

Optimized Rationality and the concept of Explainability in Artificial Intelligence

Peters Unekwu Onyilo¹

Department of Artificial Intelligence, Axiomata Technologies Limited.

Abstract

This academic inquiry delves into the intricate terrain of artificial intelligence (AI) through the lens of the widely acclaimed Generative Pretrained Transformer 3 (GPT-3). Focused on a specific computational task—calculating age relative to one's elder sister—the study showcases GPT-3's computational acuity and its adept articulation of the logical implausibility of one being older than their elder sibling.

Central to this discourse is the introduction of the concept of optimized rationality, elucidating its pivotal role in fortifying the computational system's capacity to navigate cognitive dissonance within acceptable thresholds. This conceptual framework not only economizes computing resources but also integrates ethical considerations into the explication of computational results.

Keywords: Artificial Intelligence, GPT-3, Computational Logic, Optimized Rationality, Cognitive Dissonance, Controllability, Explainability, Transparency, Ethical AI, Computational Ethics.

Introduction

An engagement with the most popular Artificial Intelligence entity for the public which is the Generative Pretrained Transformer 3 chat engine. I asked some seemingly obvious question to compute my age if twice as old as my elder sister. It sprang into action and outlined the mathematical process. The underlying logic is I cannot be older than my elder sister by any means. The concept of optimized rationality comes with enhancing ability of computational system to gravitate towards tolerable level of cognitive dissonances. It saves a lot of computing time and level of ethics into how results of computations could be explained.

Explainability enables the system to outline basic behaviour in a way that users could see reason for output and better align with such. How rational is the level of openness driving the computational change for Artificial Intelligence? The ethics of computing and Artificial Intelligence growth will determine how these questions will be answered in the future. [1] How we answer the question will also determine the maximum benefits and self-awareness we seek for computational system to adopt. Mapping human mind and understanding how it works will be no more ethical than dissecting how we engage computer and its subset to help us think. Controllability in AI refers to the degree to which an AI system can be directed, guided, or controlled by humans or other external mechanisms. Here's why controllability matters:

A rational mind is capable of creating optimal output by having traces or vivid track of events, processes that generates meaningful output. Leaving awful moment of rationality in ability to detail out explanations deepens cognitive dissonance.

Ethical and Responsible AI: Controllability is essential for ensuring that AI systems adhere to ethical and responsible behavior. It allows for the implementation of safeguards and constraints to prevent AI from making harmful or biased decisions.

Safety and Compliance: Controllability is crucial for safety and compliance in AI applications. For example, in autonomous vehicles or medical AI systems, the ability to intervene and control the system's actions is vital to prevent accidents or ensure regulatory compliance. [2]

Explainability and Transparency: Controllable AI systems are often more transparent and explainable. Users can better understand and trust AI decisions if they have the means to control and influence those decisions.

Adaptability and Learning: Controllability allows humans to guide AI systems and provide feedback. This is valuable for the learning and adaptation of AI models, as human feedback can help train the AI to make better decisions.

Bias Mitigation: In situations where AI bias is a concern, controllability enables users to modify or correct AI decisions to reduce bias and promote fairness.

User Experience: Controllability can enhance the user experience by giving users the ability to customize AI interactions according to their preferences and needs.

Human-AI Collaboration: In collaborative environments, controllability is crucial for defining the roles and responsibilities of humans and AI systems, ensuring productive cooperation.

The different model employed in integration of Explainability in Artificial Intelligence and next generation software is gathering interest from ethical point of view. Transparency will be measured by the indices of Explainability. If you can explain how got it, then you stole it.

Conclusion

In conclusion, our exploration into the realm of artificial intelligence, as exemplified by the usage of Generative Pretrained Transformer 3 (GPT-3), has illuminated multifaceted dimensions that intersect computational logic, optimized rationality, and ethical considerations. The specific task of age calculation, framed within the context of familial relationships, served as a practical canvas for GPT-3's computational capabilities. The introduction of optimized rationality, as a guiding principle in computation, unveils a strategic approach that not only streamlines computational efficiency but also integrates ethical considerations. This integration is crucial in navigating cognitive dissonance and aligning AI systems with acceptable ethical thresholds. As we delve deeper into the ethical implications of AI, these considerations become increasingly pivotal in shaping the trajectory of AI development.

References

- [1] **Adadi, A., & Berrada, M.** (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138-52160.

- [2] **Wang, H., Liu, D., Zhang, C., & Ding, C.** (2019). An Overview on Coordinated Control of Multi-Agent Systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(12), 2722-2737.