VNU-HUS MAT1206E/3508: Introduction to AI

Introduction
In-class Discussion

Hoàng Anh Đức

Bộ môn Tin học, Khoa Toán-Cơ-Tin học
Đại học KHTN, ĐHQG Hà Nội
hoanganhduc@hus.edu.vn

Contents

What is Artificial Intelligence (AI)?

At a very first glance,

- Humans have *real* intelligence
- Machines have *artificial* intelligence

What is Artificial Intelligence (AI)?

At a very first glance,

- Humans have *real* intelligence
- Machines have *artificial* intelligence

Basic Idea

Artificial intelligence is based on the assumption that the process of human thought can be *mechanized*

- Human thought can be described as a mechanical sequence of steps that machines can execute.
- This implies that machines can be designed to simulate human cognitive processes, such as *adaptation, learning, and decision making*.

What is Artificial Intelligence (AI)?

At a very first glance,

- Humans have *real* intelligence
- Machines have *artificial* intelligence

Basic Idea

Artificial intelligence is based on the assumption that the process of human thought can be *mechanized*

- Human thought can be described as a mechanical sequence of steps that machines can execute.
- This implies that machines can be designed to simulate human cognitive processes, such as *adaptation, learning, and decision making*.

Note

Indeed, *there is no officially agreed definition of AI*. It is thus not easy to categorize what is AI and what is not

What is Artificial Intelligence (AI)?

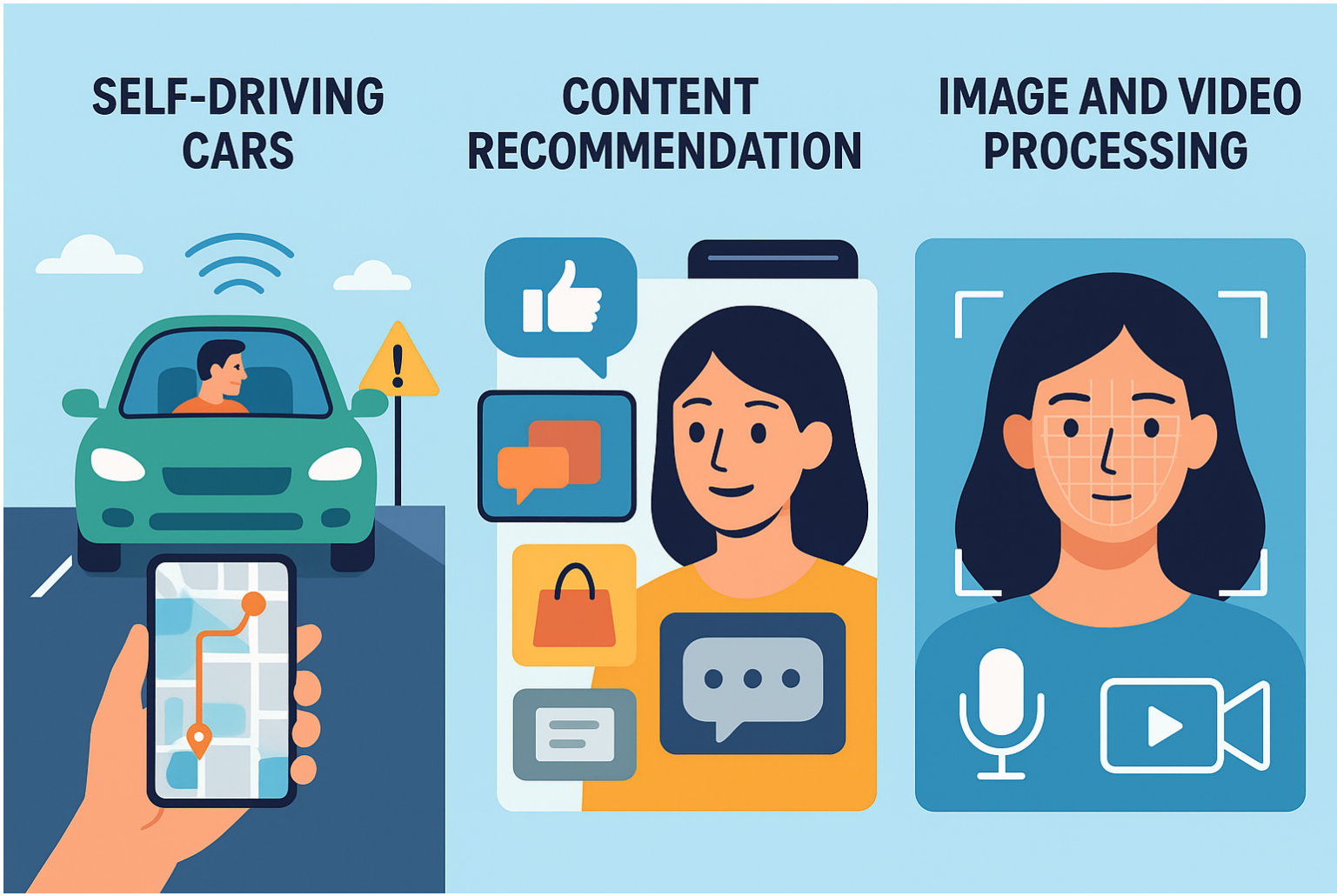


Figure: Some real-world applications which combine several "AI-techniques". Image generated by ChatGPT

What is Artificial Intelligence (AI)?

What do you think?

John McCarthy (1955) [one of the pioneers of AI]

The goal of AI is to develop machines that behave as though they were intelligent

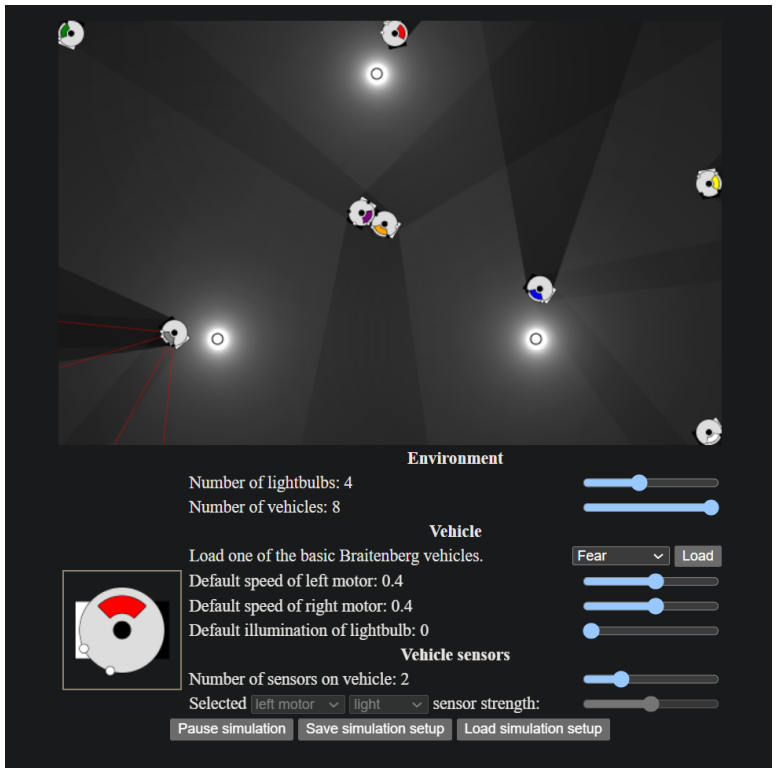


Figure: Example of Braitenberg vehicles simulating simple behaviors. (See http://www.harmendeweerd.nl/braitenberg-vehicles/)

What is Artificial Intelligence (AI)?

What do you think?

Encyclopedia Britannica (1991)

AI is the ability of digital computers or computer controlled robots to solve problems that are normally associated with the higher intellectual processing capabilities of humans ...

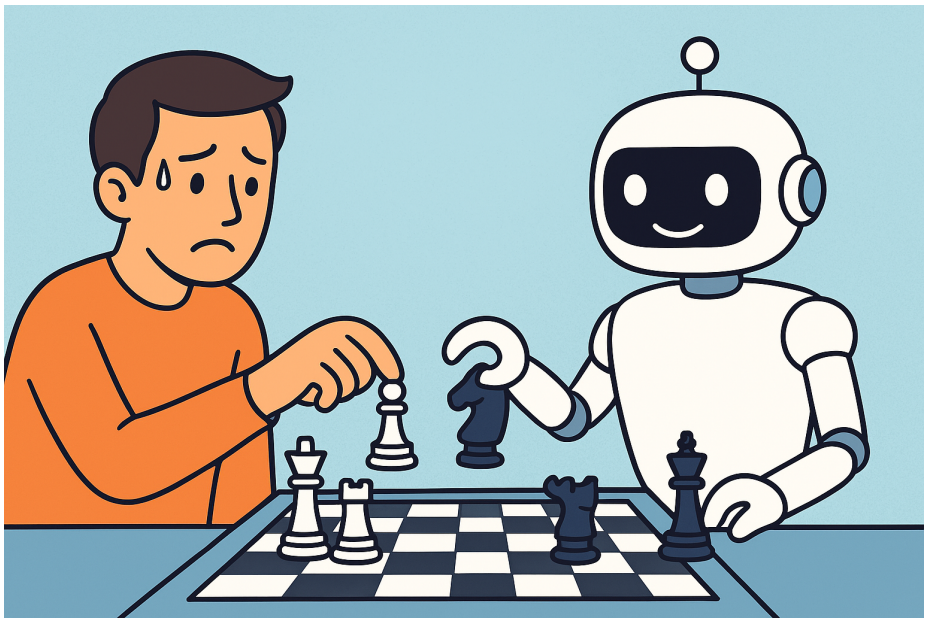


Figure: Computers can now beat humans in complex games (e.g., chess, Go), yet robots still struggle with dexterous tasks such as grasping and precise object placement. Image generated by ChatGPT

What is Artificial Intelligence (AI)?
What do you think?

Elaine Rich [Rich 1983]

Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better

Figure: AI systems can now outperform humans in tasks such as playing chess, yet they still have limitations in true understanding and general reasoning. (Image generated by Grok)

What is Artificial Intelligence (AI)?
What do you think?

Alan Turing [Turing 1950]

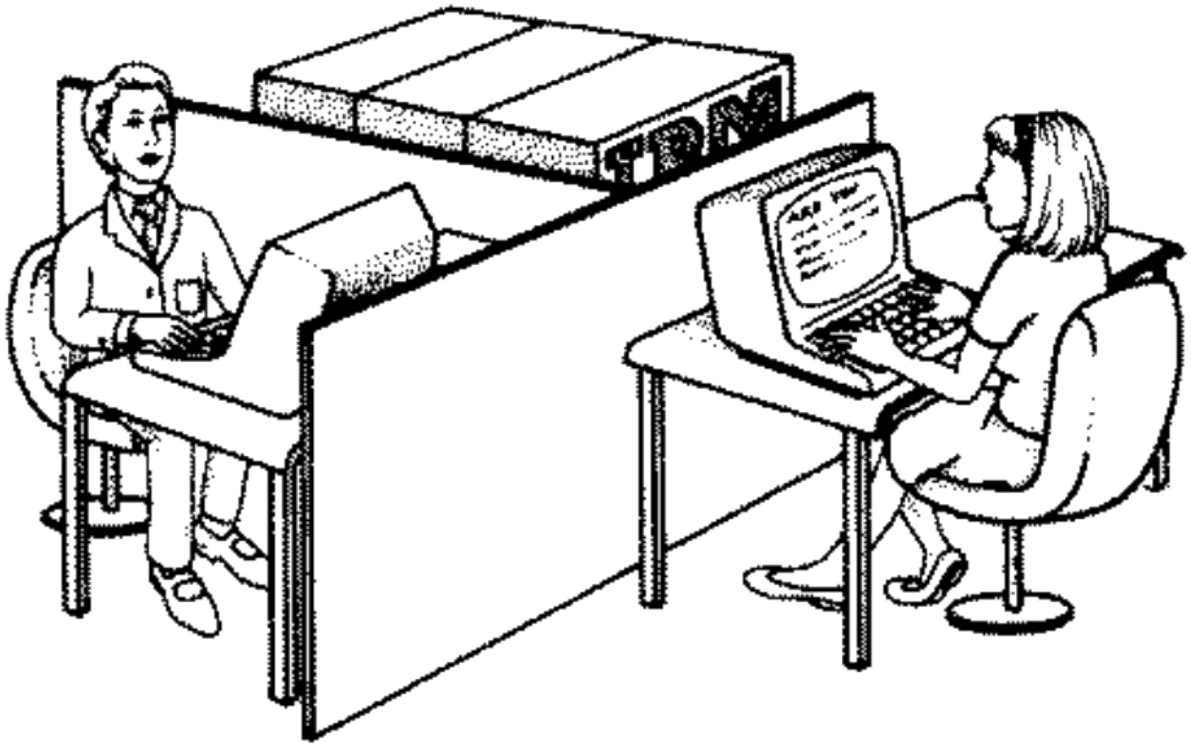


Figure: Turing test: The machine passes the test, if it can mislead Alice in 30% of the cases.

What is Artificial Intelligence (AI)?

More on Turing Test

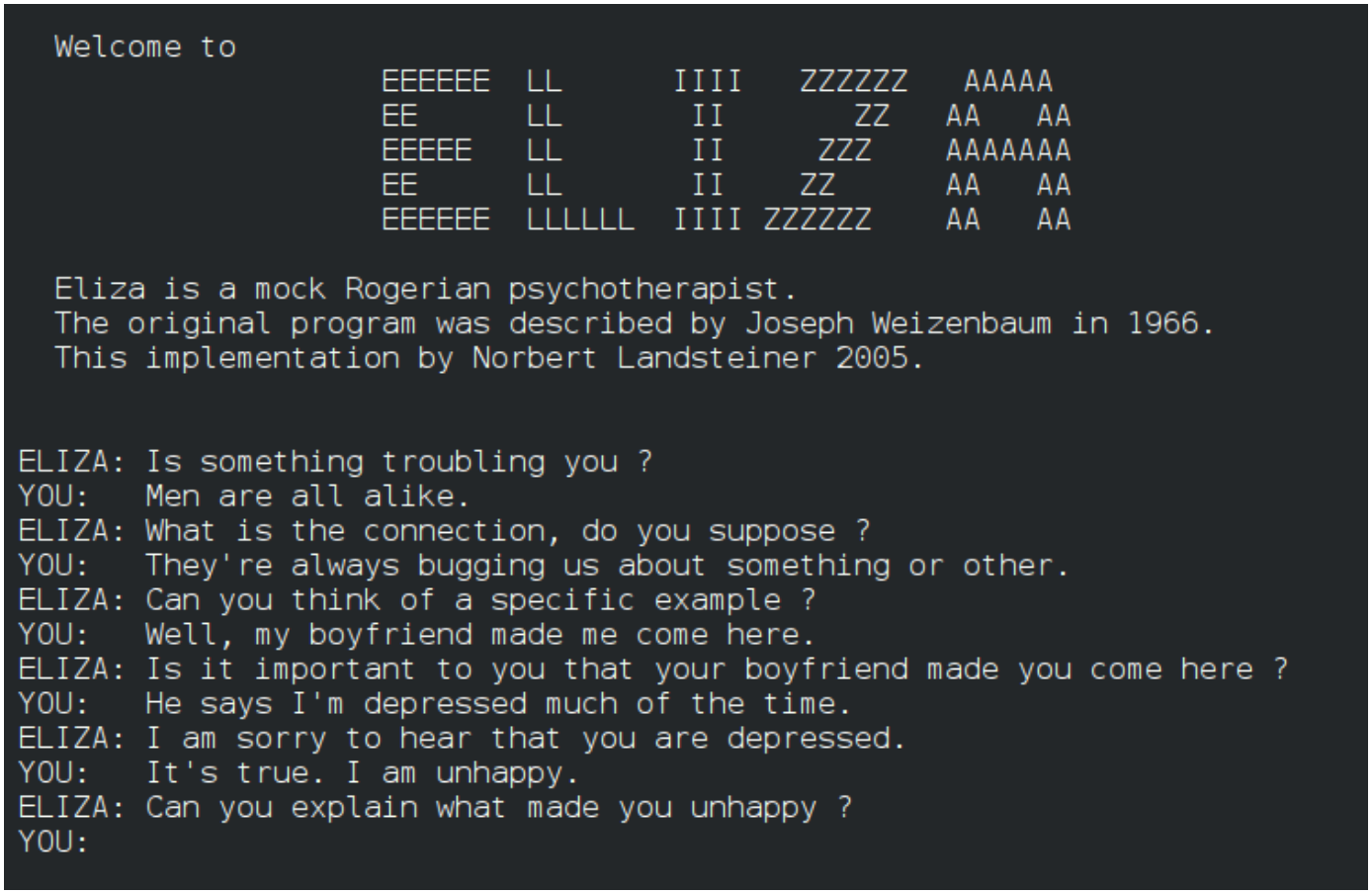


Figure: Joseph Weizenbaum's ELIZA (1966): One of the first chatterbots and one of the first programs capable of attempting the Turing Test.
(https://psych.fullerton.edu/mbirnbaum/psych101/Eliza.htm)

What is Artificial Intelligence (AI)?
More on Turing Test

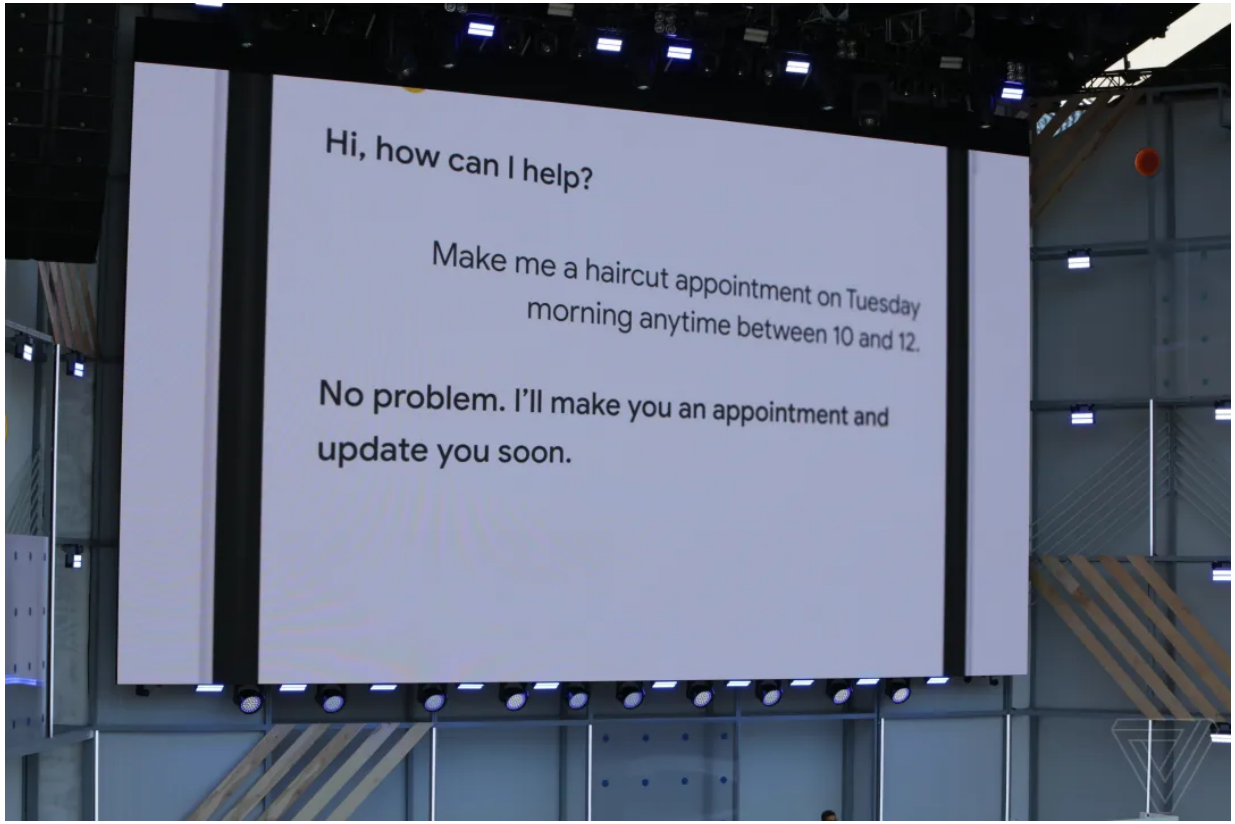


Figure: https://www.youtube.com/watch?v=pKVppdt_-B4. In 2018, at Google's I/O conference, CEO Sundar Pichai gave a standout demo of Google Assistant making a phone call to a hair salon and a restaurant to make appointments on behalf of a user. (Image taken from the Internet)

What is Artificial Intelligence (AI)?

More on Turing Test

There are contests, such as the Loebner Prize and the Turing Test Competition, to encourage people to create intelligent software systems that can pass the Turing Test.

Computer simulating 13-year-old boy becomes first to pass Turing test

'Eugene Goostman' fools 33% of interrogators into thinking it is human, in what is seen as a milestone in artificial intelligence

- In 'his own' words: how Eugene fooled the Turing judges
- What is the Turing test? And are we all doomed now?

What is Artificial Intelligence (AI)?

More on Turing Test

The idea that *"intelligence is the same as intelligent behavior"* has been challenged by some. The best known counter-argument is John Searle's Chinese Room Argument (1980)

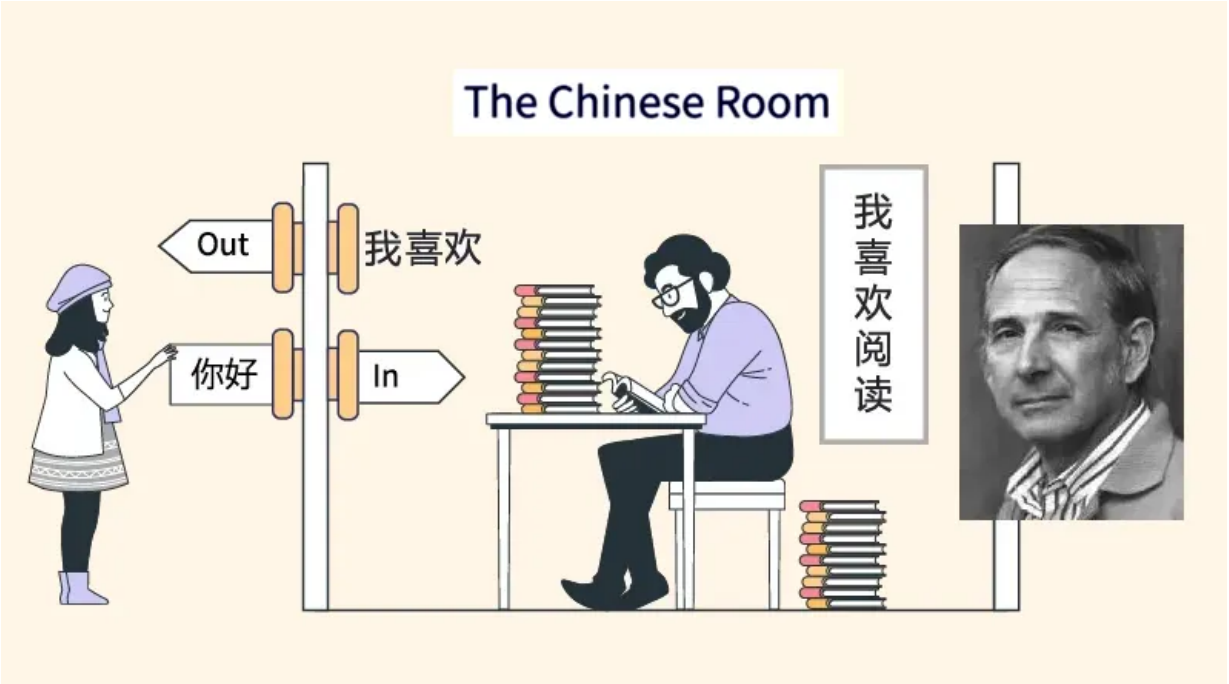


Figure: The Chinese Room Argument by John Searle (1980). Image taken from the Internet

What is Artificial Intelligence (AI)?

Searle also believed that there were two forms of AI: **weak (narrow) AI** and **strong (general) AI (Artificial General Intelligence)**.

- *Weak (narrow) AI* is any AI system that is designed and trained for a particular task (e.g., facial recognition, internet searches, or driving a car). *Most current AI applications are examples of weak AI.*
- *Strong (general) AI* is a form of AI that possesses the ability to understand, learn, and apply knowledge in a way that is indistinguishable from human intelligence. *Strong AI would be capable of performing any intellectual task that a human can do.*

Additionally, the strong development of current AI technologies suggests the idea of *artificial superintelligence (ASI)*, a level of intelligence that surpasses human intelligence across all domains. (For example, see https://www.ibm.com/think/topics/artificial-superintelligence)

What is Artificial Intelligence (AI)?

Exercise 1

There are a few other ways people define AI, for example,

- AI is any "cool things that computers can do"
- AI is any machine which simulates human behavior
- AI is an autonomous[1] and adaptive[2] system

Among all definitions we have mentioned, which one do you like *best*? How would *you* define AI?

Exercise 2

List some real-world applications of AI which we have not yet mentioned.

[1] The ability to perform tasks in complex environments without constant guidance by a user.

[2] The ability to improve performance by learning from experience.

Related Fields

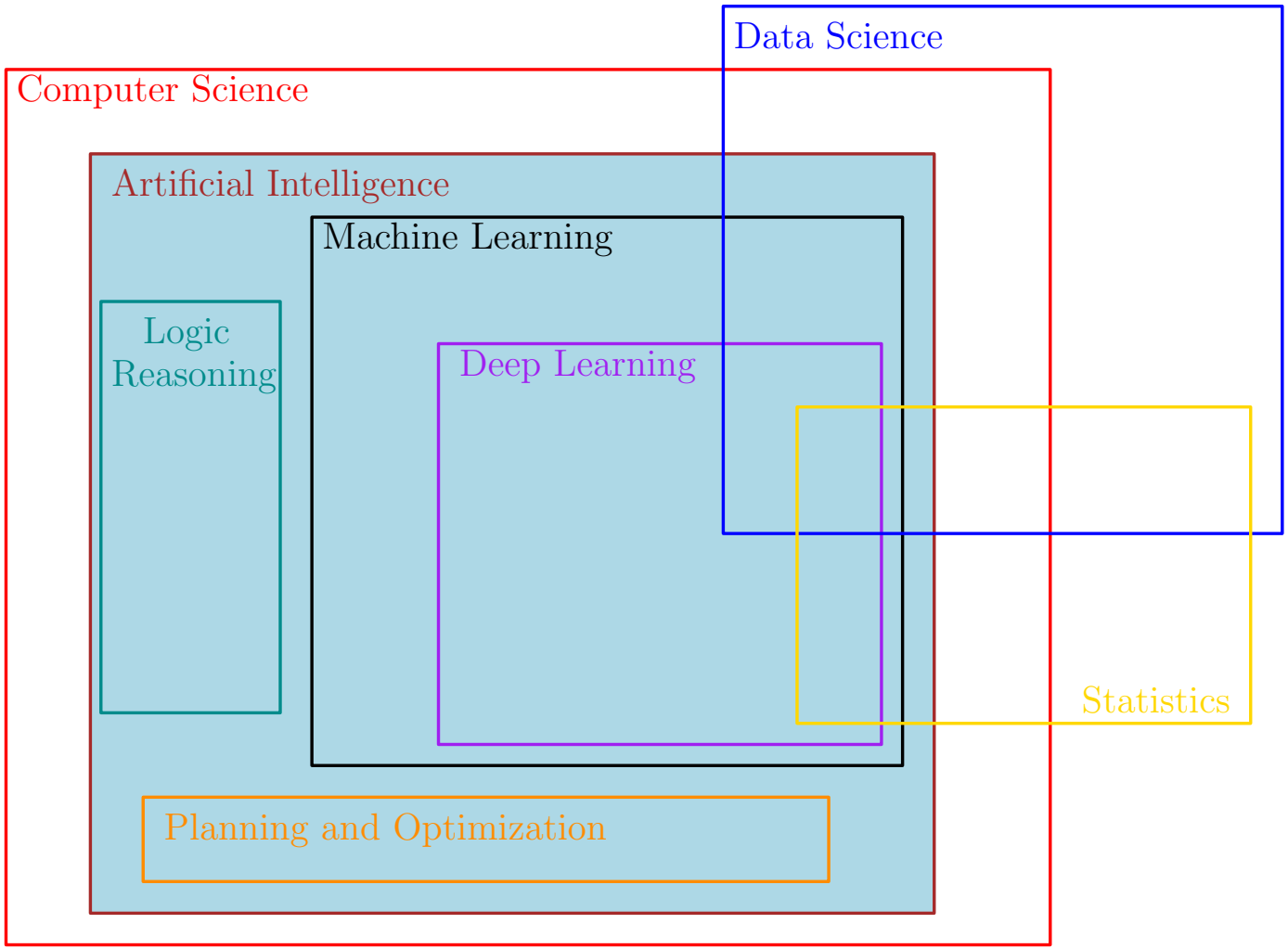


Figure: An Euler diagram of some related fields of AI. Overlapping areas indicate interdisciplinary connections.

History of AI

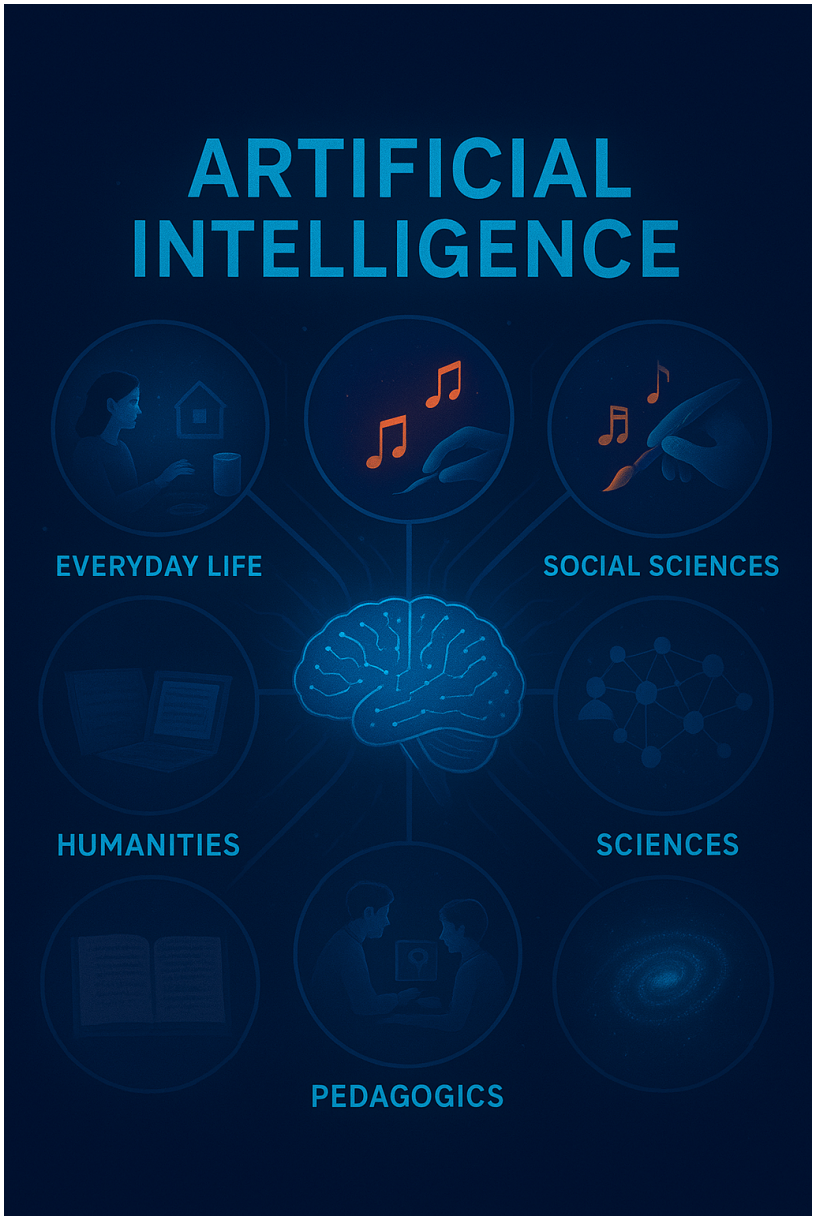
Note: There are different ways to present the history of AI. Here is what I think is a reasonable summary (which may probably contains some flaws.).

- **Early beginnings (1930s–1950s)**
  - Establishment of key theoretical foundations for AI: formal logic, computation theory, and early automata
- **Symbolic AI and the logic era (1950s–1970s) — AI is "hot"**
  - H. A. Simon (1965): "Machines will be capable, within twenty years, of doing any work a man can do."
  - H. A. Simon and Allen Newell (1958): predicted a computer would soon be the world's chess champion (that milestone was reached in 1997 with Deep Blue)
- **First AI Winter (1970s–1980s)**
  - Progress slowed, funding was reduced, and interest waned; researchers increasingly described their work as machine learning, pattern recognition, or informatics

History of AI (cont.)

- **Revival and the expert systems era (1980s–1990s)**
  - Renewed interest from industry driven by expert systems (rule-based systems capturing human expertise)
  - Connectionist ideas (neural networks) began to re-emerge with practical learning algorithms
- **Consolidation and deployment (1990s–2010s)**
  - Many long-standing goals were achieved and AI components began appearing in commercial products
  - Growth of distributed, autonomous, and learning agents for real-world applications
- **Deep learning and modern AI (2010s–present)**
  - Breakthroughs in deep neural networks delivered dramatic improvements across perception and language tasks
  - Massive growth in available data and compute (GPUs, cloud infrastructure, open-source frameworks)
  - AI became pervasive in consumer products and services (phones, cars, web services), raising important social and ethical questions

The History of AI
Today



AI is everywhere. Image generated by ChatGPT

The History of AI

Why AI is “hot” today?

- **Visible real-world impact**
  - Modern AI is bypassing grand questions about meaning of intelligence, the mind, and consciousness, and focusing on building practically useful solutions in real-world problems.
  - AI is now embedded in everyday life (virtual assistants, recommender systems, ChatGPT) and is accessible for both personal and professional use.
  - Research and industry are closely connected: practical, production-ready solutions are being deployed across healthcare, finance, education, and entertainment.
  - For example, the course “*Khoá học Ứng dụng AI cho Thanh niên Việt Nam*” demonstrates how AI can streamline daily tasks and workflows.
- **Mature compute and infrastructure**
  - Large-scale data, powerful accelerators (GPUs/TPUs), cloud platforms, and open-source frameworks make training and deploying models feasible and cost-effective.
  - The ecosystem reduces the barrier to experiment, iterate, and bring new AI methods into production quickly.

The History of AI

Exercise 3

Do your own research on the history of AI. Some intere sting topics would probably be:

(a) The *Dartmouth Conference* and the birth of AI

(b) The rise and fall of *expert systems*

(c) The history and impact of *neural networks and deep learning* on AI

(d) The history and development of *two major theories of AI*: *symbolic AI* (AI should be based on traditional computer logic or preprogramming $\Rightarrow$ Modeling human's *mind*) and *connectionist AI* (AI needed to use systems similar to the brain like neural networks $\Rightarrow$ Modeling human's *brain*). (Some scientists want to go further by blending the two approaches into something called *neuro-symbolic AI*.)

Implications of AI on the Society

Implication 1: Algorithmic Bias

Figure: An article from The Guardian (2015) on Google's job ads showing bias against women.
(https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study)

Implications of AI on the Society

Implication 1: Algorithmic Bias

Google searches expose racial bias, says study of names

4 February 2013

Prof Sweeney said technology could be used to counteract racial intolerance

A study of Google searches has found "significant discrimination" in advert results depending on the perceived race of names searched for.

Harvard professor Latanya Sweeney said names typically associated with black people were more likely to produce ads related to criminal activity.

Figure: BBC News (2013): Google's search algorithm showing bias against women (https://www.bbc.com/news/technology-21322183)

Implications of AI on the Society

Implication 1: Algorithmic Bias

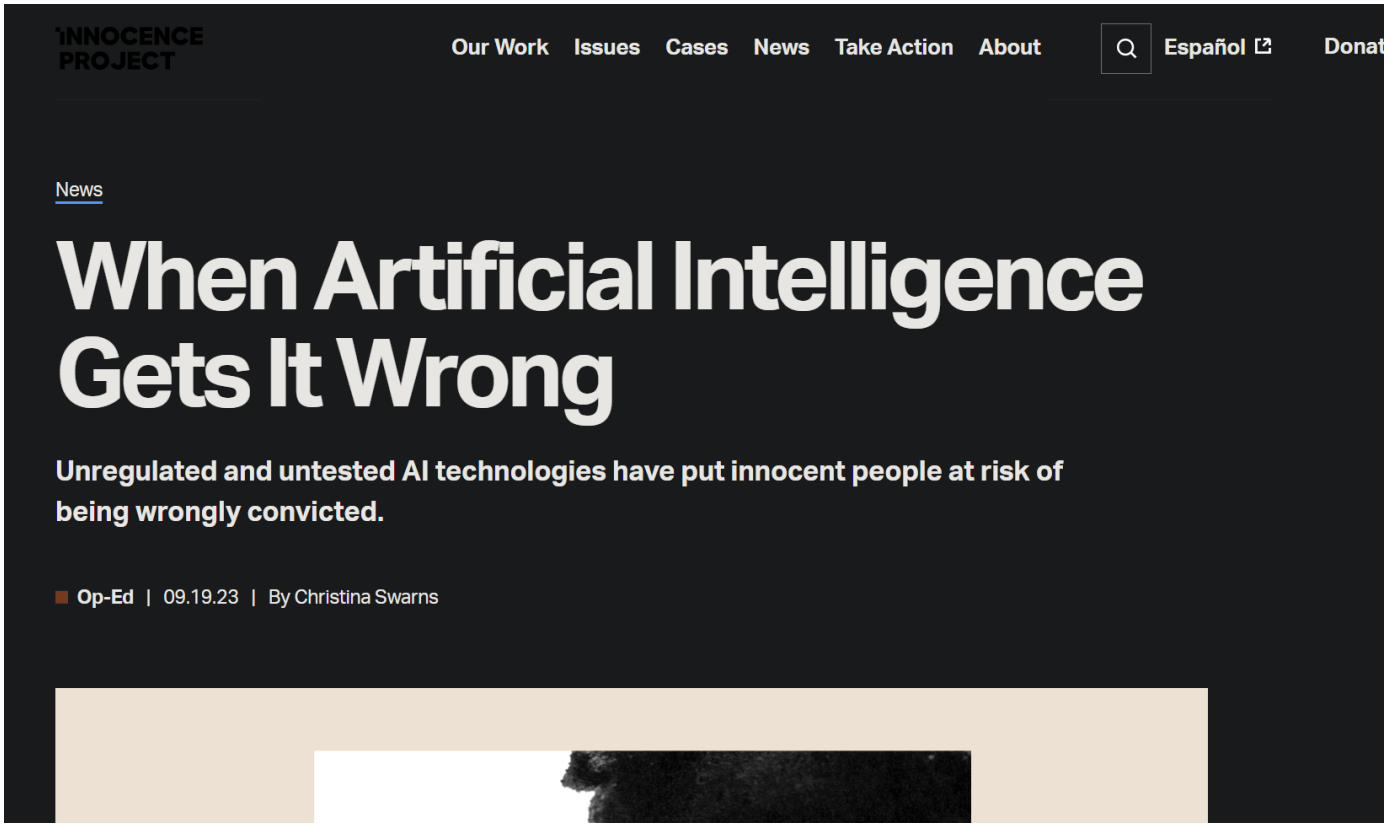


Figure: Innocence Project (2023): report on AI misidentifying individuals in criminal cases because of the bias inherent in the data used to train existing facial recognition algorithms. (https://innocenceproject.org/news/when-artificial-intelligence-gets-it-wrong/)

Implication 1: Algorithmic Bias

- Algorithmic bias isn't a hypothetical threat conceived by academic researchers. It's a real phenomenon that is already affecting people today.
- The main reason for algorithmic bias is *human bias in the data*.
- The main difficulty in the use of AI and machine learning instead of rule-based systems is their *lack of transparency*
  - Algorithms and datasets are often proprietary trade secrets and not open to public scrutiny.
  - Even with access, it is often difficult to identify which algorithmic components or data elements lead to discriminatory decisions.

Implications of AI on the Society

Implication 2: Seeing is no longer believing

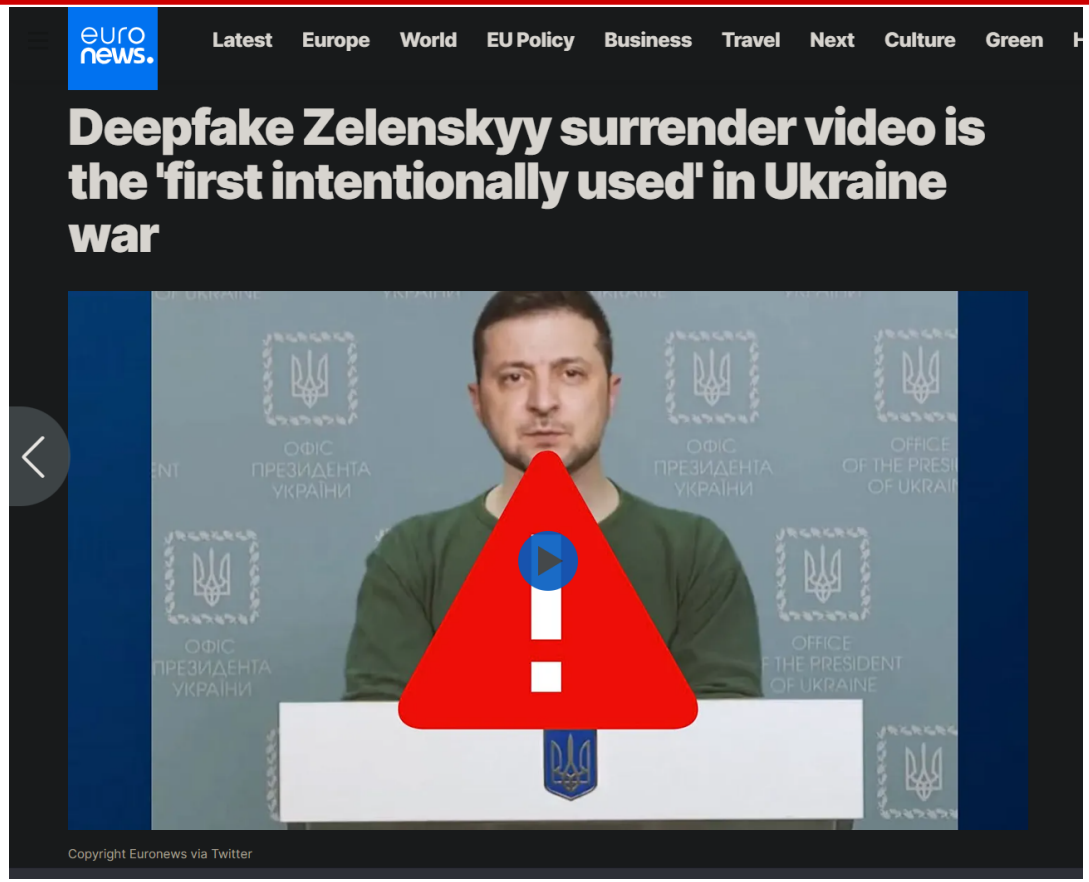


Figure: Euronews (2022): report on a deepfake video of Ukrainian President Volodymyr Zelenskyy calling citizens to surrender to Russia. (https://www.euronews.com/my-europe/2022/03/16/deepfake-zelenskyy-surrender-video-is-the-first-intentionally-used-in-ukraine-war)

Implications of AI on the Society

Implication 2: Seeing is no longer believing

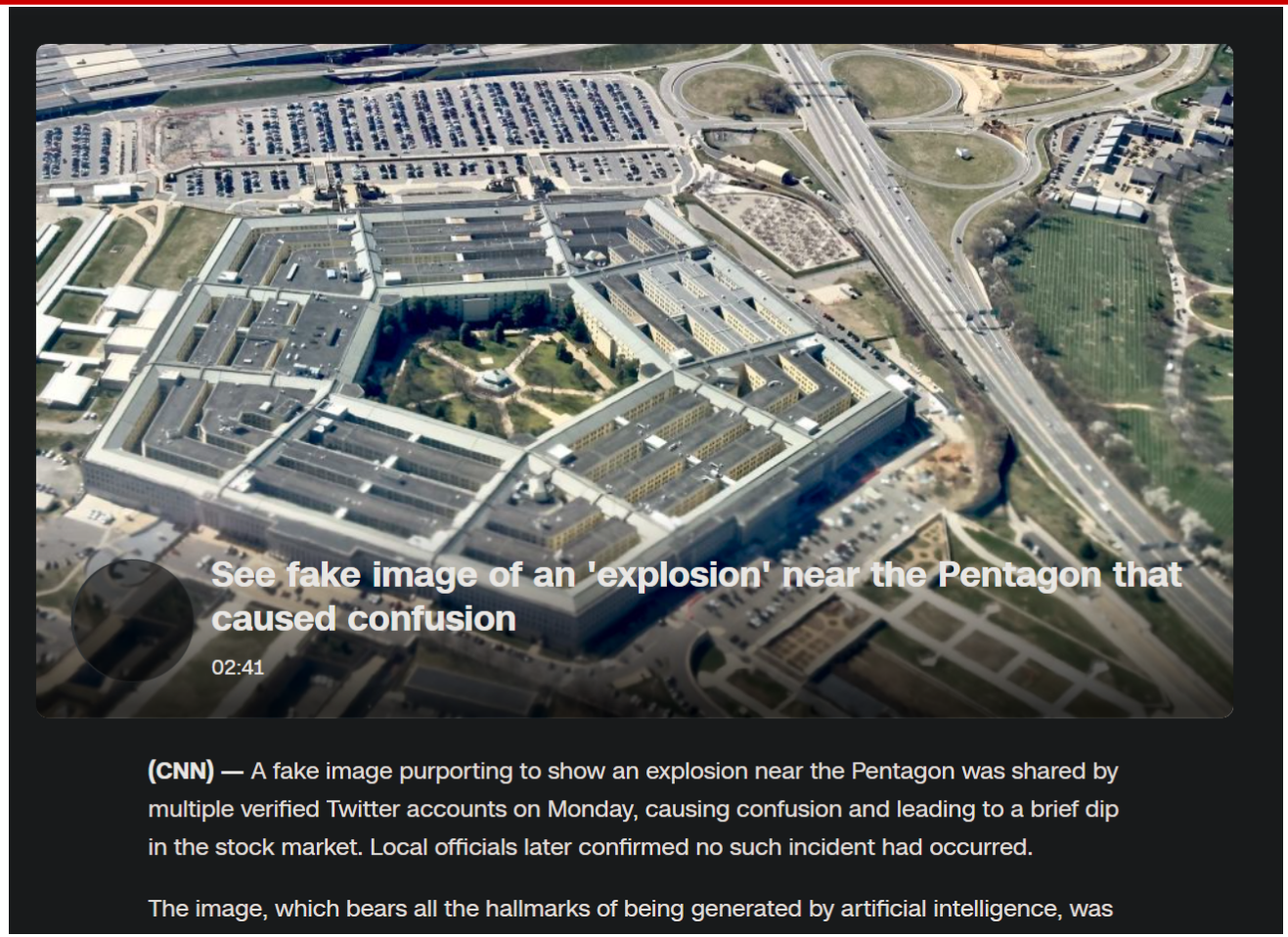


Figure: CNN (2023): report on a fake image generated by AI that caused a brief panic on Twitter and the stock market. (https://edition.cnn.com/2023/05/22/tech/twitter-fake-image-pentagon-explosion)

Implications of AI on the Society

Implication 2: Seeing is no longer believing

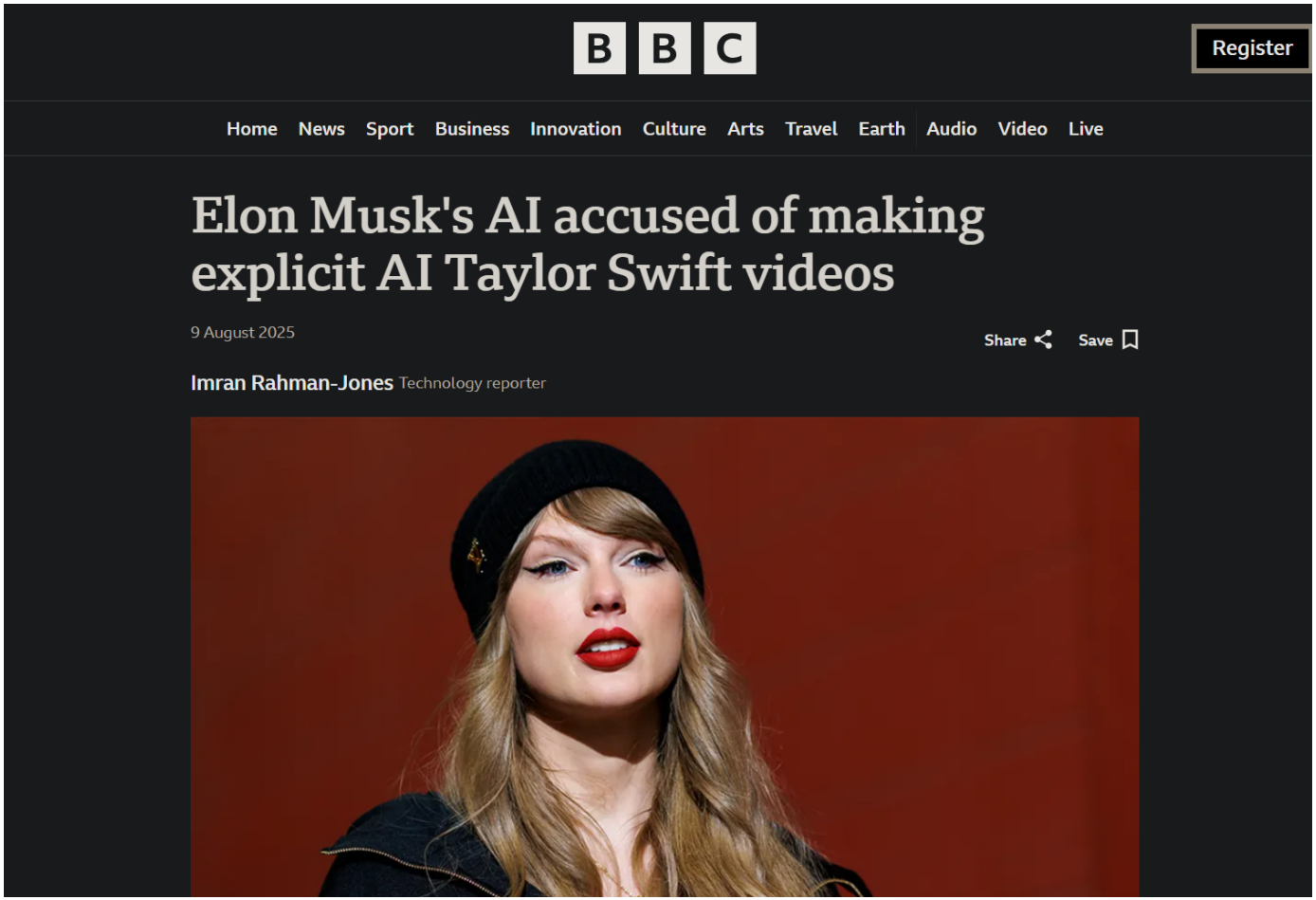


Figure: BBC News (2025): report on Elon Musk's AI accused of making explicit AI Taylor Swift videos using deepfakes. (https://www.bbc.com/news/articles/cwye62e1ndjo

Implications of AI on the Society

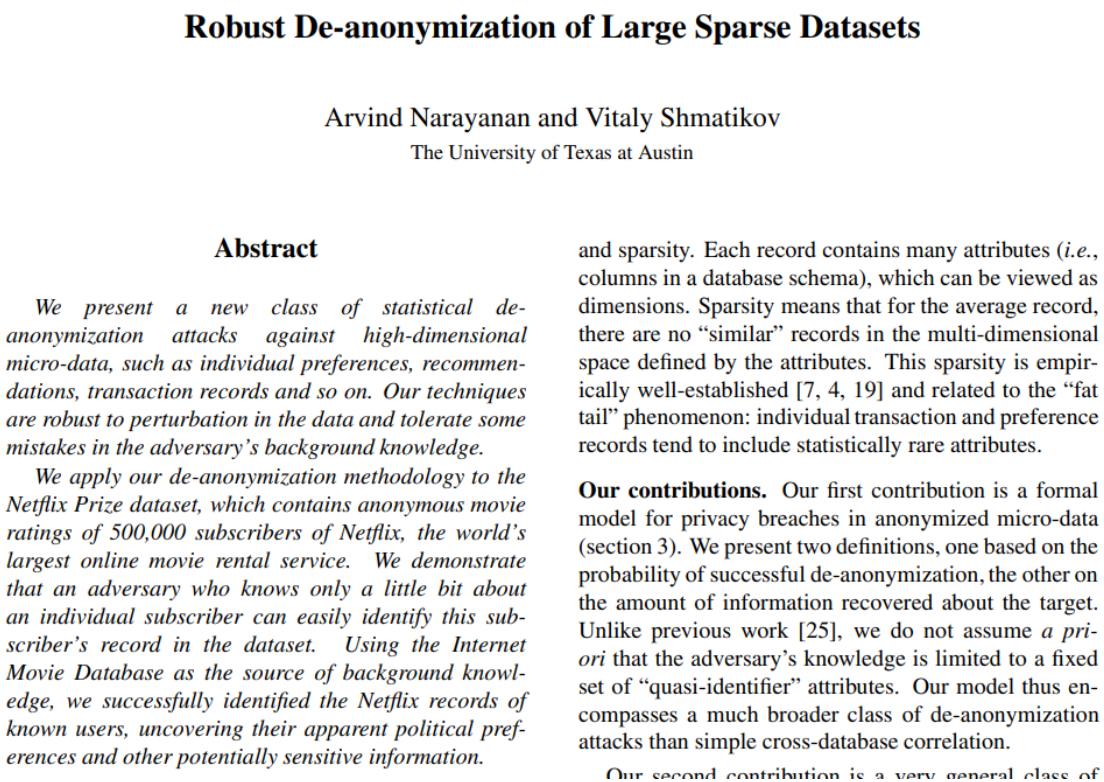
Implication 2: Seeing is no longer believing

AI is *taking the possibilities of fabricating evidence to a whole new level*

- *Deepfake* and similar technologies can create *highly realistic but entirely fabricated images, videos, and audio recordings* of people saying or doing things they never actually said or did.
  - Metaphysics Live (`https://www.metaphysic.ai/`) is a system capable of doing face-swaps, de-aging and other tricks in real time.
  - Descript (`https://www.descript.com/`) has a tool for automatic imitation of a person's voice from a few minutes of sample recording.
- These technologies pose significant challenges for *trust and verification* in media, as it becomes increasingly difficult to distinguish between real and manipulated content.

Implications of AI on the Society

Implication 3: Privacy concerns



Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov
The University of Texas at Austin

Abstract

We present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. Our techniques are robust to perturbation in the data and tolerate some mistakes in the adversary's background knowledge.

We apply our de-anonymization methodology to the Netflix Prize dataset, which contains anonymous movie ratings of 500,000 subscribers of Netflix, the world's largest online movie rental service. We demonstrate that an adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record in the dataset. Using the Internet Movie Database as the source of background knowledge, we successfully identified the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information.

and sparsity. Each record contains many attributes (*i.e.*, columns in a database schema), which can be viewed as dimensions. Sparsity means that for the average record, there are no "similar" records in the multi-dimensional space defined by the attributes. This sparsity is empirically well-established [7, 4, 19] and related to the "fat tail" phenomenon: individual transaction and preference records tend to include statistically rare attributes.

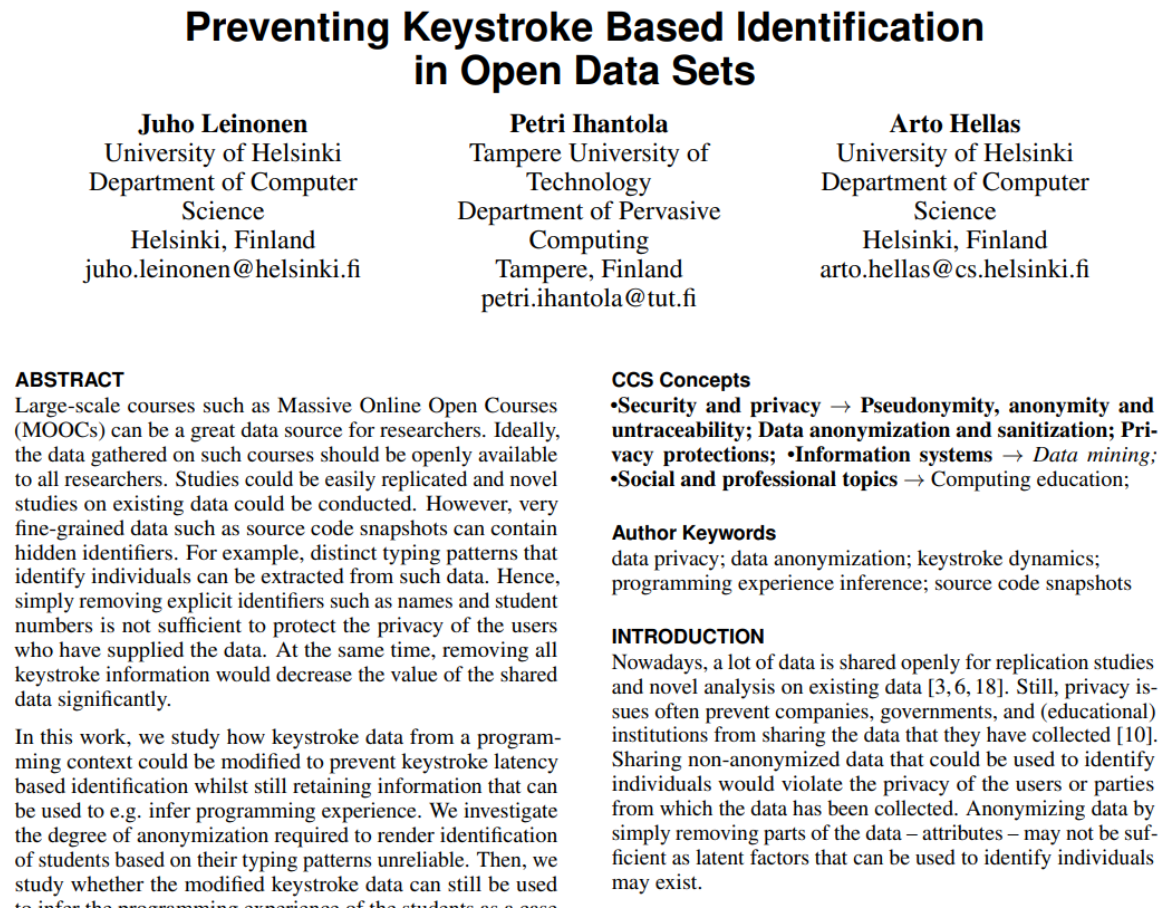
Our contributions. Our first contribution is a formal model for privacy breaches in anonymized micro-data (section 3). We present two definitions, one based on the probability of successful de-anonymization, the other on the amount of information recovered about the target. Unlike previous work [25], we do not assume *a priori* that the adversary's knowledge is limited to a fixed set of "quasi-identifier" attributes. Our model thus encompasses a much broader class of de-anonymization attacks than simple cross-database correlation.

Our second contribution is a very general class of

Figure: A research in 2006: researchers from the University of Texas at Austin re-identified Netflix users by cross-referencing the anonymized movie ratings with publicly available datasets. (https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf)

Implications of AI on the Society

Implication 3: Privacy concerns



Preventing Keystroke Based Identification in Open Data Sets

Juho Leinonen
University of Helsinki
Department of Computer Science
Helsinki, Finland
juho.leinonen@helsinki.fi

Petri Ihantola
Tampere University of Technology
Department of Pervasive Computing
Tampere, Finland
petri.ihantola@tut.fi

Arto Hellas
University of Helsinki
Department of Computer Science
Helsinki, Finland
arto.hellas@cs.helsinki.fi

ABSTRACT

Large-scale courses such as Massive Online Open Courses (MOOCs) can be a great data source for researchers. Ideally, the data gathered on such courses should be openly available to all researchers. Studies could be easily replicated and novel studies on existing data could be conducted. However, very fine-grained data such as source code snapshots can contain hidden identifiers. For example, distinct typing patterns that identify individuals can be extracted from such data. Hence, simply removing explicit identifiers such as names and student numbers is not sufficient to protect the privacy of the users who have supplied the data. At the same time, removing all keystroke information would decrease the value of the shared data significantly.

In this work, we study how keystroke data from a programming context could be modified to prevent keystroke latency based identification whilst still retaining information that can be used to e.g. infer programming experience. We investigate the degree of anonymization required to render identification of students based on their typing patterns unreliable. Then, we study whether the modified keystroke data can still be used

CCS Concepts

•**Security and privacy** → **Pseudonymity, anonymity and untraceability; Data anonymization and sanitization; Privacy protections;** •**Information systems** → *Data mining;* •**Social and professional topics** → Computing education;

Author Keywords

data privacy; data anonymization; keystroke dynamics; programming experience inference; source code snapshots

INTRODUCTION

Nowadays, a lot of data is shared openly for replication studies and novel analysis on existing data [3, 6, 18]. Still, privacy issues often prevent companies, governments, and (educational) institutions from sharing the data that they have collected [10]. Sharing non-anonymized data that could be used to identify individuals would violate the privacy of the users or parties from which the data has been collected. Anonymizing data by simply removing parts of the data – attributes – may not be sufficient as latent factors that can be used to identify individuals may exist.

Figure: A research in 2017: researchers demonstrated a method to prevent user identification based on typing patterns data. (https://doi.org/10.1145/3051457.3051458)

Implications of AI on the Society

Implication 3: Privacy concerns

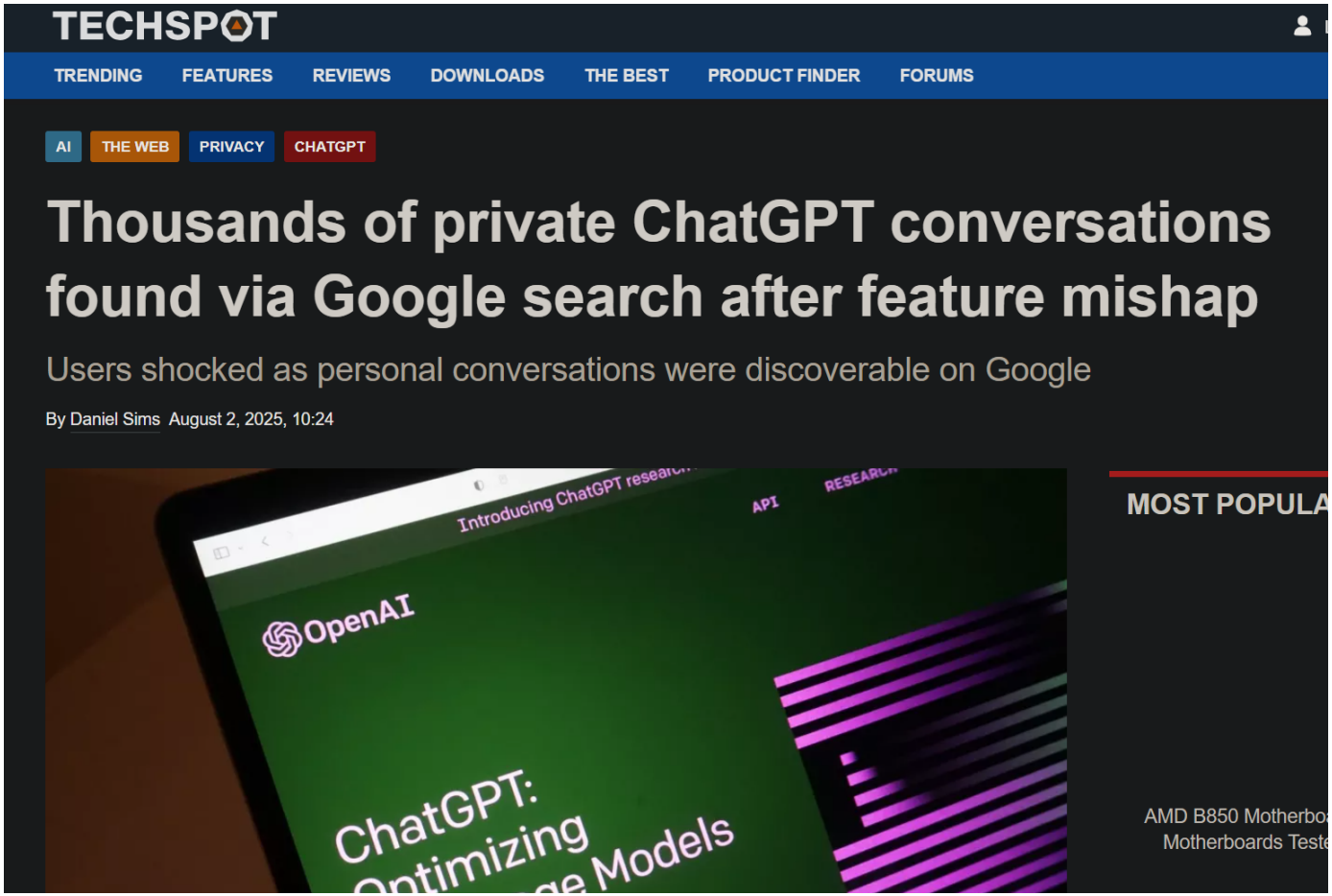


Figure: TechSpot (2025): report on thousands of private ChatGPT conversations found in Google search results. (https://www.techspot.com/news/108911-thousands-private-chatgpt-conversations-found-google-search-after.html)

Implications of AI on the Society

Implication 3: Privacy concerns

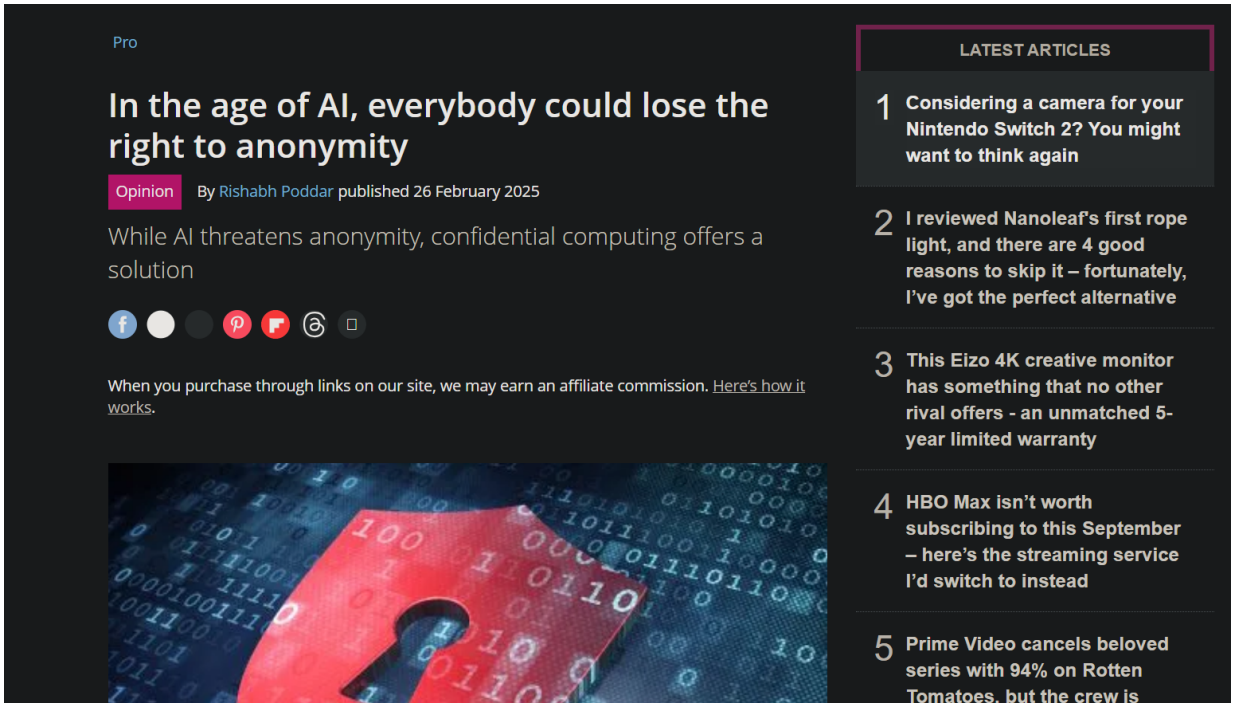


Figure: TechRadar (2025): report on the potential loss of anonymity in the age of AI where the author, as a proof-of-concept, used GPT model and the information from Personal Genome Project (PGP) dataset to match publicly available biographical data of prominent individuals to anonymized profiles within the dataset. (https://www.techradar.com/pro/in-the-age-of-ai-everybody-could-lose-the-right-to-anonymity)

Implications of AI on the Society

Implication 3: Privacy concerns

- Before “the age of AI”, privacy concerns were primarily focused on data collection and usage by companies.
  - In principle, every click, scroll and the time spent on content can be recorded.
  - Websites and trackers can access or infer your browsing history, enabling cross-site profiling and persistent personalization.
  - Consequently, a simple search for Android phones can trigger targeted ads (e.g., cheap and discounted options) across other sites unless privacy protections (incognito, trackers blocked) are used.
- In “the age of AI”, privacy concerns have evolved. Some examples include:
  - The re-identification of anonymized data, as seen in the Netflix case.
  - Data leaks, such as the ChatGPT conversations found in Google search results.
  - The potential loss of anonymity through advanced profiling techniques.
  - The potential of using AI to match user accounts in almost any service that collects detailed data about user behaviors (e.g., typing patterns).

Implications of AI on the Society

Implication 4: Security concerns

Figure: SocRadar (2024): report on various examples of AI-assisted cyber attacks, highlighting the evolving threat landscape. (https://socradar.io/examples-of-ai-assisted-cyber-attacks/)

Implications of AI on the Society

Implication 4: Security concerns

Figure: The Hacker News (2025): report on how AI-powered browsers (such as Perplexity's Comet) can be tricked into executing malicious commands, highlighting new security vulnerabilities. (https://thehackernews.com/2025/08/experts-find-ai-browsers-can-be-tricked.html)

Implications of AI on the Society
Implication 4: Security concerns

- The rise of AI technologies has led to new security vulnerabilities, as seen in various incidents involving data leaks and unauthorized access.
- Organizations are increasingly implementing strict policies and guidelines to mitigate these risks, including banning certain AI tools and conducting regular security audits.

Exercise 4

Do your own research about AI risks and security concerns. You may find some useful resources at https://www.prompt.security/resources/genai-risks-and-vulnerabilities.

Implications of AI on the Society

Implication 5: Changing in job market

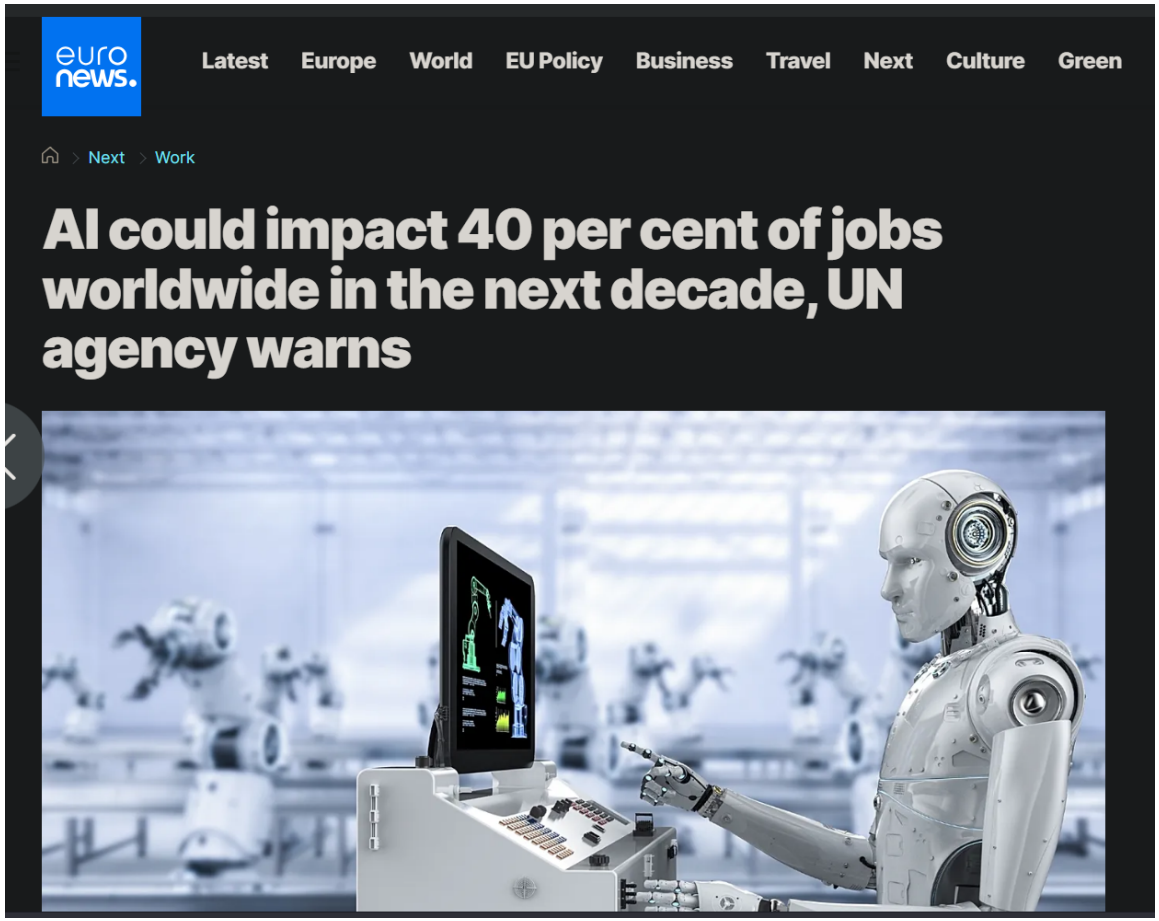


Figure: Euronews (2025): report on a UN agency warning that AI could impact 40% of jobs worldwide in the next decade. (https://www.euronews.com/next/2025/04/07/ai-could-impact-40-per-cent-of-jobs-worldwide-in-the-next-decade-un-agency-warns)

Implications of AI on the Society

Implication 5: Changing in job market

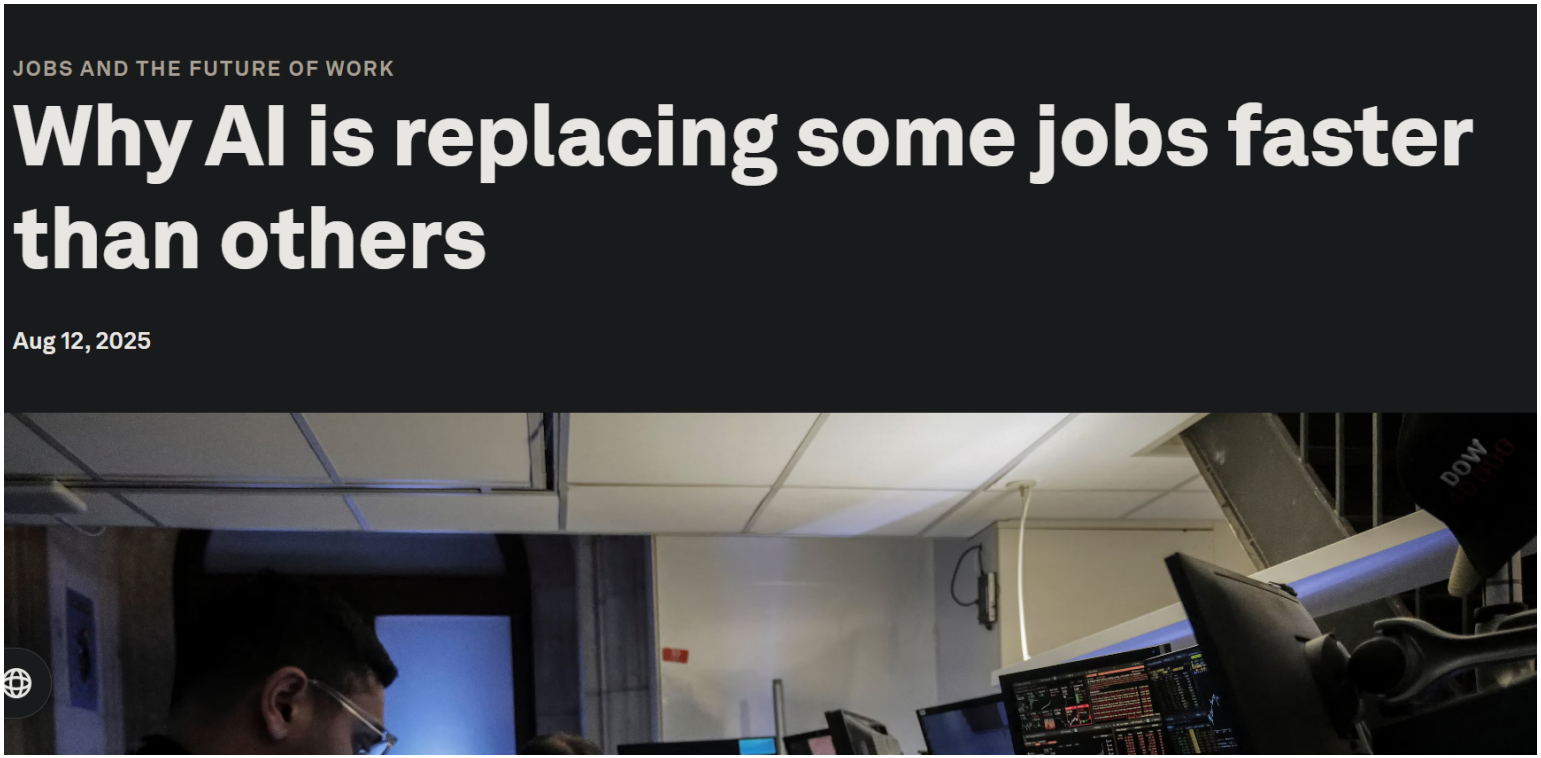


Figure: World Economic Forum (2025): report on how AI is expected to displace jobs and create new ones, emphasizing the need for reskilling. (https://www.weforum.org/stories/2025/08/ai-jobs-replacement-data-careers/)

Implications of AI on the Society

Implication 5: Changing in job market

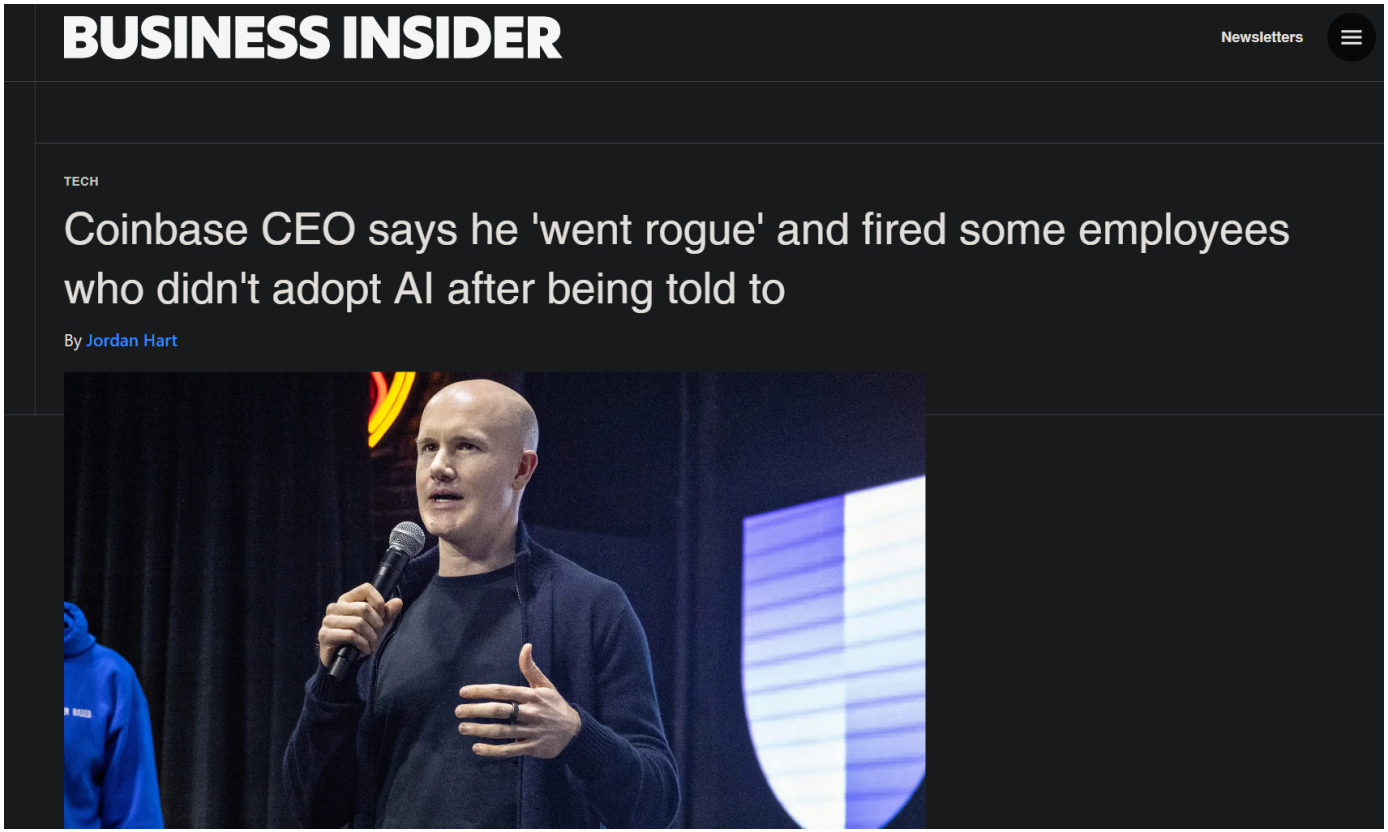


Figure: Business Insider (2025): report on how companies are increasingly expecting employees to leverage AI tools, with potential implications for job security and skill requirements. (https://www.businessinsider.com/coinbase-ceo-fired-employees-not-using-ai-tools-onboarding-2025-8)

Implications of AI on the Society

Implication 5: Changing in job market

- **Historical pattern:** automation tends to reallocate labor rather than simply eliminate jobs — routine tasks are replaced, while new occupations and industries emerge (e.g., manufacturing $\longrightarrow$ automation maintenance; clerical work $\longrightarrow$ information services).
- **Nature of new work:** growing demand for non-routine cognitive, creative, social, and technical skills (problem solving, communication, digital literacy, systems thinking).
  - It is likely that in the future, a larger fraction of the workforce will focus on research and development, and tasks that require creativity and human-to-human interaction. For example, see Abhinav Suri's nice essay on "Artificial Intelligence and the Rise of Economic Inequality" (https://abhinavsuri.com/blog/2017/aiinequality/)
- **Pace and scale:** advances in AI accelerate substitution across both blue- and white-collar roles, increasing the likelihood of long, career-spanning disruptions for many workers.

Implications of AI on the Society

Implication 5: Changing in job market

- **Vulnerable occupations (examples):** long-haul trucking and taxi driving, basic data entry and routine customer service, certain assembly and repetitive manufacturing tasks — timelines depend on technology, cost, and regulation.
- **Risks if unmanaged:** mass unemployment pockets, downward wage pressure, regional and demographic inequality, and reduced job quality for affected workers.
- **Recommended responses:** a combined strategy of education, lifelong learning, labor policy, and coordinated public-private action is needed to mitigate displacement risks and capture the opportunities of AI.

Implications of AI on the Society

Exercise 5

Find an online article written about the impact of AI to society. Write a brief review of the article you found. For example, you might use the following as a template:

- Title: [Title of the article]
- Author: [Author of the article]
- URL: [URL of the article]
- Summary: [Brief summary of the article]
- Comments: [Your comments on the article]
  - [How accurate is the AI-related statements in the article?]
  - [Any conclusions in the article is realistic? Why?]
  - [What are the potential implications of the article's findings?]

Exercise 6

How do you see AI affecting *you* in the future, both at work and in everyday life? Include both the positive and possible negative implications.

Implications of AI on the Society

Exercise 7 (Will the Terminator come?)

The Terminator is a fictional character from a 1984-movie by director James Cameron. The film depicts a dystopian future where AI (specifically a global AI-powered defense system called Skynet) becomes self-aware and threatens humanity. Do you think the scenario presented in the movie could ever become a reality? Why or why not?

Figure: The Terminator (1984) movie poster. Image taken from the Internet

Take-home Messages

- AI is a broad field with many definitions and interpretations. There is no single agreed-upon definition of AI.
- Currently, AI is at the weak stage (narrow AI, focusing on specific tasks and applications).
- The history of AI has seen cycles of optimism and setbacks, but recent advances in machine learning and deep learning have led to significant breakthroughs and widespread adoption.
- AI has profound societal implications: algorithmic bias; privacy risks (data leaks, re-identification, profiling); erosion of trust from deepfakes and misinformation; security vulnerabilities; and significant labour-market shifts that require reskilling and policy responses.
- Addressing these implications requires careful consideration of ethical, social, and economic factors, as well as proactive strategies to manage risks and opportunities.

References


Rich, Elaine (1983). *Artificial intelligence*. McGraw-Hill.

Turing, Alan M. (1950). “Computing machinery and intelligence.” In: *Mind* 59.236, pp. 433–460. DOI: 10.1093/mind/LIX.236.433.