

УДК 316.77(045)

**К ВОПРОСУ СОЦИАЛЬНОЙ ОТВЕТСТВЕННОСТИ РАЗРАБОТЧИКОВ  
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ФИЛОСОФСКИЕ АСПЕКТЫ**

**И.А. Латыпов**

Доктор философских наук, проф. кафедры истории, теории и практики  
социальных коммуникаций УдГУ.

ФГБОУ ВО Минобрнауки РФ «Удмуртский государственный университет»  
(УдГУ),

Искусственный интеллект (ИИ) стал очень популярен. Различные проблемы ИИ сейчас стоят в повестке дня всех научных и социально-политических СМИ, обсуждаются на всех влиятельных коммуникативных площадках. Возможности ИИ огромны, но и высока цена ошибок его разработчиков. Именно поэтому необходим глубокий анализ этики и социальной ответственности разработчиков искусственного интеллекта, не только с юридической, но и с философской точки зрения.

**Ключевые слова:** социальная ответственность, искусственный интеллект, философия, искусственные нейросети.

**A PROBLEM OF A SOCIAL RESPONSIBILITY OF AI DEVELOPERS:  
PHILOSOPHICAL ASPECTS**

**I.A. Latypov**

Artificial intelligence (AI) has gained great popularity. A lot of AI problems are high on the agenda of all scientific, scholar and social-political media; it is being discussed on all influential communication platforms. Abilities of AI are huge, but the price for AI developers' blunders is high. That's why a comprehensive study of not only legal, but also philosophical issues of an ethics and social responsibility of AI developers is necessary.

**Keywords:** social responsibility, Artificial intelligence, philosophy, artificial neural networks.

### **Введение**

Актуальность данной работы определяется, с одной стороны, тем, что искусственный интеллект (ИИ) развивается вместе с искусственными нейронными сетями (ИНС, нейросети), которые стали очень популярны. ИНС представляют собой интересный и перспективный инструментарий для решения сложных задач. С другой стороны, требуется тщательное изучение границ их областей применения и уровня возможного доверия к ним, а также разработки методов защиты от возможных угроз. Поэтому требуется всестороннее исследование различных, в том числе, и философско-этических вопросов социальной ответственности разработчиков ИИ.

Однако в данной работе не ставится целью обсуждение того, что такое искусственный интеллект, и в какой форме он реализуется. Как базовые определения, берутся дефиниции из широкого научного оборота, вошедшие в академические словари.

Предметная область исследования является междисциплинарной и связана с анализом проблемы этики и социальной ответственности разработчиков ИИ и нейросетей.

В качестве цели в данном исследовании ставится философский анализ различных проблем социальной ответственности разработчиков ИИ. Для достижения поставленной цели в работе решаются следующие две основные задачи: провести предварительный анализ понятия ИИ и выявить проблему социальной ответственности в сфере ИИ.

**История проблемы** междисциплинарного анализа ИИ уходит в работы Алана Тьюринга по теории алгоритмов (в т.ч. в тест Тьюринга как один из ключевых пунктов предыстории вопроса), а также литературные труды (философским образом нагруженные) Айзека Азимова, отечественных

фантастов Стругацких и многих других писателей, выступавших в роли футурологов.

В качестве исходного определения искусственной нейронной сети (ИНС), рассматривается совокупность информационно-компьютерной технологической среды (многослойной, в современности) и её математической модели, изначально имитировавшая нейронную сеть человеческого мозга и способная в результате машинного обучения обрабатывать вводимую информацию. Те разновидности и отличия ИНС от естественной сети нейронов, которые множатся в настоящее время, не входят в предметное поле данного исследования.

Предпосылки формирования ИИ, о проблематичности которого споры ведутся на протяжении почти 70 лет, постепенно складываются в работе с нейросетями как технологическими основаниями разработок ИИ. В самом деле, именно на основе ИНС и развиваются по преимуществу всё более продвинутое, но, по-прежнему, пока неудовлетворительные технологии ИИ.

Знания для потенциального ИИ могут собираться от экспертов, если речь идёт об экспертных системах, формирующихся сверху вниз. Либо знания потенциального ИИ могут закладываться в процессе машинного обучения (с учителем или без), если речь идёт о процессах, происходящих в ИНС, формирующихся снизу вверх, в ходе многократной обработки различных вводимых исходных данных.

ИНС созданы по аналогии с биологическими (или естественными) сетями нейронов, по которым передаются сигналы в головной мозг. Зарождение первых моделей ИНС ведут со статьи нейрофизиолога У. Маккалока и логика У. Питтса «Логическое исчисление идей, относящихся к нервной активности», опубликованной в 1943 г. в «Бюллетене математической биофизики» [5].

В 1949 г. нейропсихолог Д. Хебб в книге «Организация поведения» описал процесс самообучения ИНС. Математическую модель обработки информации мозгом для распознавания образов (получила название перцептрон) в 1957 г.

придумал Ф. Розенблатт, а спустя 3 года представил электронное устройство имитации работы мозга, которое с помощью «глаз»-камер распознавало отдельные символы на карточках [4].

Нет смысла даже перечислять появившееся большое количество интересных моделей и алгоритмов, связанных с искусственными нейронными сетями. Заметим только, что прорыв технологий ИНС произошёл в начале XXI в. благодаря увеличению вычислительных мощностей и появлению больших по размеру, адекватных обучающих выборок (датасетов).

Любая современная модель нейросети связана с определённым сложным алгоритмом, способным обучаться на основе вводимых или даже самостоятельно собираемых данных и принимать решения на этих знаниях. Нейросеть состоит из множества слоёв взаимосвязанных нейронов (входной слой, скрытые слои и выходной слой), которые интенсивно обмениваются информацией между собой.

Каждый отдельный нейрон в ИНС имеет свою функцию и может быть настроен на определённую задачу. Разные типы нейросетей способны решать чрезвычайно сложные задачи и используются в различных областях, например, для распознавания образов, обработки естественного языка, в робототехнике и т.д. Многие из таких задач раньше были неразрешимы для компьютеров.

В качестве исходного определения ИИ используется дефиниция российского законодателя в 123-ФЗ: «искусственный интеллект – комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека» [6].

Более подробный анализ истории проблемы и ряда смежных вопросов можно найти в источниках [1-3] и ряде других. Кодекс этики же для ИИ разработал «Альянс в сфере ИИ» (совместно с Аналитическим центром при правительстве РФ и с Минэкономразвития) в 2021г.

**Содержание и результаты исследования.**

Ширящееся множество разработчиков вновь создающихся нейросетей формирует растущие сообщества в разных формах и на разных коммуникативных площадках, например, в социальных медиа и др.

С философской точки зрения, пора ставить вопрос о субъекте работы нейросети ИИ. Слово «работа» не требует решения проблемы субъекта. В самом деле, многие автоматические устройства работают без частого вмешательства человека. А для термина «деятельность» неизбежна постановка проблемы субъекта. Можно ли говорить о «деятельности» ИИ и о результатах интеллектуальной деятельности нейросетей? И если можно, то в какой степени? И в какой степени эту деятельность можно назвать интеллектуальной?

И только в этом контексте можно ставить вопрос о вменении ответственности за неизбежные сбои, «зависания», нарушения авторских тех людей, произведения которых использовались для обучения ИИ и т.д. и т.п. Часть проблемы подобного вменения ответственности в нашей стране ставится в связи с подписанием Кодекса этики для ИИ многими российскими госкорпорациями и различными федеральными структурами. Например, подписание состоялось в ТАСС в рамках первого международного форума "Этика искусственного интеллекта: начало доверия", который проходил 26 октября 2021г. в Москве.

«Кодекс задуман частью федерального проекта "Искусственный интеллект" и Стратегии развития информационного общества на 2017-2030гг. Он устанавливает общие этические принципы и стандарты поведения, которыми следует руководствоваться участникам отношений в сфере искусственного интеллекта в своей деятельности. Кодекс распространяется на отношения, связанные с этическими аспектами создания (проектирования, конструирования, пилотирования), внедрения и использования технологий ИИ на всех этапах жизненного цикла, которые в настоящее время не урегулированы законодательством РФ или другими актами регулирования. ...Присоединение к

кодексу является добровольным. Подписи под документом поставили входящие в альянс Сбербанк, "Газпром нефть", "Яндекс", VK, МТС и Российский фонд прямых инвестиций, а также представители Сколково, "Ростелекома", Росатома, InfoWatch, ЦИАН и др.» [7].

Подробное обсуждение Кодекса этики ИИ не входит в задачи данной работы и требует отдельного дальнейшего исследования.

Следует лишь отметить, что этот кодекс «поделен на два раздела, куда входят, в том числе такие темы, как стимулирование развития ИИ, повышение осведомленности об этике применения ИИ, идентификация ИИ в общении с человеком и информационная безопасность» [7].

К основным предварительным результатам относятся: постановка проблемы социальной ответственности разработчиков ИИ, а также постановка вопроса об идентификации различных версий ИИ как новых социально значимых агентов информационной деятельности.

### **Выводы.**

Итак, даже беглое ознакомление с некоторыми примерами свидетельствует о том, что проблема социальной ответственности разработчиков ИИ в России осознаётся многими специалистами, вплоть до самого высокого уровня.

Необходимо расширение круга философских вопросов, обновление понятийного аппарата и междисциплинарного языка современной философии. Как это уже бывало ранее, философия замыкалась на себя в стабильных предсказуемых условиях. Но ускорение и непредсказуемость изменений современного общества настоятельно требуют развития теории и методологии философии по проблеме социальной ответственности разработчиков ИИ.

Обсуждение вопросов, поднятых или затронутых в данном исследовании, будет полезно для специалистов в сфере социальной философии, философии права, философии сознания, компьютерных наук, этики, а также пользователей ИИ, требует отдельного дальнейшего исследования.

**Список литературы:**

1. Верижникова, А. А. Латыпов, И.А. Роль и место искусственного интеллекта в современном медиакоммуникативном поле / А. А. Верижникова, И. А. Латыпов // Актуальные тенденции социальных коммуникаций: история и современность: сб. науч. ст.: [материалы Междунар. науч.-практ. конф. "Актуальные тенденции социальных коммуникаций: история и современность", 21-25 нояб. 2022 г., Ижевск] / М-во науки и высш. образования РФ, ФГБОУ ВО "Удмуртский государственный университет", Ин-т соц. коммуникаций, УО "Минский инновационный университет"; науч. ред. Г. В. Мерзлякова; отв. ред. Е. И. Михалева; редкол.: Л. В. Баталова, С. А. Васюра, С. А. Даньшина [и др.]. - Ижевск: Удмуртский университет, 2023. - + Электрон. ресурс. - С. 392-398. - URL: <http://elibrary.udsu.ru/xmlui/handle/123456789/21672>. (Дата посещения: 6 марта 2024 г.).
2. Латыпов, И. А. Некоторые аспекты самоорганизации онлайн-научно-образовательных сообществ на основе интереса к социально-гуманитарным проблемам / И. А. Латыпов // Вестник Восточно-Сибирской Открытой Академии. – 2016. – № 22. – С. 4. – EDN VNTUPB. - URL: <http://elibrary.ru/item.asp?id=25579744> (Дата посещения: 6 марта 2024 г.).
3. Латыпов, И. А. Мультифрактальность сетевых медиакоммуникаций в информационной цивилизации: философский аспект / И. А. Латыпов // Вестник Удмуртского университета. Серия Философия. Психология. Педагогика. – 2020. – Т. 30, № 4. – С. 333-343. – DOI 10.35634/2412-9550-2020-30-4-333-343. – EDN GATXYI. – URL: <https://cyberleninka.ru/article/n/multifraktalnost-setevyih-mediakommunikatsiy-v-informatsionnoy-tsivilizatsii-filosofskiy-aspekt> (Дата посещения: 6 марта 2024 г.).

4. Нейронные сети: как работают и где используются. 19.12.2021. – [дата обращения: 06.03.2024]. Режим доступа: <https://gb.ru/blog/nejronnye-seti/>
5. *Пиковер К.А.* Искусственный интеллект. Иллюстрированная история. От автоматов до нейросетей. Москва: Синдбад, 2021. ISBN 978-5-00131-330-4.
6. Федеральный закон от 24 апреля 2020 г. № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации - городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона "О персональных данных"». <https://base.garant.ru/73945195/> (дата обращения 06.03.2024).
7. Федеральный проект «Искусственный интеллект» - <https://ai.gov.ru/strategy/federalnyy-proekt-ii/?ysclid=lsg73k8nil799612894> (дата обращения 06.03.2024).