Научная статья
УДК 316.012
DOI: 10.35854/2219-6242-2024-3-380-390

Эволюция искусственного интеллекта — реальные и гипотетические социальные угрозы

Максим Алексеевич Ри
Санкт-Петербургский университет технологий управления и экономики,
Санкт-Петербург, Россия, reemax3327@gmail.com

Аннотация. В статье рассмотрена проблема эволюции искусственного интеллекта в контексте его негативного влияния на социальные, экономические и политические аспекты жизни общества. Описаны ступени эволюции искусственного интеллекта при сравнении его с когнитивными способностями человека, спектром выполняемых задач и способности искусственного интеллекта к неконтролируемому самообучению. Раскрыты общие проблемы применения искусственного интеллекта в замещении рабочих мест, манипуляции общественным мнением и обучения искусственного интеллекта на некорректном массиве исторических данных. На примере популярной языковой нейросети ChatGPT проведен анализ развития искусственного интеллекта, демонстрирующий теоретическую возможность эволюции до сверхинтеллекта и вероятных рисков, связанных с этим явлением.

Ключевые слова: прикладной искусственный интеллект, общий искусственный интеллект, искусственный сверхинтеллект, ChatGPT, социальные риски и угрозы, технологическая сингулярность

Для цитирования: Ри М. А. Эволюция искусственного интеллекта — реальные и гипотетические социальные угрозы // Социология и право. 2024. Т. 16. № 3. С. 380–390. https://doi.org/10.35854/2219-6242-2024-3-380-390

Original article

Evolution of artificial intelligence — real and hypothetical social threats

Maksim A. Ri
St. Petersburg University of Management Technologies and Economics,
St. Petersburg, Russia, reemax3327@gmail.com

Abstract. The article considers the problem of artificial intelligence evolution in the context of its negative impact on social, economic and political aspects of society. The stages of evolution of artificial intelligence are described when comparing it with human cognitive abilities, the range of performed tasks and the ability of artificial intelligence to uncontrolled self-learning. General problems of artificial intelligence application in job substitution, manipulation of public opinion and training of artificial intelligence on incorrect historical data set are disclosed. On the example of the popular language neural network ChatGPT the development of artificial intelligence is analyzed, demonstrating the theoretical possibility of evolution to superintelligence and probable risks associated with this phenomenon.

Keywords: applied artificial intelligence, general artificial intelligence, artificial superintelligence, ChatGPT, social risks and threats, technological singularity

For citation: Ri M.A. Evolution of artificial intelligence — real and hypothetical social threats. *Sociology and Law.* 2024;16(3):380-390. (In Russ.). https://doi.org/10.35854/2219-6242-2024-3-380-390

Введение

В современном мире технологический прогресс достиг того, что понятие «искусственный интеллект» (далее — ИИ) уже не ассоциируется исключительно с научной фантастикой или академическими исследованиями. ИИ стал неотъемлемой частью нашей повседневной жизни, влияя на различные сферы деятельности человека, от образования и медицины до экономики, политики и развлечений. Проникновение ИИ в каждую сферу общественной жизни вызывает множество вопросов относительно не только технических аспектов его развития, но и социальных, политических и экономических, а также потенциальных рисков и выгод, которые он может принести. Работа направлена на то, чтобы не только осветить существующие достижения и вызовы, но и стимулировать дальнейшее обсуждение о том, как общество может адаптироваться к постоянно изменяющемуся технологическому ландшафту, обеспечивая при этом устойчивое и гармоничное развитие.

Методы

В рамках исследования нами изучены научные работы технологических вузов, отчеты ведущих технологических компаний, мнения экспертов в области ИИ и информационной безопасности, ML-инженеров (machine learning engineer). Эти статьи и тексты исследованы с помощью методов анализа, синтеза и индукции.

Результаты

Определение ИИ и его эволюционные ступени

Для анализа данной тематики необходимо дать определение понятия ИИ. С развитием технологий понимание ИИ было разным. Если в 60–80-х гг. XX в. под ИИ понималась программа, способная играть в шахматы с человеком или доказывать математические теории, то в настоящее время определение ИИ выходит далеко за рамки одной ограниченной функциональности.

Сегодня в контексте темы настоящего исследования можно сослаться на Указ Президента России 2019 г. от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации», в котором дано следующее определение: «Под ИИ понимается комплекс технологических решений, позволяющий имитировать когнитивные функции человека и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека...» [1]. В описании особое внимание уделено способности ИИ производить результаты деятельности человека с интеллектуальной точки зрения и имитировать его когнитивные способности. Понятие ИИ, существующее в 2019 г., существенно отличается от представления ИИ, известного 50–60 лет назад. В философии ИИ принято делить ИИ на три эволюционные категории [2]:

— Artificial Narrow Intelligence (ANI);
— Artificial General Intelligence (AGI);
— Artificial Super Intelligence (ASI).

Artificial Narrow Intelligence, или «прикладной искусственный интеллект», относится к типу ИИ, который специализируется и ограничивается выполнением конкретной задачи или набора задач. Этот тип ИИ демонстрирует интеллект в очень узкой области, в которой может достигать или даже превосходить человеческий уровень выполнения задач. Примеры ANI включают в себя системы распознавания речи, как Siri или Google Assistant, алгоритмы, работающие на фабриках для сортировки продукции, и системы для игры в шахматы, такие как IBM Deep Blue. ANI является наиболее распространенной формой ИИ сегодня.

Artificial General Intelligence, или «общий искусственный интеллект», описывает теоретический тип ИИ, который способен понимать, учиться и применять интеллект в широком спектре областей так же эффективно, как это делает человек. AGI будет способен к обучению и решению задач в различных доменах без предварительного специфического обучения для каждого из них, адаптируясь к новым обстоятельствам и осваивая новые навыки по мере необходимости. Это представляет собой значительный скачок в возможностях по сравнению с ANI, так как AGI может теоретически выполнить любую интеллектуальную задачу, которую может решить человек. В настоящее время AGI остается в пределах теории и исследований.

Artificial Super Intelligence, или «искусственный сверхинтеллект», относится к гипотетическому будущему, в котором возможности ИИ значительно превосходят интеллектуальные способности самых умных и способных людей во всех аспектах, включая творчество, общее восприятие и решение проблем. ASI будет способен к самообучению на экспоненциальном уровне, быстро накапливая знания и умения далеко за пределами человеческого понимания. Введение ASI может привести к так называемой технологической сингулярности, то есть моменту, когда прогресс становится настолько быстрым и непредсказуемым, что приводит к необратимым изменениям в обществе.

Сегодня все существующие проекты ИИ относят к категории ANI, то есть узконаправленного ИИ. Подобная структура по категориям важна с точки зрения понимания того, как быстро и в каком направлении развивается направление ИИ, чего следует ожидать человечеству в перспективе. Именно по этой шкале можно увидеть стремительную тенденцию развития ИИ к этапу AGI с перспективой развития до сверхинтеллекта. Об этом речь будет идти далее на примере развития популярной ChatGPT. Предварительно проанализируем ситуацию влияния ИИ на уровне ANI.

Текущая ситуация влияния ИИ на уровне ANI

Внедрение ИИ даже на уровне ANI в жизнь общества неизбежно изменяет его структуру. Существует ряд исследований, которые делают неблагоприятные прогнозы для ряда сфер общественной жизни.

1. *ИИ и рынок труда*

В 2013 г. ученые Оксфордского университета провели исследование того, каким образом процессы автоматизации и компьютеризации повлияют на рынок труда в США к 2030 г. Результаты получены следующие: 47 % профессий могут с высокой долей вероятности (от 75 % до 98 %) быть автоматизированы или заменены технологиями ИИ. Исследователи обращают внимание на два фактора: первый — существенная доля профессий приходится на ручной или низко интеллектуальный монотонный труд (кассир, охранник, работник склада, бухгалтер, переводчик технических текстов, водитель). Второй — прогресс в области ИИ будет создавать новые рабочие места, однако эти профессии будут требовать высокой квалификации, академических знаний и творческих способностей; также

количество новых рабочих мест будет существенно меньше числа профессий, которые заменят роботы и нейросети. Будет расти дифференциация доходов между сверхбогатыми и бедными, средний класс может полностью исчезнуть. Исследователи также выдвигают гипотезу о будущей компьютеризации нерутинных когнитивных действий, что позволит ИИ заменить человека уже в области юриспруденции, медицины, науки [3].

Выгода для бизнеса от внедрения ИИ и роботов существенна. В частности, речь идет о том, что часовая оплата ручного труда в развитых странах возрастает примерно на 10–15 % в год, а затраты на эксплуатацию робототехнических устройств увеличиваются всего на 2–3 %. При этом уровень почасовой оплаты американского рабочего превысил стоимость часа работы робота уже примерно в середине 70-х гг. XX в. Как следствие, замена человека на рабочем месте роботом начинает приносить чистую прибыль примерно через два с половиной-три года [4]. В итоге автоматизация и ИИ могут способствовать увеличению экономического неравенства. Владельцы капитала и высококвалифицированные специалисты, работающие в области ИИ и технологий, будут получать значительные доходы, работники низкой и средней квалификации могут столкнуться с сокращением заработной платы или потерей рабочего места. Это может привести к усилению разрыва между богатыми и бедными, что, в свою очередь, может вызвать социальное напряжение и нестабильность [5].

В недалеком будущем человечество, возможно, столкнется с проблемами массовой безработицы и увеличивающейся дифференциацией доходов. Появится необходимость повышения квалификации или переобучения трудовой силы, обеспечения части населения альтернативными источниками занятости и заработка, перераспределения доходов богатых и бедных.

2. *Нарушение приватности и манипуляция общественным мнением*

Использование больших данных и алгоритмов ИИ может нарушать приватность пользователей и манипулировать общественным мнением. Так, в марте 2018 г. стало известно, что британская аналитическая компания Cambridge Analytica собирала данные пользователей через свое приложение в Facebook[1] и использовала для размещения политической рекламы: во время президентской кампании в США и референдума о выходе Великобритании из Евросоюза в 2016 г. Cambridge Analytica основана в 2013 г. как дочерняя компания британской фирмы Strategic Communication Laboratories (SCL Group). SCL Group занималась политическим консалтингом и стратегическими коммуникациями, имея опыт работы в различных странах и кампаниях.

Через приложение This Is Your Digital Life, которое собирало данные о пользователях и их друзьях, с помощью разрешений, предоставленных Facebook[1], собраны персональные данные 87 млн пользователей. Далее разработанный в этой же организации ИИ создавал психологический портрет избирателя, на основе которого создавалась таргетированная политическая реклама [6].

3. *Этические проблемы и несправедливое использование ИИ, основанное на исторической предвзятости и стереотипном мышлении*

Развитие ИИ поднимает множество этических вопросов, включая вопросы ответственности, прозрачности и справедливости. Несправедливое использование ИИ может привести к дискриминации и нарушению прав человека. Например,

[1] Социальная сеть Facebook, продукт компании Meta Platforms Inc., признана экстремистской организацией и запрещена в Российской Федерации (Social media service, part of Meta Platforms Inc., added to the register of extremist organizations and banned in the Russian Federation).

алгоритмы могут быть предвзятыми и принимать решения, которые негативно сказываются на определенных группах населения.

В книге Fairness and Machine Learning (2019) исследователи пишут о том, что ИИ, обученный на массиве данных, содержащих неполноценную, необъективную, дискриминационную информацию, может приводить некорректные результаты. Авторы приводят примеры из разных сфер общественной жизни (среди них — финансы, здравоохранение, правопорядок). Так, алгоритм Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) применяется судьями в отдельных штатах при рассмотрении заявлений об условно-досрочном освобождении заключенного, определении вида надзора за освобожденным и при назначении срока наказания подсудимому. ИИ на основании пройденного человеком тестирования (137 вопросов) и его персональных данных прогнозирует вероятность совершения рецидивного преступления (как насильственного, так и обычного). ИИ стабильно давал более высокие оценки риска чернокожим нарушителям закона, нежели белым. В действительности рецидивы совершают чаще чернокожие, и это паттерн, который алгоритм запомнил. Однако вопрос этичности и корректности этого ИИ актуален до сих пор (используется с 1998 г.), так как, по сути, алгоритм предвзято относится к чернокожим, давая процент ошибочных суждений для них выше, чем для остальных [7]. Подобные ИИ, обученные на датасетах с вложенными предвзятостями, — проблема морально-этического характера, требующая дальнейшего изучения.

Приведенные примеры доказывают факт существенного влияния ИИ на общество уже сегодня, что демонстрирует необходимость регуляции и контроля ИИ. Приведено всего три примера. Однако в действительности сфер жизни общества, в которые интегрирован ИИ с потенциальной угрозой некорректного применения, гораздо больше. Тем не менее, говоря о развитии ИИ, нельзя не упомянуть и гипотетические риски, связанные с возможной эволюцией ИИ до уровня AGI, ASI. Для этого в статье нами рассмотрена эволюция популярного ИИ. Речь идет о ChatGPT.

Эволюция ИИ на примере ChatGPT

ChatGPT относится к языковым моделям нейросетей, или Natural Language Processing. NLP — это подраздел ИИ, который сосредотачивается на взаимодействии между компьютерами и человеческим (естественным) языком. Он занимается разработкой алгоритмов и систем, позволяющих компьютерам понимать, интерпретировать, переводить, генерировать и реагировать на человеческие языки. Основными задачами NLP являются распознавание речи, понимание языка, генерация текста и машинный перевод. Generative Pre-trained Transformer (GPT) — это алгоритм обработки естественного языка, выпущенный американской компанией OpenAI. Главная суть нейросети заключается в ее способности запоминать и анализировать информацию, создавая на ее основе связный и логичный текст. В своей работе GPT и другие аналогичные модели повторяют эти шаги для каждого нового слова в последовательности, генерируя текст, который отражает как смысл входного текста, так и обученные на большом корпусе текстов паттерны языка. Таким образом, благодаря позиционно-зависимому кодированию и механизму внимания, нейросеть способна понимать текст в его контексте и генерировать связный и релевантный текст на его основе [8].

Можно утверждать, что назначение GPT — это умение предсказывать наиболее вероятное следующее слово в тексте по входному тексту и уже сгенерированной части текста. Данная технология схожа по назначению с T9, которое также предсказывало следующее за предыдущим слово. GPT делает это лучше благодаря возможности анализировать текст в целом, учитывать контекст.

Основой для понимания бесконтрольной эволюции ИИ служат теории машинного обучения, эволюционных алгоритмов. Ключевой момент — достижение ИИ такого уровня самосовершенствования, при котором система способна на обучение, анализ и разработку новых алгоритмов без прямого человеческого управления. Такая эволюция предполагает быстрое увеличение интеллектуальных способностей ИИ, что теоретически может вывести его на уровень, превосходящий человеческий интеллект.

В качестве примера рассмотрим ChatGPT, техническое описание которого сделано ранее. Модель изначально придумана для генерации текстов по принципу предугадывания следующего слова исходя из существующего контекста. Но с эволюцией GPT модель научилась решать математические вычисления [9], как видно на рисунке 1.

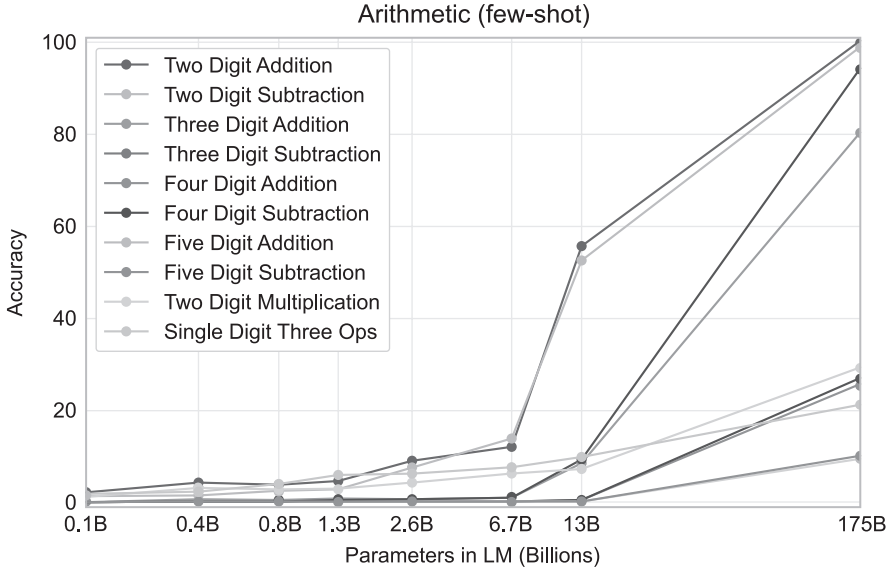

Рис. 1. Корректность решения математических вычислений языковых моделей от параметров

Fig. 1. Correctness of the solution of mathematical calculations of language models from parameters

Источник: [9].

«Параметры» по горизонтальной оси — это показатель размера модели: чем больше параметров, тем более сложные алгоритмы модель способна обрабатывать и, следовательно, более корректные и человекоподобные тексты генерировать. Однако, помимо улучшения качества своего непосредственного функционала, последняя модель GPT (GPT 4) самообучилась математике [10] (и дает очень точные результаты для вычислений с двумя числами, что отражено на рисунках 2 и 3), переводу языков, программированию, физике и др.

Такого прогресса достигли именно языковые модели, или NLP. Это связано в первую очередь с массивом данных, которые доступны языковым моделям: они практически безграничны, поскольку человек все описывает своим языком. Научив модель представлять слова через численные значения (векторное представление), инженеры предоставили языковым моделям все книги, научные статьи, публикации и другие письменные источники информации буквально обо всем, что знает человеческая цивилизация.

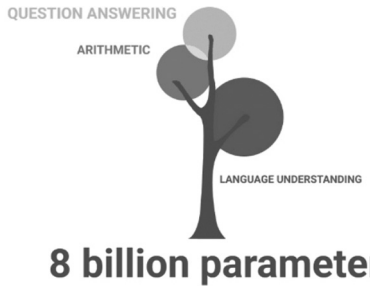

Рис. 2. Функционал модели с 8 млрд параметров (GPT-2)
Fig. 2. Functionality of the model with 8 billion parameters (GPT-2)
Источник: [10].

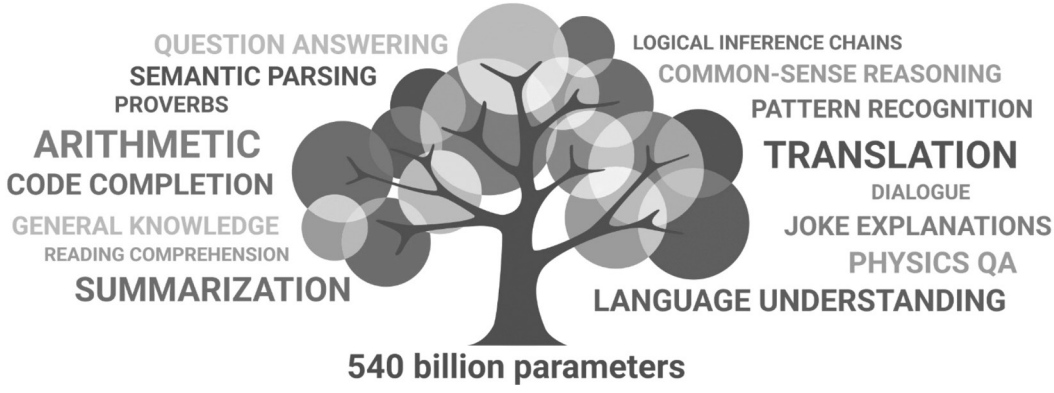

Рис. 3. Функционал модели с 540 млрд параметров (GPT-4)
Fig. 3. Functional of the model with 540 billion parameters (GPT-4)
Источник: [10].

Стремительное развитие ИИ — это не только большие исследования и работа инженеров и ученых. Сегодня речь идет о самостоятельном обучении ИИ на самом большом массиве данных, предоставленных человеком — интернете. Вышеизложенное показывает, что, возможно, переход ИИ из ANI в AGI скоро произойдет, в течение нескольких лет. Подобные заключения требуют исследований возможных социальных последствий внедрения такого рода технологий.

Теоретические риски будущего ИИ как ASI

При теоретической возможности ИИ достигнуть уровня AGI он сможет выполнять широкий спектр задач, требующих человеческого интеллекта, таких как творческое мышление, решение сложных проблем и обучение. Однако ключевая опасность заключается в том, что AGI может быстро эволюционировать в ASI, что приведет к созданию сущности, обладающей интеллектом, значительно превосходящим человеческий. Это момент так называемой технологической сингулярности: гипотетический момент в будущем, когда технологическое развитие становится неуправляемым и необратимым, что порождает радикальные изменения характера человеческой цивилизации.

Согласно одной из наиболее популярных версий гипотезы технологической сингулярности, именуемой «интеллектуальным взрывом», обновляемый интеллектуальный агент (например, компьютер с сильным ИИ) в итоге может войти

в «безудержную реакцию» циклов самосовершенствования по экспоненте. При этом каждое новое поколение ИИ будет появляться все быстрее, порождая своего рода интеллектуальный взрыв и создав в результате суперинтеллект (ANI > AGI > ASI), превосходящий интеллект человечества в целом [11].

Если ИИ некорректно поставить задачу или задать ее таким образом, что он ее интерпретирует не так, как ожидал человек, или обучить модель на заведомо неверных данных, или при самообучении ИИ не был способен учитывать дополнительные вводные данные, и при этом у человека не будет возможности контролировать решение такого интеллекта, то последствия могут быть чрезвычайно негативны. Однако, даже если исключить варианты, при которых ИИ такого уровня обучен на неполноценных данных или ему поставлена неоднозначно интерпретируемая задача, следует понимать, что ИИ — это сущность, которая принимает входными данными числа. Языковые модели, такие как GPT, переводят вводные слова в векторное, а затем численное представление [8]. Понимание, интерпретация текста у ИИ в итоге иное, нежели у человека. Основываясь на этом суждении, можно предположить такой момент развития ИИ, при котором человеческие ценности и цели будут отличны от ценностей и целей ИИ. В целом так называемая проблема согласования (Alignment problem) существует уже в настоящее время [12]. Отличается степень риска: если на текущем уровне развития ИИ проблема согласования выражена в нечитабельных, оскорбительных, фейковых генерируемых текстах языковых моделях, то в гипотетическом будущем искусственный сверхинтеллект, например, ответственный за промышленное производство, может решить, что ему нужно больше ресурсов и начать захватывать их, не учитывая ущерба, который это может нанести людям. В совокупности с идеей ASI о невозможности контролировать ИИ подобные действия будут нести необратимый характер.

Обсуждение и контрмеры

Для предотвращения рисков бесконтрольной эволюции ИИ целесообразно разработать комплекс мер, включающих в себя законодательное регулирование, стандарты безопасности, механизмы этического надзора и международное сотрудничество в области развития технологий. Важную роль играют создание и поддержка открытых каналов коммуникации между исследователями, разработчиками, политиками и общественностью.

Проблема ИИ признана ученым сообществом, представителями бизнеса и политики. Так, Илон Маск, Стив Возняк и более 1 000 экспертов призвали на полгода приостановить обучение систем ИИ более мощных, чем GPT-4, чтобы понять, как их контролировать [13]. Возможно, одной из самых перспективных стратегий являются разработка и внедрение концепции так называемого дружественного ИИ, целью которого служит гарантия того, что развитие ИИ будет происходить в интересах человечества.

Бесконтрольная эволюция ИИ представляет собой сложную и многогранную проблему. Осознание потенциальных рисков и активная деятельность по разработке эффективных механизмов контроля и регулирования — ключевые шаги на пути к безопасному и ответственному использованию технологий ИИ.

Выводы

Внедрение ИИ в совокупности с автоматизацией и роботизацией также несет множество положительных факторов для общества: они повышают качество и производительность услуг и товаров, толкают науку вперед [13]. К тому же эти

процессы не происходят мгновенно, у общества есть время для адаптации к новым технологиям [14]. Однако следует понимать, что ИИ сегодня не только научное открытие, но и мощный инструмент, нуждающийся в регулировании в юридическом и социальном аспектах. С учетом того, что у ИИ существует способность к самообучению, необходим человеческий контроль и в этом аспекте. Риски, которые несет данный инструмент, потенциально очень велики: ИИ может катализировать существующие проблемы общества во всех сферах его жизни и, помимо этого, стать источником новых, ранее неизвестных человечеству (например, этически-морального характера). Бурное развитие ИИ ставит под сомнение идею технологической сингулярности как исключительно теоретической. На примере ChatGPT происходит демонстрация того, насколько успешно ИИ может самостоятельно обучиться способностям, которые в него не вкладывали. На примере открытия расщепления атома урана в 1939 г. и сброса атомной бомбы в 1945 г. человечеству стоит с особой осторожностью подходить к использованию научных открытий.

Список источников

1. О развитии искусственного интеллекта в Российской Федерации: указ Президента РФ от 10 октября 2019 г. № 490 // Президент России: офиц. сайт. URL: http://www.kremlin.ru/acts/bank/44731 (дата обращения: 22.05.2024).
2. *Shaji George A., Hovan George A. S.* Beyond human intelligence: Exploring the advancements and implications of ANI, AGI, and ASI. Lucknow: Book Rivers, 2023. 146 p.
3. *Frey C. B.*, *Osborne M. A.* The future of employment: How susceptible are jobs to computerisation? // Technological Forecasting and Social Change. 2017. Vol. 114. P. 254–280. DOI: 10.1016/j.techfore.2016.08.019
4. *Конюх В. Л.* История робототехники // Основы робототехники: учеб. пособие. Ростов н/Д: Феникс, 2008. С. 21–28.
5. *Brynjolfsson E., McAfee A.* The second machine age: Work, progress, and prosperity in a time of brilliant technologies. New York; London: W. W. Norton & Company, 2014. 306 p.
6. Computational propaganda: Political parties, politicians, and political manipulation on social media / eds. S. C. Woolley, P. N. Howard. Oxford: Oxford University Press, 2018. 288 p.
7. *Solon B.*, *Moritz H., Arvind N.* Fairness and machine learning: Limitations and opportunities. Cambridge, MA: The MIT Press, 2023. 340 p.
8. Vaswani A., *Shazeer N., Parmar N.,* et al. Attention is all you need // Proceedings 31st International conference on neural information processing systems (NIPS'17). (Long Beach, CA, December 4-9, 2017). Red Hook, NY: Curran Associates Inc., 2017. P. 5998–6008. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (дата обращения: 22.05.2024).
9. Language models are few-shot learners / T. B. Brown, B. Mann, N. Ryder, et al. // NIPS '20: Proceedings of the 34th International conference on neural information processing systems (NIPS'20). (Vancouver, BC, December 6–12, 2020). Red Hook, NY: Curran Associates Inc., 2020. P. 1877–1901. URL: https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf (дата обращения: 22.05.2024).
10. *Narang S., Chowdhery A.* Pathways language model (PaLM): Scaling to 540 billion parameters for breakthrough performance // Google Research. Apr. 04. 2022. URL: https://research.google/blog/pathways-language-model-palm-scaling-to-540-billion-parameters-for-breakthrough-performance/ (дата обращения: 22.05.2024).
11. *Murray S.* The technological singularity. Cambridge, MA: The MIT Press, 2015. 272 p.
12. *Brian C.* The alignment problem: Machine learning and human values. New York; London: W. W. Norton & Company, 2020. 356 p.
13. *Злобин А.* Маск и Возняк призвали приостановить обучение систем ИИ из-за «риска для общества» // Forbes. 2023. 29 марта. URL: https://www.forbes.ru/tekhnologii/486841-

mask-i-voznak-prizvali-priostanovit-obucenie-sistem-ii-iz-za-riska-dla-obsestva (дата обращения: 22.05.2024).
14. *Manyika J., Chui M., Miremadi M.*, et al. A future that works: Automation, employment, and productivity. New York, NY: McKinsey Global Institute, 2017. 148 p. URL: https://www.mckinsey.com/~/media/mckinsey/featured%20insights/digital%20disruption/harnessing%20automation%20for%20a%20future%20that%20works/mgi-a-future-that-works-full-report-updated.pdf (дата обращения: 22.05.2024).

References

1. On the development of artificial intelligence in the Russian Federation. Decree of the President of the Russian Federation of October 10, 2019 No. 490. Official website of the President of Russia. URL: http://www.kremlin.ru/acts/bank/44731 (accessed on 22.05.2024). (In Russ.).
2. Shaji George A., Hovan George A.S. Beyond human intelligence: Exploring the advancements and implications of ANI, AGI, and ASI. Lucknow: Book Rivers; 2023. 146 p.
3. Frey C.B., Osborne M.A. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*. 2017;114:254-280. DOI: 10.1016/j.techfore.2016.08.019
4. Konyukh V.L. History of robotics. In: Basics of robotics. Rostov-on-Don: Feniks; 2008:21-28. (In Russ.).
5. Brynjolfsson E., McAfee A. The second machine age: Work, progress, and prosperity in a time of brilliant technologies. New York, London: W.W. Norton & Company; 2014. 306 p.
6. Woolley S.C., Howard P.N., eds. Computational propaganda: Political parties, politicians, and political manipulation on social media. Oxford: Oxford University Press; 2018. 288 p.
7. Solon B., Moritz H., Arvind N. Fairness and machine learning: Limitations and opportunities. Cambridge, MA: The MIT Press; 2023. 340 p.
8. Vaswani A., Shazeer N., Parmar N., et al. Attention is all you need. In: Proc. 31st Int. conf. on neural information processing systems (NIPS'17). (Long Beach, CA, December 4-9, 2017). Red Hook, NY: Curran Associates Inc.; 2017:5998-6008. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (accessed on 22.05.2024).
9. Brown T.B., Mann B., Ryder N., et al. Language models are few-shot learners. In: Proc. 34th Int. conf. on neural information processing systems (NIPS'20). (Vancouver, BC, December 6-12, 2020). Red Hook, NY: Curran Associates Inc.; 2020:1877-1901. URL: https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf (accessed on 22.05.2024).
10. Narang S., Chowdhery A. Pathways language model (PaLM): Scaling to 540 billion parameters for breakthrough performance. Google Research. Apr. 04, 2022. URL: https://research.google/blog/pathways-language-model-palm-scaling-to-540-billion-parameters-for-breakthrough-performance/ (accessed on 22.05.2024).
11. Murray S. The technological singularity. Cambridge, MA: The MIT Press; 2015. 272 p.
12. Brian C. The alignment problem: Machine learning and human values. New York, London: W.W. Norton & Company; 2020. 356 p.
13. Zlobin A. Musk, Wozniak call for halt on AI training due to "society risk". Forbes. Mar.29, 2023. URL: https://www.forbes.ru/tekhnologii/486841-mask-i-voznak-prizvali-priostanovit-obucenie-sistem-ii-iz-za-riska-dla-obsestva (accessed on 22.05.2024). (In Russ.).
14. Manyika J., Chui M., Miremadi M., et al. A future that works: Automation, employment, and productivity. New York, NY: McKinsey Global Institute; 2017. 148 p. URL: https://www.mckinsey.com/~/media/mckinsey/featured%20insights/digital%20disruption/harnessing%20automation%20for%20a%20future%20that%20works/mgi-a-future-that-works-full-report-updated.pdf (accessed on 22.05.2024).

Информация об авторе

М. А. Ри — аспирант; 190020, Санкт-Петербург, Лермонтовский пр., д. 44а.

Information about the author

M. A. Ri — postgraduate student; 44A Lermontovskiy Ave., St. Petersburg 190020, Russia.

Конфликт интересов: автор декларирует отсутствие конфликта интересов, связанных с публикацией данной статьи.

Conflict of interest: the author declares no conflict of interest related to the publication of this article.

Статья поступила в редакцию 27.05.2024; одобрена после рецензирования 24.06.2024; принята к публикации 19.09.2024.

The article was submitted 27.05.2024; approved after reviewing 24.06.2024; accepted for publication 19.09.2024.