AI Ethics and Privacy

ITAIA1-33

Eduvos

Your Education. Your Future.

Disclaimer

Please note that content made available on myLMS may deviate only slightly from what is covered in lecturer-led sessions, however, the content uploaded to myLMS represents all the material that you need to refer to in preparation for your assessments.

Week 2: Lesson 1
Introduction

Current State of AI

Learning Outcome	Assessment Criteria
2. Analyze the Present State of AI	2.1 Evaluate the defining characteristics and capabilities of contemporary AI. 2.2 Compare and contrast AI as a form of agency with traditional notions of human intelligence. 2.3 Critically assess the implications of AI's current state on various industries and domains.

What will be covered in Today's Lesson?

We will be learning about:

What is AI
Definitions

AI has numerous definitions;

- No single agreed-upon definition.

- Legg and Hutter (2007) listed 53 definitions for 'intelligence' and 18 for AI

- The number of definitions continues to grow (Russell and Norvig 2018)

Wikipedia's Tautology:

- Tautology - the saying of the same thing twice over in different words, generally considered to be a fault of style (e.g. they arrived one after the other in succession)

- Wikipedia defines AI as "intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans."

- This definition is seen as true but unhelpful

What is AI
Definitions Continued

Challenges in Defining AI;

- Lack of a standard definition can be problematic in ethical discussions.

- Endless debates on the definition often lead to frustration and no consensus

Comparison to Other Concepts:

- Important concepts like friendship, health, and love are also hard to define but understood intuitively.

- AI, like these concepts, is recognized when seen

Historical Reference:

- US Supreme Court Justice Potter Stewart's famous phrase *"I know it when I see it"* (1964) about obscenity is used to illustrate the intuitive understanding of AI

What is AI
Definitions Continued

Scientific Definition;

- Unlike precise scientific terms (e.g., triangle, planet), AI is not strictly definable
- AI is a generic term encompassing various disciplines, services, and products

AI as Family:

- AI is a family of related technologies with some common traits
- Disagreements among experts about the categorization and terminology within AI

Conclusion:

- AI can be understood through its applications and common traits
- Proposed Definition: Artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving. (McCarthy et al, 2006)

AI as a Counterfactual
AI Behaviour and Counterfactual Nature

Defining AI Through Behaviour;

- AI is defined as making a machine behave in ways that would be considered intelligent if a human did them.
- Focus is on *behaviour*, not *thinking*

Counterfactual Nature of AI:

- AI does not imply actual intelligence or consciousness.
- It's about outcomes, not internal processes

Turing's Viewpoint:

- Turing dismissed the question "Can machines think?" as meaningless.
- Proposed the Turing Test: if a machine's responses are indistinguishable from a human's, it passes

AI as a Counterfactual
Performance, Process and Outcomes

<u>Performance vs Process;</u>

- Success in task completion doesn't imply similarity in method or intelligence
- Example: A dishwasher vs. a human cleaning dishes—same result, different processes

<u>Fallacy of Outcome-Based Equivalence:</u>

- Identical outcomes do not mean identical agents or processes
- Mistaking performance for intelligence is misleading

AI as a Counterfactual
Brains vs Machines

Turing's 1951 BBC Broadcast:

- Suggested digital computers could be described as brains

Metaphorical vs Scientific Language:

- Describing brains as machines is metaphorically acceptable but scientifically imprecise

- Overly broad definitions dilute meaningful distinctions

The Machine Analogy Problem:

- If everything is "kind of" a machine, the term loses precision

- Important differences between brains and computers become obscured

AI as a Counterfactual
Logical Insight

Key Logical Insight:

- Same input and output do not imply same internal workings
- Example: Clean dishes don't reveal whether Bob or a dishwasher cleaned them

Conclusion:

- Turing's metaphorical framing is insightful but scientifically flawed
- The counterfactual definition of AI supports an *engineering* rather than *cognitive* approach

The Two Souls of AI
Engineering vs Cognitive

Dual Goals of AI Research:

- Reproductive AI: Replicates intelligent behavior using non-biological systems
- Productive AI: Aims to recreate the source of intelligence itself

Engineering Success – Reproductive AI:

- Example: Deep Q-network (2015) learned to play 49 Atari games using only pixel and score data
- Smart technologies now outperform humans in many tasks
- AI replaces human intelligence in effectiveness and efficiency

Cognitive Disappointment – Productive AI:

- Still far from replicating true human-like intelligence
- Systems like Watson and Deep Blue showcase engineering skill, not cognitive understanding
- John McCarthy criticized chess-playing AI as not true AI

The Two Souls of AI
Comparisons

Terminology vs Confusion:

- Various labels:
  - Weak vs. Strong AI
  - GOFAI vs. Nouvelle AI
  - Artificial General Intelligence (AGI)
  - Light vs. Full AI
- Misalignment of goals leads to confusion and debate

Speculation and Singularity:

- Confusion between engineering and cognitive AI fuels fears of superintelligence
- Much of the debate is intellectually entertaining but often unproductive

The Two Souls of AI
Comparisons Continued

AlphaGo vs AlphaFold:

- AlphaGo (2017) retired from Go to focus on broader scientific challenges
- AlphaFold 2 (2020) solved the protein folding problem—major breakthrough in biology

The Two Souls of AI
Outcomes Over Intelligence

Outcome Over Intelligence:

- AI is not about replicating biological intelligence
- It's about achieving results without human-like cognition

Current AI Capabilities:

- Machines today have the intelligence of a toaster—functional but not cognitive
- Example: "Printer not found" error despite the printer being present—annoying, but expected

Smart Without Being Intelligent:

- Artefacts can be smart (effective) without being intelligent (aware or understanding)
- Reproductive AI is a continuation of human intelligence by other means

The Two Souls of AI
Outcomes Over Intelligence Continued

Decoupling Intelligence from Performance:

- AI succeeds by separating task performance from intelligent behavior
- It thrives in domains where understanding and awareness are unnecessary

Replacing, Not Replicating Intelligence:

- Success comes from replacing human intelligence, not mimicking it
- Example: AlphaGo doesn't play like a human but still wins

Smart Without Being Intelligent:

- Artefacts can be smart (effective) without being intelligent (aware or understanding)
- Reproductive AI is a continuation of human intelligence by other means

Future Directions for AI

Gamification of Tasks:

- AI should evolve from winning games to gamifying tasks
- Anything that can be turned into a game becomes solvable by AI

AI vs AI:

- AI systems will increasingly compete with each other in complex, gamified environments
- Human observers may not fully understand these interactions—like enjoying a Bach performance

New Role for Human Intelligence:

- Humans will focus less on solving problems and more on:
  - Choosing which problems to solve
  - Understanding the purpose and implications
  - Managing trade-offs and consequences

Questions for Discussion
AI Usage and Impact

Discuss the following in your groups:

1. Which area of society you deem to have been impacted the most by the emergence and usage of AI?
2. How has it been impacted in terms of:
    1. Ethics?
    2. Usage of Data?
    3. Any other areas you feel are impacted

The Human Use of Humans as Interfaces
AI's Environments and Capabilities

AI Friendly Environments:
- The world is increasingly enveloped into a digitally friendly infosphere
- We now share our space with Artificial Agents (AAs), not just humans and nature

Smart but Not Intelligent:
- Machines lack true intelligence, understanding, or emotions
- Yet, syntactic technologies (based on data/statistics) deliver meaningful outcomes

AI's Practical Capabilities:
- AI can recommend books, diagnose illnesses, find deals, and more—without understanding
- Memory and algorithms outperform human intelligence in many tasks

The Human Use of Humans as Interfaces
AI's Environments and Capabilities Continued

Automation in Enveloped Environments:
- In data-rich domains (e.g., e-health, finance), decisions are often made automatically
- Machines execute tasks like scanning patients or trading bonds without human input

The Human Use of Humans as Interfaces
Humans as Part of the Mechanism

Human-Based Computation:
- Humans are used to perform tasks AI cannot yet handle
- Example: Amazon Mechanical Turk—humans complete HITs (Human Intelligence Tasks)
- This is called "Artificial Artificial Intelligence"—machines rely on hidden human input

Historical Analogy – The Mechanical Turk:
- 18th-century chess-playing automaton secretly operated by a human
- Modern systems similarly use humans behind the scenes to simulate intelligence

The Human Use of Humans as Interfaces
Humans as Part of the Mechanism Continued

<u>Humans as Interfaces in Advertising:</u>
- Customers are treated as interfaces between suppliers and bank accounts
- Free services are used to extract data for targeted manipulation
- AI tailors and optimizes advertising strategies using recommender systems
- Customers are nudged and influenced based on data-driven insights

<u>Means vs Ends:</u>
- Humans are increasingly treated as means, not ends—violating Kantian ethics (The end does not justify the means)
- This shift raises concerns about autonomy, dignity, and digital exploitation

What is AI?
Summation

Definition:
- Oxford English Dictionary:
  - *"The capacity of computers, or other machines, to exhibit intelligent behaviour."*

Key Characteristics:
- Mimics or exceeds human capabilities.
- Can think, learn, solve problems, make decisions, and even create

Why AI Matters:
- AI is a multi-use technology with applications across industries.
- It's transforming how we live, work, and interact

Why AI is Booming Now:
- Advanced algorithms
- Massive data availability
- Powerful computing resources

Algorithms
Brains of AI

Data
Fuel that drives AI

Computing Power
Machines behind AI
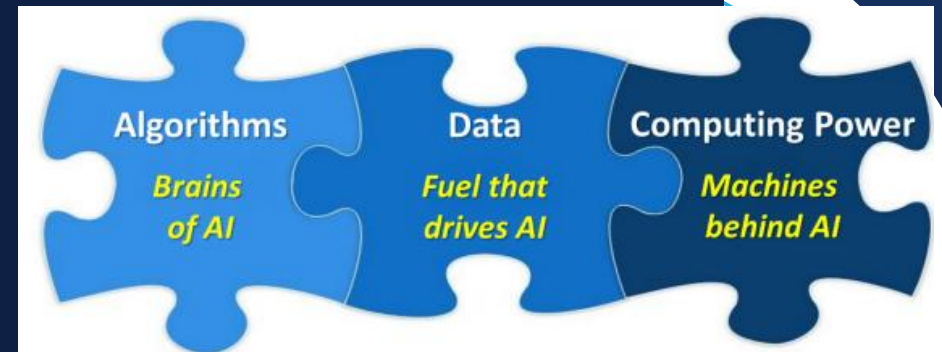
Activity

1. Watch the following YouTube Video to help yourself consolidate your knowledge for this lesson, and to prepare for the next:

https://youtu.be/bknUn7yMwNI?si=1RICnJ-jmGs-M0vU

2. Complete the Week 2 – 3.2. Practice Quiz on MyLMS