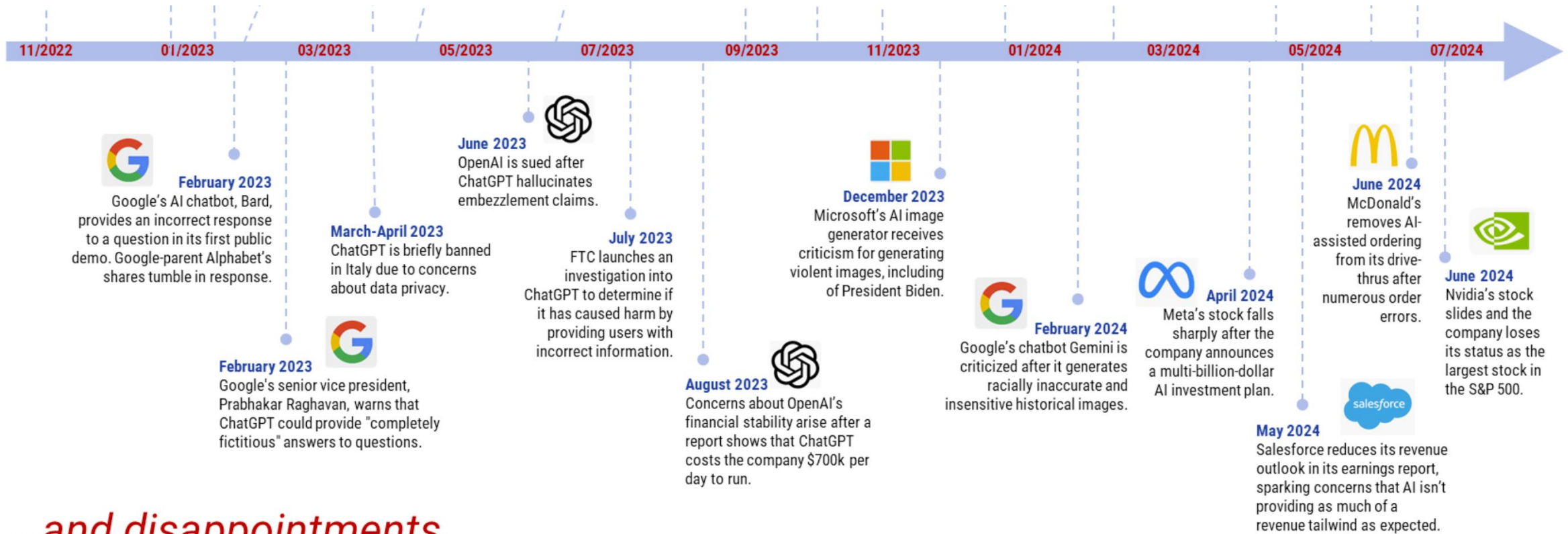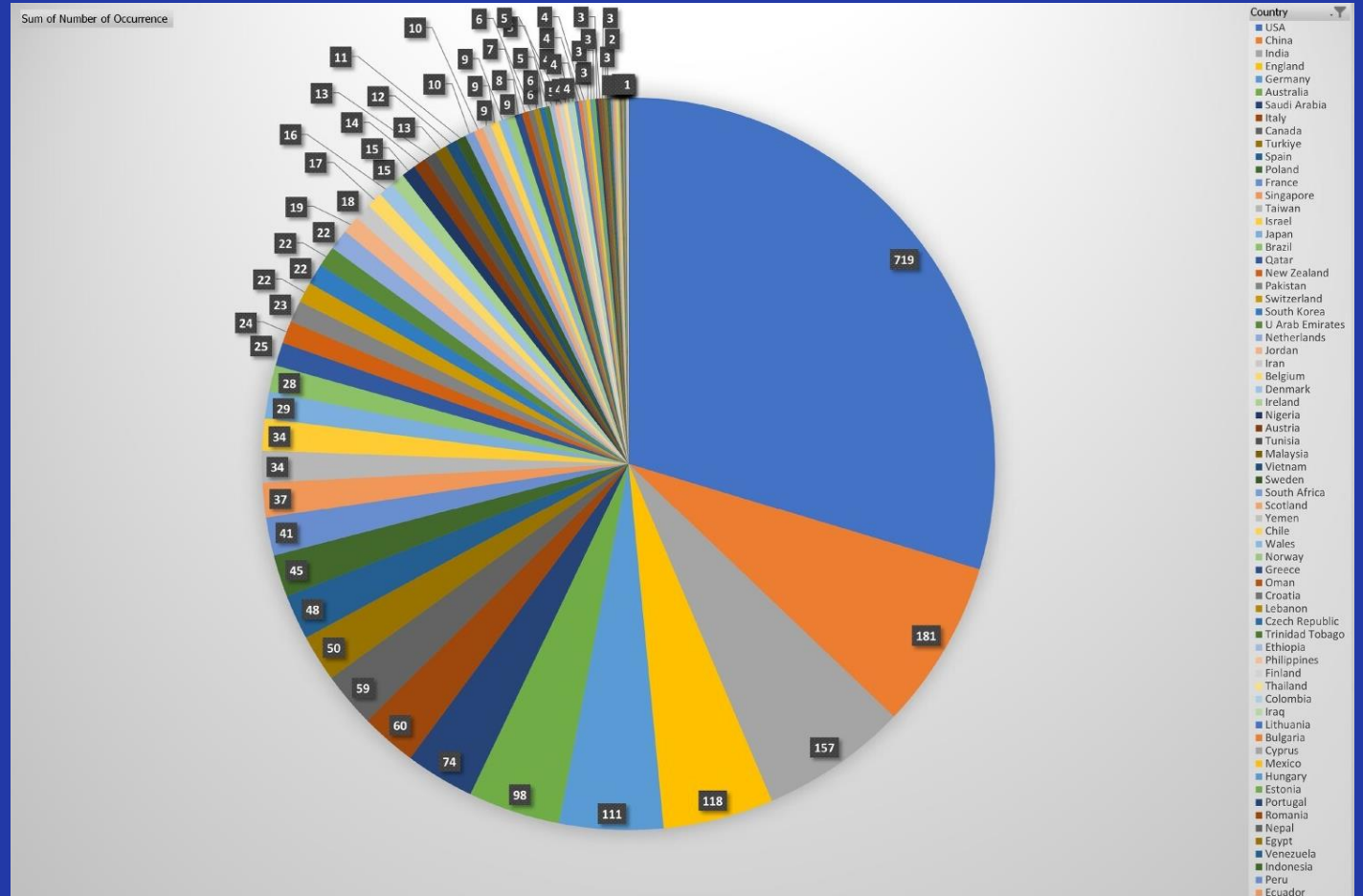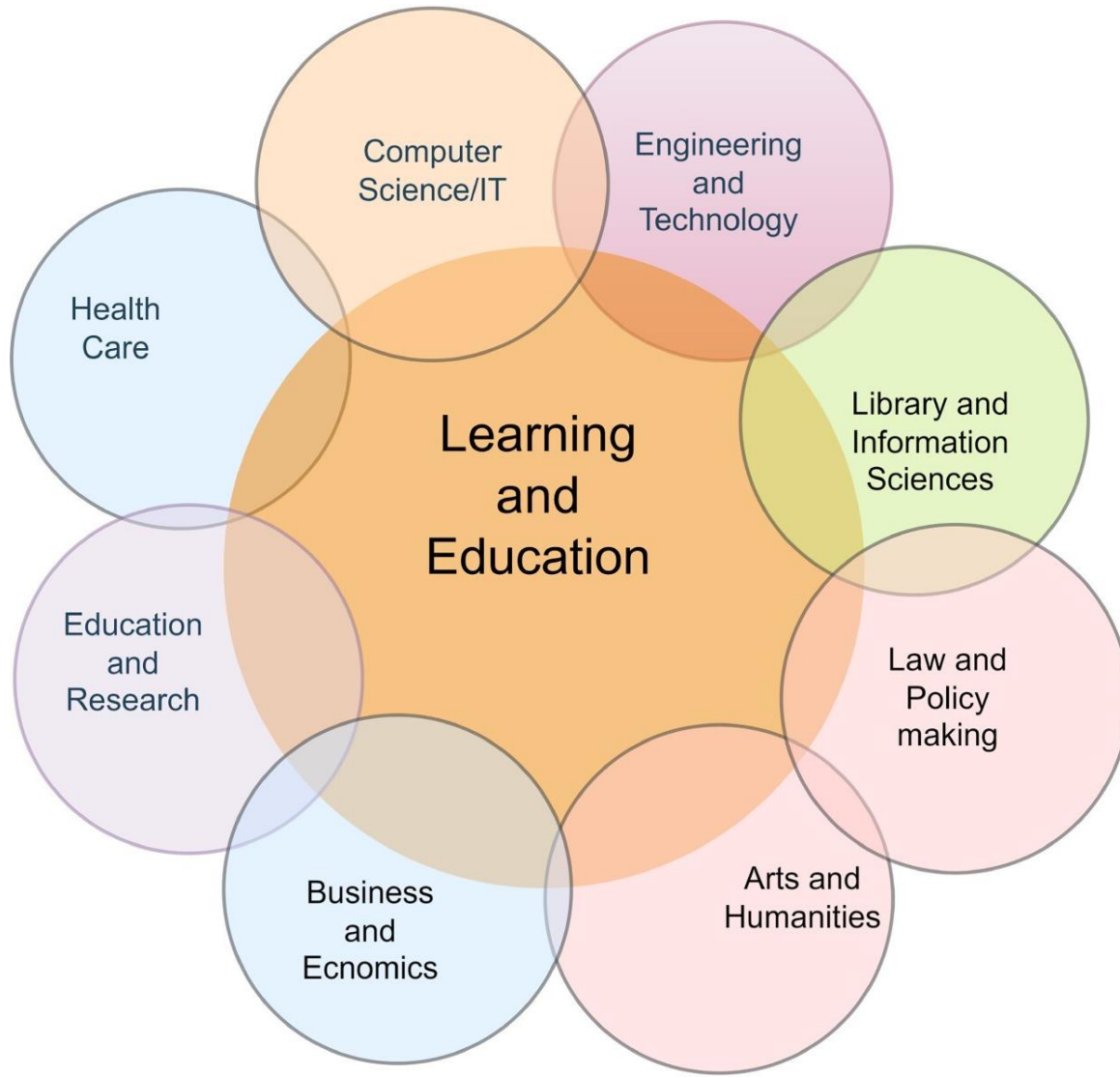
...and disappointments

Facts of ChatGPT Usage

Khan, N. et al. (2024) 'Global insights and the impact of generative AI-ChatGPT on multidisciplinary: a systematic review and bibliometric analysis', Connection Science, 36(1). doi: 10.1080/09540091.2024.2353630.
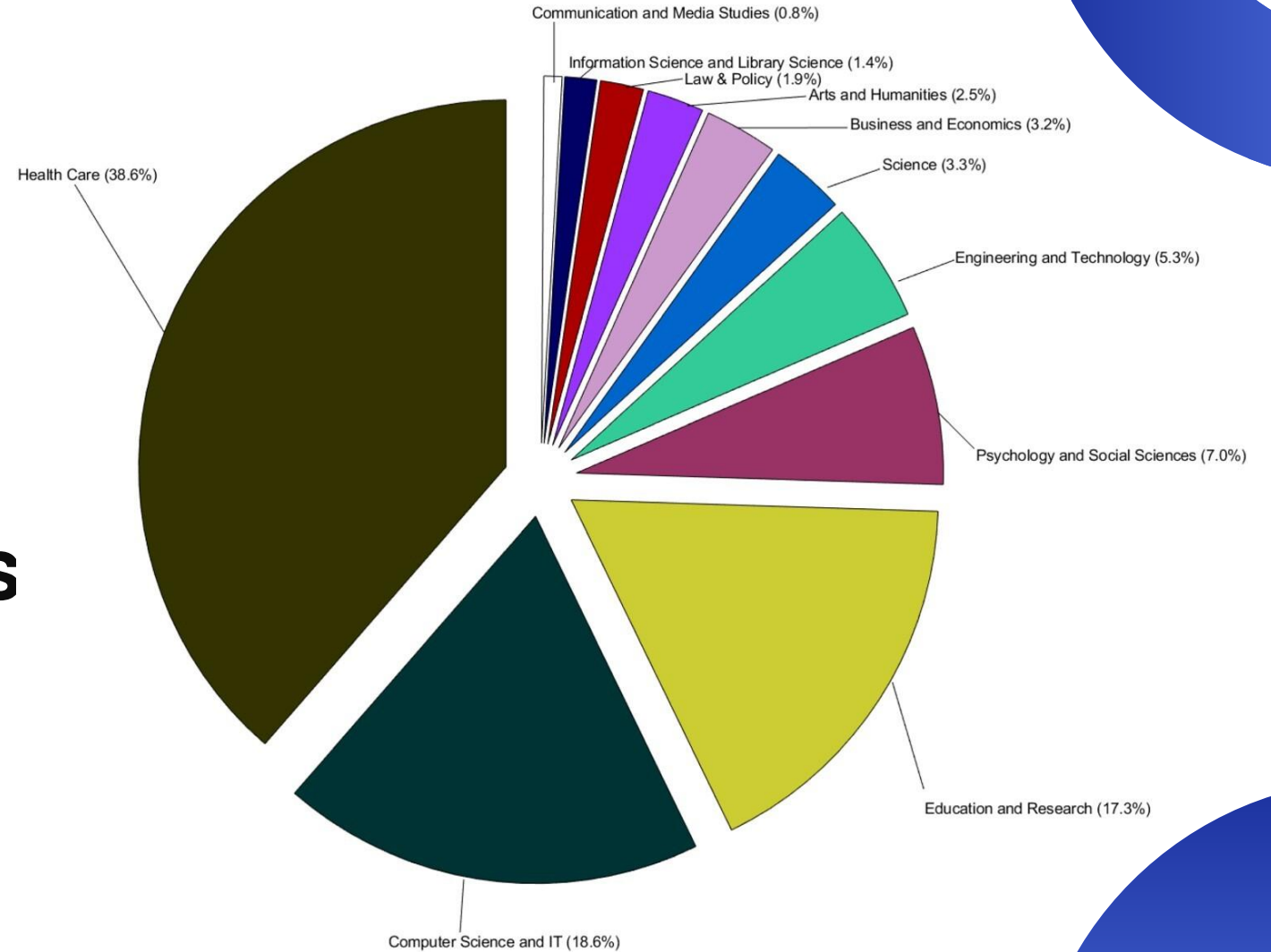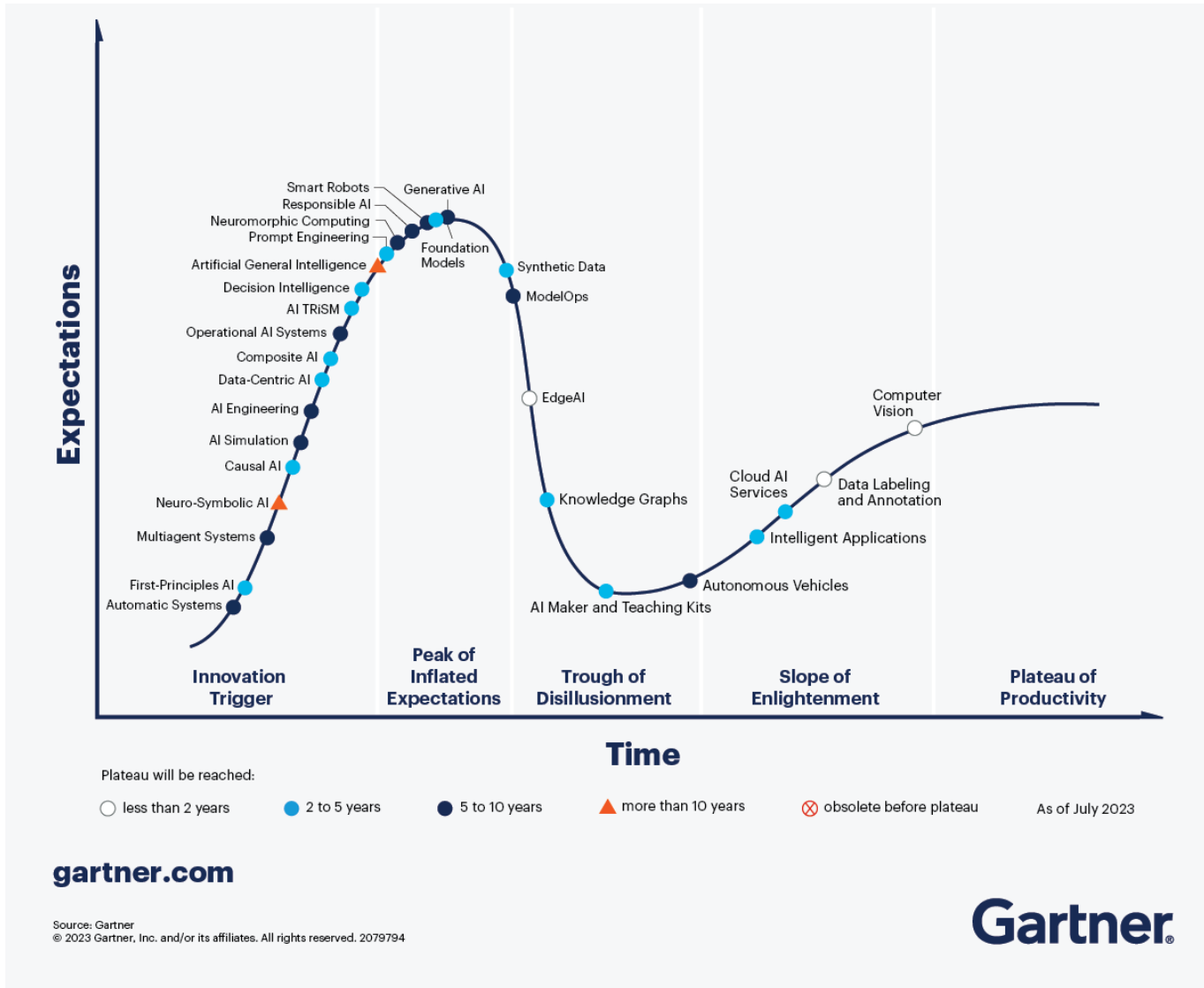
ChatGPT Usage Across All Disiplines

Khan, N. et al. (2024)

Distribution of Publication Across Disciplines

Khan, N. et al. (2024)

Communication and Media Studies (0.8%)
Information Science and Library Science (1.4%)
Law & Policy (1.9%)
Arts and Humanities (2.5%)
Business and Economics (3.2%)
Science (3.3%)
Engineering and Technology (5.3%)
Psychology and Social Sciences (7.0%)
Health Care (38.6%)
Education and Research (17.3%)
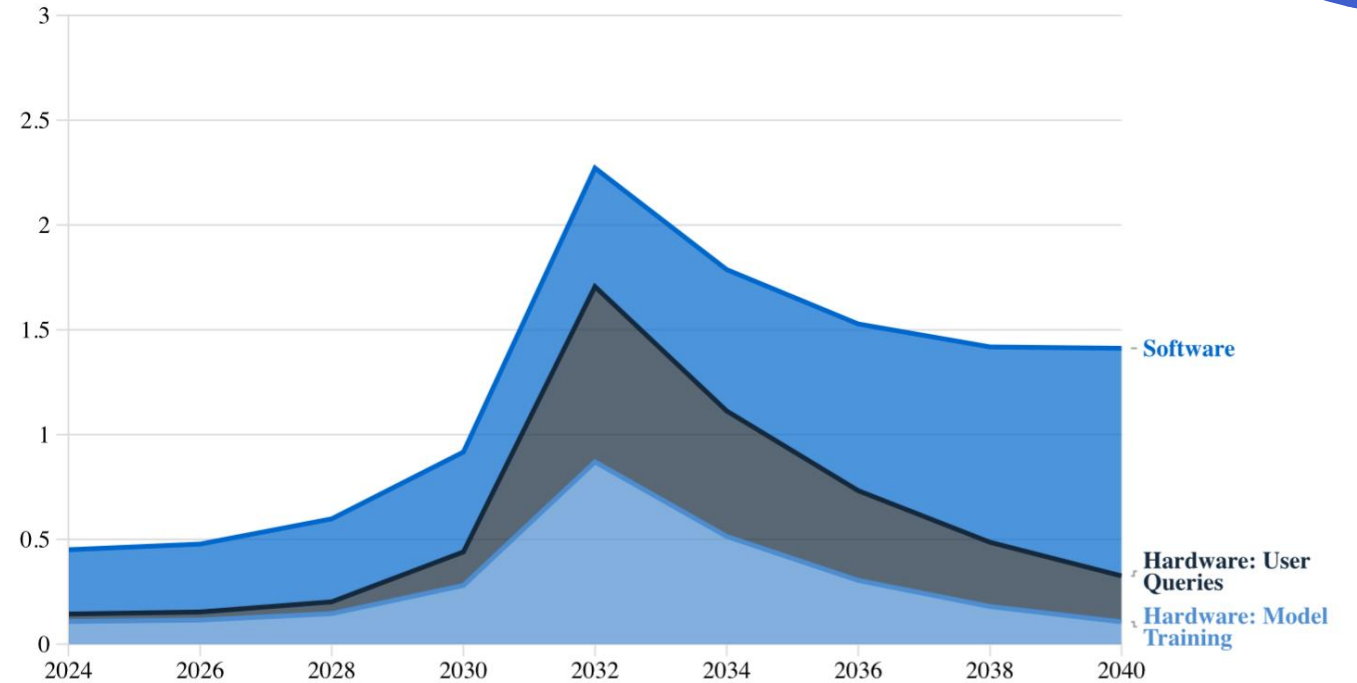Computer Science and IT (18.6%)
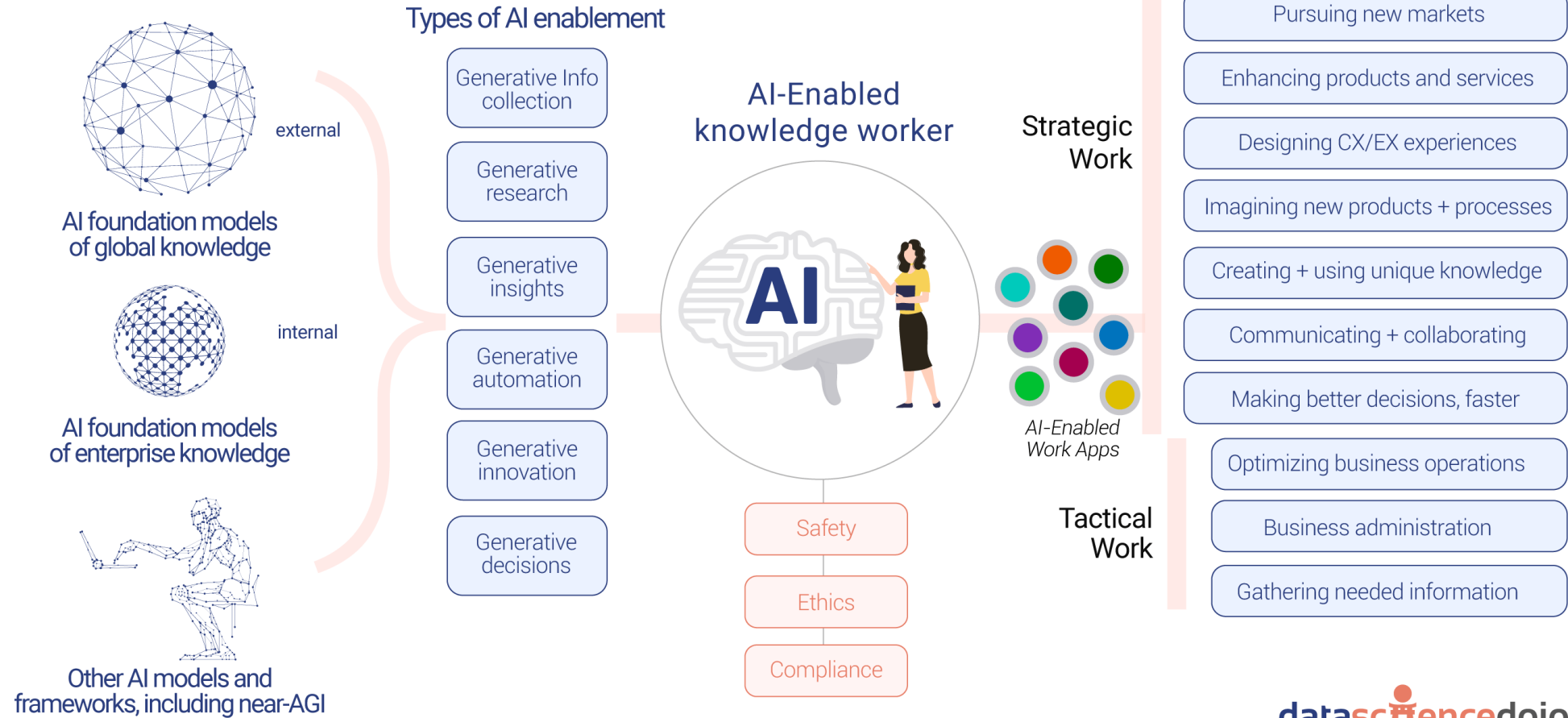
Hype Cycle for AI 2023

Stylized Us AI Investment Cycle, Percent of GDP

AI-related investment could peak at more than 2% of GDP, while software investment persists

Generative AI
Reshaping the future

How Artificial Intelligence will reshape the workplace and employee experience

AI foundation models of global knowledge

external

AI foundation models of enterprise knowledge

internal

Other AI models and frameworks, including near-AGI

Types of AI enablement

- Generative Info collection
- Generative research
- Generative insights
- Generative automation
- Generative innovation
- Generative decisions

AI-Enabled knowledge worker

AI

- Safety
- Ethics
- Compliance

AI-Enabled Work Apps

Strategic Work

- Pursuing new markets
- Enhancing products and services
- Designing CX/EX experiences
- Imagining new products + processes
- Creating + using unique knowledge
- Communicating + collaborating
- Making better decisions, faster

Tactical Work

- Optimizing business operations
- Business administration
- Gathering needed information

datasciencedojo
data science for everyone

THANK YOU

Universitas Pelita Harapan