Introduction to Artificial Intelligence - Homework Assignment 1: Introduction/Agents

October 22, 2023

Group 6 - Castagnotto Alessandro, Coceani Elisa, Majer William, Mingrone Tommaso, Secci Marco

1 Exercise

Read Alan Turing's original paper on AI (Turing, 1950). In the paper, he discusses several objections to his proposed enterprise and his test for intelligence. Which objections still hold weight? Are his refutations valid? Can you think of new objections arising from developments since he wrote the paper? In the paper, he predicts that, by the year 2000, a computer will have a 30 chance of passing a five-minute Turing Test with an unskilled interrogator. What chance do you think a computer would have today? In another 25 years?

1.1 Introduction

Alan Turing's 1950 paper, "Computing Machinery and Intelligence," marked a pivotal moment in the history of artificial intelligence (AI). In this article, Turing proposed the idea of artificial intelligence and introduced the concept of the Turing Test, a measure of a machine's ability to exhibit intelligent behavior indistinguishable from that of a human. While Turing's work laid the foundation for artificial intelligence research, it also sparked objections and debates that continue to shape the field today. We will explore the objections raised in the Turing era, the validity of his rebuttals, and the new challenges facing artificial intelligence in the modern world.

1.2 Historical Objections to AI

Theological Objection One of the earliest objections to AI posited that machines could not possess true intelligence because they lacked essential human attributes, such as a soul or consciousness. Critics argued that the essence of human thought and understanding could never be replicated by machines. Over time, this objection has lost prominence as AI research shifted its focus from

metaphysical questions to practical functional capabilities. The emphasis has shifted toward understanding how machines can exhibit intelligent behaviors, regardless of whether they possess consciousness or a soul.

Mathematical Objection The mathematical objection questioned whether machines could genuinely "think" in any well-defined sense. Critics argued that the very notion of "thinking" by machines was problematic and that Turing's work faced limitations in this regard. Turing responded by emphasizing that the definition of "thinking" should be grounded in observable behaviors, a stance that continues to be central to discussions about AI and consciousness. While this objection occasionally resurfaces in contemporary AI debates, it has evolved with a greater focus on the relationship between AI and consciousness.

The Argument from Informality of Behavior The objection from the informality of behavior posited that human behavior is too complex and informal to be replicated by machines following rigid rules. This argument raised concerns about whether AI systems could ever truly understand or mimic human thought processes. Today, we have witnessed AI systems performing complex tasks, but they often do so differently from humans. The question of whether these systems genuinely understand human thought processes or merely mimic them remains open, highlighting the ongoing relevance of this objection.

The Argument from Extrasensory Perception Turing briefly mentioned the argument from extrasensory perception, which posited that paranormal phenomena might challenge scientific understanding. However, this objection is largely unrelated to AI and has had no significant impact on the field.

1.3 Turing's Refutations and Their Validity

Turing provided reasoned responses to these historical objections. He argued that theological concerns were subjective and not relevant to the practical goal of AI research. In addressing the mathematical objection, he asserted that the definition of "thinking" should be based on observable behaviors, setting the foundation for the field's empirical focus. Turing's refutations have remained persuasive to many in the AI community, guiding contemporary research in the discipline.

1.4 New Challenges in AI

The Objection of Consciousness While Turing primarily focused on intelligent behavior, discussions about machine consciousness have gained prominence in contemporary AI discourse. Some argue that true AI should encompass consciousness, sparking debates about whether machines can possess subjective experiences. The question of consciousness remains a subject of ongoing philosophical and scientific inquiry, influencing the trajectory of AI research.

The Threat to Human Employment As AI systems become increasingly capable, concerns about their impact on employment and the economy have grown. Critics worry that AI could lead to widespread job displacement, challenging traditional labor markets and necessitating new approaches to workforce development and job creation.

The Ethics and Bias Objection The rise of machine learning and deep learning has brought concerns about AI bias and ethical decision-making to the forefront. Objections center on AI systems reinforcing societal biases or making ethically questionable decisions. Ethical considerations have become integral to AI development, demanding increased scrutiny of AI technologies' societal implications.

1.5 Turing's Prediction for 2000 and Beyond

Turing predicted that by the year 2000, a computer would have a 30 chance of passing a five-minute Turing Test with an unskilled interrogator. By 2023, we have indeed witnessed significant advancements in AI and chatbot technologies, such as OpenAI's GPT-3, which can produce human-like text responses in various contexts. These models can pass simple Turing Tests with a reasonably high success rate, although they are not yet at the level of complete human-like understanding.

1.6 Speculating the Future of AI

Predicting the chance of a computer passing a Turing Test in another 25 years remains highly speculative. It depends on future breakthroughs in AI, the development of more sophisticated models, and our ability to address challenges like consciousness, ethics, and bias in AI systems. Given the rapid pace of AI advancements, it is plausible that AI systems will continue to improve, but it remains uncertain when, or if, they will pass a comprehensive Turing Test with flying colors.

1.7 Conclusion

Alan Turing's pioneering work laid the groundwork for AI research and continues to inspire contemporary discussions on machine intelligence. While historical objections have evolved, new challenges have arisen, and Turing's predictions have partially materialized, the field of AI remains a dynamic and multidisciplinary domain. As AI researchers navigate these challenges, the enduring legacy of Turing's contributions serves as a guiding beacon in the pursuit of artificial intelligence.

2 Exercise

Many researchers have pointed to the possibility that machine learning algorithms will produce classifiers that display racial, gender, or other forms of bias. How does this bias arise? Is it possible to constrain machine learning algorithms to produce rigorously fair predictions? **Biases in machine learning classifiers can emerge from several factors. Here are some of the main causes:**

1. Bad training data: If the dataset used to train a machine learning algorithm contains racial, gender, or other biases, the algorithm can learn and reflect those biases. For example, if historical data contains gender or race discrimination in the past, the algorithm can learn to reproduce such discrimination.

2. Data selection: The selection of training data can influence the presence of bias. If the data used to train a model is collected in ways that reflect existing biases or is incomplete, the model may reflect those biases.

3. Discriminating features: The features selected for model training can lead to discrimination. For example, if you use characteristics that are related to racial or gender factors, the model can learn to make predictions based on those characteristics, even if it is not appropriate.

To mitigate biases in machine learning classifiers, several strategies can be adopted:

1. Data Cleaning: It is important to carefully examine your training data to identify and remove existing biases. This may require removing problematic data or reducing the importance of certain features.

2. Balanced representation: Make sure that the training dataset equally represents the different categories in play, so as to avoid overfitting or underfitting of some categories.

3. Fairness-aware learning: Use fairness-aware machine learning approaches that explicitly seek to reduce bias in models and produce fairer and more balanced predictions.

4. Continuous auditing and evaluation: Continuously monitor machine learning models in production to identify and correct any emerging biases. This may include the use of fairness metrics to evaluate model performance.

5. Regulation and legislation: Authorities and organizations can introduce regulations and guidelines to ensure justice and fairness in the use of machine learning algorithms.

While it is not possible to completely eliminate all biases, it is possible to reduce them and manage them to produce more rigorously fair and balanced forecasts. Awareness and commitment to addressing this issue are key to developing and using AI responsibly.