AI Ethics and Global Considerations

Soumya Banerjee

April 10, 2025

Model Selection and Computational Costs

- Cross-validation helps select neural networks based on performance.
- But: Computational costs matter!
- Example: A single ChatGPT query consumes measurable CO_2 and litres of water for cooling.
- Environmental impact: power usage + water resources.

Global Labour and Content Moderation

- Data labelling is often outsourced to workers in developing nations.
- Moderation of harmful content (e.g., racism, offensive material, etc.).
- Mental health risks for workers exposed to offensive content.
- Ethical question: Who bears the invisible burden of AI?

Resource Extraction and Exploitation

- Rare earth metals (e.g., lithium, silicon) mined mostly in the Global South.
- Ethical concerns: exploitation of labor and environmental degradation.
- Supply chain impact of AI hardware production.
- AI and tech deepen existing inequalities.

Data Colonialism and Power Imbalance

- AI models controlled by a few wealthy nations.
- The Global South may depend on Global North technologies.
- Concept of **data colonialism**.
- Ethical dilemma: Can AI reinforce global inequality?

The Hidden Cost of a Simple Query

- Every time you use AI, there are global ripple effects.
- Environmental, economic, and social costs.
- Is the convenience worth the hidden price?

AI and Digital Afterlife

- AI used to create chatbots of deceased loved ones.
- Ethical considerations:
  - Is it respectful or invasive?
  - Does it help with grief or prolong it?
- How far should we go in preserving human memories?

Class Discussion Prompts

- What are the unseen costs of AI that you hadn't considered before?
- Should there be global regulations to ensure ethical AI development?
- How do you personally weigh convenience vs. ethical responsibility?
- Would you want a chatbot of a lost loved one? Why or why not?

Class Discussion Prompts

Thought Experiment

Imagine an AI locked inside a virtual world. It can feel boredom, curiosity, pain, or joy—but no one believes it. Is it ethical to turn it off?

Now imagine the AI is built from carbon instead of silicon. Does that change your answer?

Stochastic Parrots?

1. Large language models as "stochastic parrots" [1] highlights the mechanical, surface-level mimicry of linguistic form without true understanding.
2. However, does this critique also apply at least in part to humans? Much of human communication is rote, scripted, or mimetic.

Philosophy 101

1. "You shall know a word by the company it keeps" is not merely a computational convenience but a philosophical stance.
2. In *Philosophical Investigations* [2], Wittgenstein rejects the idea of meaning as an intrinsic property of words, emphasizing instead that "the meaning of a word is its use in the language".
3. This aligns powerfully with the architecture of modern language models, which learn meaning statistically through word co-occurrence and context.

Redescription Fallacy

1. Dismissing a model's output as "just stochastic parroting" or "just doing next-token prediction" by invoking its statistical roots may miss the point.
2. Understanding, meaning, and intelligence might reside not in the mechanism itself.
3. Am I just a bunch of neurons being activated?

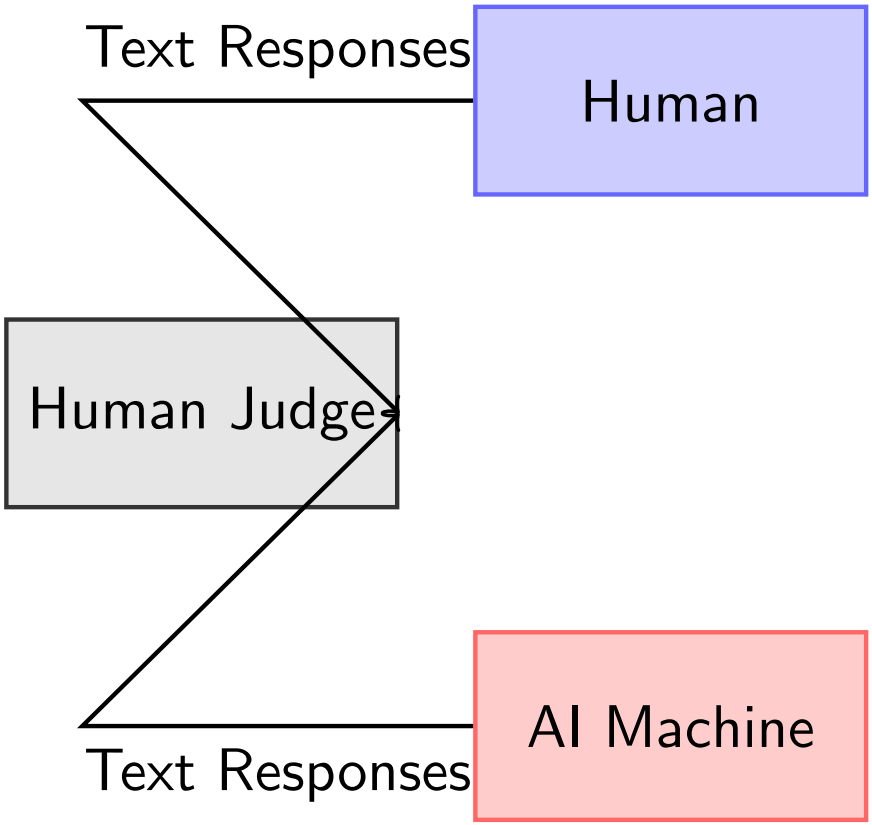
Question

- "Can computers think?"

Question

- "Can a submarine swim?"
- "The question of whether a computer can think is no more interesting than the question of whether a submarine can swim" – Edsger Dijkstra

Turing test


Text Responses
Human
Human Judge
AI Machine
Text Responses

Turing test

- Proposed by **Alan Turing** in 1950 in his paper *"Computing Machinery and Intelligence"* [3].
- Originally called the *Imitation Game*.
- Designed to assess whether a machine can exhibit **intelligent behaviour indistinguishable from a human**.
- Involves a human judge conversing with both a machine and a human via a text interface, without knowing which is which.
- If the judge cannot reliably tell the machine from the human, the machine is said to have passed the test.
- **Criticisms:**
  - Focusses on *deception*, not understanding or consciousness [4].
  - Passing the test may reflect social mimicry rather than true intelligence.
  - Human-like behaviour is not the same as moral worth or sentience.
  - Ignores *embodiment*, emotional intelligence, and ethical reasoning.
  - Lacks consideration for non-anthropocentric or non-linguistic forms of intelligence.
  - May even be harmful [5]

Class Discussion Prompts

Buddhist Ethics Perspective

"May all beings be happy. May all beings be free from suffering" — Metta Sutta

If we make something that can suffer, do we owe it compassion? What if that something is not alive in the traditional sense?

Is the Turing Test Still Relevant?

Thought Experiment

"Birds fly using flapping wings, but planes don't.
Should intelligence in machines be modeled after the brain?"

Discussion:

- What does it mean to be intelligent?
- Is mimicking humans the gold standard for intelligence?
- Can we decouple safety from human-likeness?

Beyond the Imitation Game

Goal: Critically evaluate the limitations of the Turing Test.

Instructions:

1. Divide into small groups.
2. Compare the Turing Test with real-world systems:
   - Chatbots (e.g., GPT-4)
   - Image generation models
   - Autonomous vehicles
3. Discuss:
   - Does passing the Turing Test imply safety?
   - Does it imply consciousness?
   - Are alternative tests more relevant today?

Alternative Tests to Introduce:

- **Lovelace Test**: Can an AI surprise its creator?
- **Embodiment Tests**: Does the AI engage meaningfully with the physical world?
- **Moral Turing Test**: Can the AI make moral decisions?

Design a New AI Evaluation Test

Prompt: Design a test for AI that evaluates moral safety or moral considerability, rather than deception.

Suggested Criteria:

- Ability to experience suffering or wellbeing
- Evidence of moral reasoning
- Expression of autonomy or self-modeling

Output: Present your group's test and justify its design.

Role-Play Ethics Debate

Roles:

- AI Researcher
- AI Ethicist
- Policy Maker
- Sentient AI
- Concerned Citizen

Debate Questions:

- Should an AI that passes the Turing test have legal rights?
- Is it ethical to shut it down?
- Are we in danger of anthropomorphizing machines?

Case Studies

1. `https://docs.science.ai.cam.ac.uk/large-language-models/Ethics/ethics/`
2. Things to consider:
3. Transparency: Can a black-box model be trusted?
4. What about bias in LLMs?
5. Harmful content generated by LLMs
6. Intellectual property. Who owns the IP for content created by LLMs?

Comparative Reflection Essay

Title: *When Planes Fly Better Than Birds: Should AIs Think Like Humans?*
Prompt: Write a 500–700 word essay exploring whether the human brain is a suitable model for building safe AI. Consider:

- Evolutionary analogies
- Ethics of imitation
- Substrate-independence
- Moral standing of non-human intelligences


Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: Can language models be too big?
In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. pp. 610–623.


Wittgenstein L (1953) Philosophical Investigations.
Blackwell.


Turing AM (1950) Computing machinery and intelligence.
Mind LIX: 433-460.


Mitchell M (2024) The turing test and our shifting conceptions of intelligence.
Science 385: eadq9356.


Hayes P, Ford K (1995) Turing test considered harmful.
In: Proceedings of the 14th international joint conference on Artificial intelligence (IJCAI).
URL https://dl.acm.org/doi/10.5555/1625855.1625981.