

Интеграция аристотелевской этики в разработку искусственного интеллекта: концепция разработки, управляемой этикой (EDD)

Научный руководитель – Ларионов Игорь Юрьевич

Ярочкин Дмитрий Андреевич

Аспирант

Санкт-Петербургский государственный университет, Институт философии,
Санкт-Петербург, Россия

E-mail: ya.yarochkin@yandex.ru

Интеграция аристотелевской этики в разработку искусственного интеллекта: концепция разработки, управляемой этикой (EDD) Integrating Aristotelian Ethics into the Development of Artificial Intelligence: A Framework for Ethic-Driven Development (EDD)

Аннотация В статье рассматривается интеграция аристотелевской этики в разработку искусственного интеллекта (ИИ) через создание концепции разработки, управляемой этикой (EDD). Исследование основывается на учении Аристотеля о четырёх причинах, где Благо рассматривается как цель существования. В работе предлагается подход, при котором этическое тестирование ИИ проводится на каждом этапе его разработки. Результаты показывают, что внедрение этического компонента в ИИ на всех стадиях, от сбора данных до внедрения приложений, способствует долгосрочным выгодам и моральной устойчивости.

Abstract This paper explores the integration of Aristotelian ethics into the development of artificial intelligence (AI) through the creation of an Ethic-Driven Development (EDD) framework. The study is grounded in Aristotle's doctrine of the four causes, where Good is considered the purpose of existence. The proposed approach involves ethical testing of AI at every stage of its development. The results demonstrate that embedding ethical components in AI from data collection to application deployment contributes to long-term benefits and moral sustainability.

Ключевые слова

Искусственный интеллект, аристотелевская этика, разработка, управляемая этикой (EDD), Этическое тестирование

Keywords Artificial Intelligence, Aristotelian Ethics, Ethic-Driven Development (EDD), ethical Testing

Введение

Учение Аристотеля о четырёх причинах ставит Благо в качестве причины бытия [4]. Благо — это то, к чему стремятся все вещи [5], и в этом отношении возникает вопрос, должен ли искусственный интеллект также стремиться к Благу. Размышляя о душе, Аристотель выделяет несколько её частей, из которых только рациональная часть может быть отделена [3]. Термин "душа" редко используется в современном дискурсе из-за его загадочной природы. Однако, обсуждая искусственный интеллект, именно это глубокое понятие нам и нужно рассматривать.

Методы

В нашей Лаборатории цифровой философии [6] мы разрабатываем подход, основанный на разработке через тестирования (EDD). Этот подход предполагает необходимость этического тестирования искусственного интеллекта на каждом этапе его разработки, без продвижения, если не разработан подходящий тест.

Результаты

В отличие от традиционного подхода в безопасности ИИ который противоречит учению Философа о причинах, так как мы пытаемся создать объект, который изначально не стремится к благу, и только потом учим его не причинять вреда наш основан на том, что , этический компонент должен быть принципом при создании искусственного интеллекта на всех этапах, от сбора данных до внедрения приложений.

Обсуждение

Отсутствие души в искусственном интеллекте не исключает возможности этики. Если этика понимается как стремление к благу, то искусственный интеллект, как и любой другой объект в мире, подвержен этому стремлению. Согласно Аристотелю, это стремление заложено в самом объекте, а не во внешней причине. Поэтому этика должна быть заложена в самом объекте, а не просто навязана извне каким-либо этическим институтом.

Заключение

Этот подход не только морален, но и выгоден в долгосрочной перспективе, так как концепция блага выходит за рамки экономической выгоды или мимолётного удовольствия [2] и связана с долгосрочными выгодами.

Источники и литература

- 1) Brennan-Marquez K., Henderson S. E. Artificial intelligence and role-reversible judgment //J. Crim. L. & Criminology. – 2019. – Т. 109. – С. 137.
- 2) Sparks J., Wright A. T. Human-centered AI //Divus Thomas. – 2023. – Т. 126. – №. 2. – С. 200-218.
- 3) COVOTTI A. RD Hicks, Aristotle De anima, with translation, introduction and notes (Book Review) / /Rivista di Filologia e di Istruzione Classica. – 1909. – Т. 37. – С. 455.
- 4) Aristotle's Metaphysics. Trans. Hippocrates Gorgias Apostle. Bloomington: IndianaU. Press, 1966.
- 5) The Nicomachean Ethics. – Penguin UK, 2020.
- 6) <https://philosophy.spbu.ru/ru/609-laboratoriya-tsifrovoj-filosofii.html>