УДК [101.1+17.02]:004.8
DOI: 10.36979/1694-500X-2024-24-2-57-61

ЭТИЧЕСКИЕ И ФИЛОСОФСКИЕ ПРОБЛЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Н.И. Осмонова, А.В. Иванская

Аннотация. Рассматривается возрастающая значимость решения этических и философских проблем в сфере искусственного интеллекта, а также проводится анализ причин, затрудняющих проблему урегулирования этических и правовых норм в данной области. Несмотря на активную разработку на государственном и международном уровнях соответствующих законов, существующие противоречия ещё не решены. В этой связи акцентируется внимание на важности философского осмысления искусственного интеллекта как универсального способа видения современной ситуации и новых тенденций развития. На основе изучения различных точек зрения на данную проблему обосновывается возможность снятия сложившихся между человеком и искусственным интеллектом антиномий путём их коэволюции, направленной на решение противоречий между стремительным развитием новых технологий и обеспечением безопасности человека.

Ключевые слова: искусственный интеллект; этические проблемы искусственного интеллекта; философское осмысление искусственного интеллекта; коэволюция.

ЖАСАЛМА ИНТЕЛЛЕКТТИН ЭТИКАЛЫК ЖАНА ФИЛОСОФИЯЛЫК КӨЙГӨЙЛӨРҮ

Н.И. Осмонова, А.В. Иванская

Аннотация. Макалада жасалма интеллект чөйрөсүндөгү этикалык жана философиялык маселелерди чечүүнүн өсүп бара жаткан актуалдуулугу каралат, ошондой эле бул жаатагы этикалык жана укуктук ченемдерди жөнгө салууну кыйындаткан себептерге талдоо жүргүзүлөт. Мамлекеттик жана эл аралык деңгээлдерде тиешелүү мыйзамдардын жигердүү иштелип чыккандыгына карабастан, учурдагы карама-каршылыктар чечиле элек. Ушуга байланыштуу заманбап кырдаалды жана өнүгүүнүн жаңы тенденцияларын көрүүнүн универсалдуу жолу катары жасалма интеллектти философиялык жактан түшүнүү маанилүүлүгүнө басым жасалат. Бул маселе боюнча ар кандай көз караштарды изилдөөнүн негизинде, жаңы технологиялардын тез өнүгүшү менен адамдын коопсуздугун камсыз кылуунун ортосундагы карама-каршылыктарды чечүүгө багытталган адамдар менен жасалма интеллекттин ортосунда пайда болгон антиномияларды алардын коэволюциясы аркылуу жоюу мүмкүнчүлүгү негизделет.

Түйүндүү сөздөр: жасалма интеллект; жасалма интеллекттин этикалык маселелери; жасалма интеллекттин философиялык түшүнүгү; коэволюция.

ETHICAL AND PHILOSOPHICAL PROBLEMS OF ARTIFICIAL INTELLIGENCE

N.I. Osmonova, A.V. Ivanskaya

Abstract. The article discusses the growing importance of solving ethical and philosophical problems in the field of artificial intelligence, and also analyzes the reasons that complicate the settlement of ethical and legal norms in this area. Despite the active development of relevant laws at the state and international levels, existing contradictions have not been resolved. In this regard, attention is focused on the importance of philosophical understanding of artificial intelligence as a universal way of seeing the modern situation and new development trends. Based on the study of various points of view on this problem, the possibility of removing the antinomies that have developed between humans and artificial intelligence, through their co-evolution, aimed at solving the contradictions between the rapid development of new technologies and ensuring human security, is substantiated.

Keywords: artificial intelligence; ethical problems in the field of artificial intelligence; philosophical understanding of artificial intelligence; coevolution.

Ввиду стремительного развития новых технологий можно подумать, что человек XXI века преодолел все те проблемы, которые ставили его дальнейшее развитие под угрозу. Замена физического труда машинным, интеллектуального – программным; существенное увеличение объёма свободного времени, которое можно потратить как на отдых, так и на выполнение большего числа задач, – разве не об этом мечтали наши предки, пытаясь заглянуть в неизведанное будущее? Но при таком рассмотрении современной ситуации игнорируются влияние каждодневного нахождения человека в информационной среде, его попытки постоянной актуализации получаемых данных, а также изменение и самой формы бытия человека: его существование, неразрывно связанное с индивидуальными качествами как субъекта деятельности, заменяется синтезом человеческого и искусственного. В результате известная с античности дихотомия "фюсис – технэ" теряет свою актуальность, и на первый план выходит изучение не только неразрывного влияния человека и его творений, сколько, ввиду интеграции в нашу жизнь искусственного интеллекта (далее – ИИ), анализ последним оснований нашей человечности. Но какие угрозы в данной ситуации готовит нам научный прогресс? И почему именно ИИ становится краеугольным камнем рассмотрения этических проблем современности?

Для ответа на эти вопросы следует определиться с самим понятием "искусственный интеллект". Как известно, на сегодняшний день "сфера применения ИИ достаточно обширна: образование, медицина, охрана окружающей среды, экономика, вооружение и др.", при этом ИИ может представлять собой как отдельную программу или аппарат, так и комплекс алгоритмов или даже теоретических наработок в рамках данной темы [1].

Так что же в таком случае необходимо понимать под ИИ? Из-за несводимости ИИ к частным понятиям, а также его абстрактности учёные считают уместным считать ИИ "зонтичным" понятием и определять его как "научное направление, в рамках которого ставятся и решаются задачи аппаратного или программного моделирования тех видов человеческой деятельности, которые традиционно считаются интеллектуальными" [2, с. 38–39]. То есть под понятием "искусственный интеллект" стоит понимать не конкретную программу или же машину, работающую по определённым алгоритмам, а "комплекс технологических решений, имитирующих когнитивные процессы человека" [3, с. 2].

Но если ИИ лишь имитирует деятельность человека, то почему такое большое внимание уделяется этическим проблемам, возникающим в данной области? Ведь можно подумать, если ИИ – инструмент, используемый в ходе человеческой деятельности и не более, то этика должна ограничиваться регулированием человеческих отношений и никак не затрагивать работу машины.

На самом деле ситуация не так проста, и этому есть две причины. Во-первых, из-за стремительного развития ИИ. Ещё в середине прошлого века интерес к ИИ преимущественно ограничивался научной фантастикой и инженерными исследованиями. Сегодня этот термин применяется не только относительно долгосрочных перспектив по моделированию человеческого интеллекта, но и стремительно развивающихся технологий (таких, как сложные нейронные сети, использующие огромные объёмы данных), которые всё больше влияют на финансы, транспорт, здравоохранение, национальную безопасность, рекламу и социальные сети, а также множество других областей [4].

Таким образом, даже несмотря на то что сегодняшний ИИ является "слабым" и направлен на решение конкретных задач в той или иной области, зависящих от определённых моделей поведения, это не препятствует ему становиться всё более значимым компонентом нашей жизни. Чат-GPT, голосовые помощники по типу "Алиса", нейронные сети в сфере искусства, создающие неповторимый продукт, – все эти приложения дают нам колоссальную экономию времени и сил, но что делать, если ИИ "ошибётся"? Кто тогда будет нести ответственность за уникальность и достоверность полученной информации? И не значит ли это, что "технологии ИИ имеют огромный потенциал не только для развития научно-технического прогресса, но и для злоупотребления ими или некорректного их использования?" [1].

Кроме этого, встаёт вопрос и о создании “сильного” ИИ, способного к саморазвитию и совершенствованию своих систем, универсального подхода к выполнению любой задачи и, по сути, имеющего искусственное сознание. Как быть здесь? Не будет ли даже само существование данного субъекта противоречить утвердившимся этическим нормам?

Во-вторых, из-за негативного влияния друг на друга человека и ИИ. С одной стороны, это находит своё проявление как в антропоморфизации строго выверенных алгоритмов, так и в роботизации самого человека. В первом случае имеет место популярное заблуждение, когда человек за каждым ответом системы начинает видеть проявление осознанности или даже чувств. Результатом второго становится автоматизация жизни человека, что чаще всего находит своё отражение в следовании определённым моделям поведения, исключающим индивидуальный выбор. С другой стороны, ИИ тоже подвергается прямому воздействию человека, приобретая от него значительную долю субъективности. Можно подумать, как такое возможно, если ИИ лишь оперирует полученными данными и, значит, должен быть непредвзятым? Здесь большую роль играет человеческий фактор, заключающийся как в случайном, так и умышленном допущении непроверенной или искажённой информации для использования ИИ. В результате получается так, что, когда ИИ выполняет задачу, поставленную перед ним пользователем, он невольно может привести его к ложным выводам. Но одно дело, когда такие “ошибки” могут произойти при нахождении ответа в поисковой системе, и совсем другое, если на кону, например, жизнь человека, находящегося в критическом состоянии в больнице. Что тогда делать? Кого считать виновным? И как предотвратить появление аналогичных ситуаций?

На подобные вопросы в сфере ИИ обратили внимание такие известные учёные, как А. Азимов [5] и Дж. Вейценбаум [6], среди наших современников – Р. Курцвейл [7] и Юдковский [8]. Они первыми стали разрабатывать морально-этическую проблематику искусственного интеллекта, необходимую для решения непростых дилемм.

Самым известным результатом такой работы считаются сформулированные ещё в прошлом веке А. Азимовым 3 закона робототехники:

1) робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинён вред;
2) робот должен повиноваться всем приказам, которые даёт человек, кроме тех случаев, когда эти приказы противоречат Первому закону;
3) робот должен заботиться о своей безопасности в той мере, в которой это не противоречит Первому или Второму законам [9].

Позже он вывел так называемый Нулевой закон, который объединил в себе все вышеприведённые: “Робот не может причинить вред человечеству или своим бездействием допустить, чтобы человечеству был причинён вред” [5, с. 486].

Безусловно, для своего времени А. Азимов вывел основополагающие принципы, на которых строились как этические, так и правовые нормы разработки и использования ИИ. Но применимы ли сегодня законы робототехники к ИИ, особенно учитывая всё большее его вовлечение в различные сферы нашей жизнедеятельности? Как, например, оценивать использование ИИ в военных целях, где его и действие, и бездействие приведут к человеческим жертвам? Быть может, дать ИИ свободу, чтобы он самостоятельно смог разрешить подобного рода противоречия? Но не окажется ли под угрозой сам человек, как наиболее непредсказуемое, а значит, опасное существо планеты Земля?

Именно такого рода неразрешимые дилеммы привели к активной разработке не только на государственном, но и на международном уровнях универсальных принципов по регулированию правовых и этических норм в сфере ИИ. Так, в 2020 году был подписан документ – Call for the AI Ethics, – направленный на недопущение замены человеческой деятельности технологическими инновациями [10]. А в 2021 году ЮНЕСКО была принята первая всеохватывающая конвенция, посвящённая рекомендациям в области этики ИИ, которая была поддержана 192 странами мира [11].

Стоит отметить, что, несмотря на попытку целостного рассмотрения современной ситуации в сфере ИИ, данные документы сконцентрировали своё внимание на юридической стороне вопроса. Поэтому для более глубокого понимания взаимоотношений человека и ИИ, а также выходящих из этого перспектив развития, будет уместно обратиться к философии.

В философском понимании данной проблемы можно выделить три основные точки зрения:

ИИ как субъективный инструмент. Так как "только люди (а не машины) являются конечным источником и определителем ценностей, от которых зависит любой искусственный интеллект", технологизация морали последнего является невозможной [3, с. 2]. Поэтому "этико-философская интерпретация ИИ предполагает в первую очередь дилемму ответственности, раскрывающуюся, с одной стороны, с точки зрения ответственности разработчиков и владельцев ИИ и, с другой – с точки зрения его потребителей [12]".

ИИ как объективный инструмент. Работая на большом массиве данных и не имеющий чувств и предпочтений ИИ является более объективным инструментом принятия решений, к тому же лишённым ошибок в связи с так называемым человеческим фактором [13]. При этом обращается внимание на несовершенство работы алгоритмов и качества обучения систем ИИ, что, однако, может найти своё решение в будущем.

Коэволюция человека и ИИ. Решая вопрос о строгом разграничении сфер влияния человека и ИИ, российские учёные Н.Н. Кожевников и В.С. Данилова приходят к тому, что "только взаимодействие человека с искусственным интеллектом, их коэволюция могут обеспечить устойчивое и оптимальное развитие того и другого" [14]. При этом предлагаются следующие принципы работы ИИ, а именно:

1) он должен быть автономным образованием, обладающим значительной устойчивостью по отношению к внешним воздействиям;
2) искусственный интеллект должен быть не просто исполнителем чьей-то воли, но и созидателем, элементом творческой активности;
3) он должен всегда защищать человека;
4) он не должен разрушать окружающую среду, а должен всегда защищать её [14].

Бесспорно, на сегодняшний день выполнение вышеприведённых принципов может считаться фантастикой как в силу возможностей технического прогресса, так и в связи с необходимостью возникающего парадокса о соотношении свободы ИИ и безопасности человека. Но, даже несмотря на это, понимание значимости общего развития вместо строгой дихотомии выглядит разумным выходом в свете стремительного развития новых технологий. Но, если в будущем ИИ станет не только инструментом в наших руках, что будет тогда с человеком? Является ли человеческое чисто человеческим или моральный закон – это нечто большее, стоящее на страже гармонии и созидания?

Безусловно, все эти вопросы остаются открытыми на сегодняшний день, а пока, подводя итоги, стоит отметить возрастающую значимость решения этических проблем, связанных с искусственным интеллектом. Этому способствуют как стремительное развитие новых технологий, так и расширение областей применения ИИ. И хотя на сегодняшний день разрабатываются различные международные соглашения, направленные на регулирование нормативно-правовых вопросов в данной сфере, однако невозможно полностью предусмотреть возникшие противоречия. В философском же осмыслении наиболее разумным подходом является не ограничение ИИ или возведение его в абсолют, а идея коэволюции ИИ и человека. Эта идея направлена на решение противоречий между развитием новых технологий и обеспечением безопасности человека, а также выработку таких объективных моральных принципов, которые имели бы огромное значение для дальнейшего гармоничного развития того и другого.

Поступила: 28.12.23; рецензирована: 15.01.24;
принята: 17.01.24.

Литература

1. *Афанасьева Ж.С.* Этические аспекты применения технологий искусственного интеллекта / Ж.С. Афанасьева, А.Д. Афанасьев // Информационные технологии и математическое моделирование в управлении сложными системами. 2022. № 3 (15). С. 24–32.

2. *Аверкин А.Н.* Толковый словарь по искусственному интеллекту / А.Н. Аверкин, М.Г. Гаазе-Рапопорт, Д.А. Поспелов. М.: Радио и связь, 1992. 256 с.
3. *Дзялошинский И.М.* Когнитивные процессы человека и искусственный интеллект в контексте цифровой цивилизации / И.М. Дзялошинский. М.: Ай Пи Ар Медиа, 2022. 583 с. URL: https://publications.hse.ru/pubs/share/direct/575368147.pdf (дата обращения: 08.11.2023).
4. *Дядченко А.А.* Этический и правовой аспекты регулирования искусственного интеллекта / А.А. Дядченко, И.И. Карташов // Ius publicum et privatum. 2023. № 2 (22). С. 34–44.
5. *Asimov I.* Foundation and Earth / I. Asimov. New York: Spectra, 2004. 528 p.
6. *Kurzweil R.* The Singularity Is Near: When Humans Transcend Biology / R Kurzweil. New York: Penguin Books, 2005. 672 p.
7. *Weizenbaum J.* Computer Power and Human Reason: From Judgement to Calculation / J. Weizenbaum. New York: W.H. Freeman & Company, 1976. 300 p.
8. *Yudkowsky E.* Creating Friendly AI 1.0: The Analysis and Design of Benevolent Goal Architectures / E. Yudkowsky. San Francisco: The Singularity Institute, 2001. 278 p.
9. *Азимов А.* Хоровод [пер. с англ.] / А. Азимов. URL: https://asimovonline.ru/short-stories/khorovod/read/ (дата обращения: 08.11.2023).
10. *Грязнов С.А.* Искусственный интеллект: этический аспект / С.А. Грязнов // Modern science. 2021. № 2–1. С. 365–267.
11. *Рязанова А.А.* Исследования и разработки в области искусственного интеллекта: достижения, выводы, перспективы / А.А. Рязанова // Вестник современных цифровых технологий. 2022. № 13. С. 38–46.
12. *Назарова Ю.В.* Актуальные проблемы цифровизации информационного общества / Ю.В. Назарова // Гуманитарные ведомости ТГПУ им. Л.Н. Толстого. 2023. № 2 (46). С. 39–47.
13. *Камалова Г.Г.* Правовые и этические принципы регулирования искусственного интеллекта и робототехники / Г.Г. Камалова // Право и государство: теория и практика. 2021. № 10 (202). С. 181–184.
14. *Кожевников Н.Н.* Этические аспекты искусственного интеллекта / Н.Н. Кожевников, В.С. Данилова // Наука и техника в Якутии. 2020. № 2 (39). С. 28–31.