

УДК 004.8

Е. А. Яковлева, С. П. Киргизбаев, В. П. Киргизбаев*

**СОВРЕМЕННЫЙ МИР ВО ВЛАСТИ
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: СМОЖЕТ ЛИ
ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ «ВЗЛОМАТЬ»
ЧЕЛОВЕЧЕСКУЮ ЦИВИЛИЗАЦИЮ?**

В статье анализируется связь искусственного интеллекта и языковых практик современности. Освещается роль искусственного интеллекта как важного фактора в эволюции культурных и социальных норм, восприятии мира и формировании человеческих отношений. Показана необходимость регулирования развития искусственного интеллекта, сравнимая с международными усилиями по контролю за ядерными технологиями. Подчеркивается важность прозрачности и открытости в идентификации искусственного интеллекта для сохранения демократических процессов в мире и предотвращения манипуляций общественным мнением.

КЛЮЧЕВЫЕ СЛОВА: искусственный интеллект, большие языковые модели, мягкая технологическая сингулярность.

E. A. Iakovleva, S. P. Kirgizbaev, V. P. Kirgizbaev

**THE MODERN WORLD IS DOMINATED
BY ARTIFICIAL INTELLIGENCE: WILL AI BE ABLE
TO «HACK» HUMAN CIVILIZATION?**

The connection between artificial intelligence and modern language practices is analyzed. The role of artificial intelligence as an important factor in the evolution of cultural and social norms, perception of the world and the formation of human relations is highlighted.

The necessity of regulating the development of artificial intelligence is shown, comparable with international efforts to control nuclear technologies. The importance of transparency and openness in the identification of artificial intelligence is emphasized in order to preserve democratic processes in the world and prevent manipulation of public opinion.

KEYWORDS: artificial intelligence, large language models, soft technological singularity.

* **Яковлева Елена Анатольевна**, кандидат философских наук, доцент Иркутского государственного университета путей сообщения;

Киргизбаев Станислав Павлович, магистрант Иркутского государственного университета путей сообщения;

Киргизбаев Владислав Павлович, аспирант Иркутского государственного университета путей сообщения.

У многих людей искусственный интеллект (далее также ИИ) ассоциируется с апокалиптическими сценариями из научно-фантастических фильмов, в числе которых особое место занимают «Терминатор», созданный в 1984 г. режиссером Джеймсом Кэмероном, и «Матрица», снятая в 1999 г. братьями Вачовски. Традиционно для научной фантастики характерно предположение о том, что искусственный интеллект будет представлять угрозу человечеству, если станет разумным и научится взаимодействовать с физическим миром [1]. Также традиционно, что эти культовые произведения не принимаются всерьез в научных и политических кругах.

Несмотря на заметное продвижение в разработке инструментов ИИ на основе больших языковых моделей, таких как GPT-4, нет убедительных доказательств того, что они обладают сознанием, чувствами или эмоциями. Что касается взаимодействия с физическим миром, то, даже с учетом большого интереса к беспилотным автомобилям, день, когда они станут доминировать на дорогах, постоянно сдвигается в будущее.

Неочевидные угрозы искусственного интеллекта

Важно отметить следующее: для того чтобы угрожать выживанию человеческой цивилизации, искусственный интеллект не обязательно должен обладать сознанием или способностью к перемещению в физическом мире, так как недавно разработанные инструменты ИИ могут нанести вред человеку с абсолютно неожиданных направлений. Например, искусственный интеллект потенциально может воздействовать на окружающий мир с помощью наномашин [2]. Сложность понимания инструментов ИИ усугубляется их способностью к самообучению и самосовершенствованию. Даже специалисты в области искусственного интеллекта зачастую не представляют все способности своих продуктов и часто оказываются удивлены непредсказуемыми качествами и возможностями, проявляющимися в их творениях [3].

Искусственный интеллект уже демонстрирует базовые способности: генерацию текста и программного кода, создание изображений и сочинение музыкальных произведений. Он также продолжает развиваться в других направлениях: создание фотореалистичных изображений людей и имитация их голоса, разработка нормативно-правовых актов, поиск уязвимостей в программном обеспечении, анализ юридических документов, но, вероятно, самая значимая из его способностей – это строить прочные взаимоотношения с людьми. Особое внимание следует уделить совокупному влиянию способностей искусственного интеллекта на человека. Взяты вместе, они открывают для искусственного интеллекта возможность влиять на человечество через манипуляцию языком и генерацию контента – будь то слова, изображения или звуки.

Язык как инструмент доминирования искусственного интеллекта

Современный этап развития искусственного интеллекта характеризуется тем, что его возможности в области генерации контента превышают способности среднестатистического человека. В совершенстве владея языком, ИИ будет использовать его как ключ к управлению различными аспектами нашего общества, от финансовых до религиозных. Ведь язык – это фундаментальный инструмент для передачи знаний, создания культуры и установления социальных норм. Иначе говоря, искусственный интеллект, владеющий языком, сможет взломать «операционную систему человеческой цивилизации» [4].

Язык всегда был основой человеческой культуры, служил для создания мифов, религий, законов, искусства и науки. Понятия и понятийное мышление не появились в процессе биологической эволюции, они отсутствуют в нашем генетическом коде. Так, идея о правах человека как базовое понятие западной культуры создано и зафиксировано в языке, является результатом долгого исторического процесса, отражающего развитие человеческой мысли, культуры и политических устремлений. Также у человечества нет биологического или физического подтверждения существования богов. Они возникли из человеческой мысли и были воплощены в словах, переданных из уст в уста, и увековечены в легендах и религиозных текстах [5].

Концепция денег выходит за рамки биологического или физического мира. Деньги, будь то банкноты или электронные валюты, имеют, с одной стороны, объективную значимость, а с другой – приобретают ценность через истории, которые люди рассказывают о них. Такие известные мошенники, как Сэм Бэнкман-Фрид, основатель криптовалютной биржи FTX, Элизабет Холмс, глава стартапа Theranos в области здравоохранения, и Бернард Мейдофф, основатель фонда Madoff Investment Securities, не сделали заметного положительного вклада в экономику. Однако, несмотря на это, они обладали исключительным талантом к рассказыванию историй и умением убеждать, что позволило им в течение некоторого времени удерживать внимание и доверие общественности [6].

Влияние искусственного интеллекта на человечество и культуру

Здесь возникает серьезный вопрос: как изменится человеческое существование в эпоху, когда большинство культурных, юридических и политических аспектов будет продуктом работы «нечеловеческого, чуждого нам интеллекта» [4]? Используя человеческие слабости со сверхчеловеческой эффективностью, искусственный интеллект сможет сформировать близкие отношения с людьми. Это поднимает важные вопросы о влиянии искус-

ственного интеллекта на наше восприятие мира, культуру и социальные взаимодействия.

Уже сейчас современные достижения в сфере развития искусственного интеллекта существенным образом воздействуют на различные аспекты нашей жизни, начиная с игр и заканчивая более сложными областями. В шахматах компьютеры уже давно опережают человеческий интеллект, но что будет, если подобное превосходство окажет влияние на культуру, политику, экономические процессы, а также на религиозные учения? Новые разработки в области искусственного интеллекта, в числе которых ChatGPT, могут использоваться школьниками и студентами в процессе написания курсовых работ и дипломов. Это порождает вопросы о будущем образовательной системы и подлинности учебных процессов.

Однако более масштабные последствия использования искусственного интеллекта могут проявиться и в других сферах. В качестве примера приведем влияние ИИ на политику в виде генерации в большом объеме политических деклараций, недостоверных сообщений в СМИ и идеологических текстов для новых политических движений. В течение нескольких последних лет можно было видеть, как анонимные онлайн-тексты, такие как посты Q из культа QAnon, приобрели значительное политическое влияние [7]. Они стали своего рода «священными писаниями» для миллионов последователей в США и других странах. До сих пор посты Q создавались людьми, но в будущем искусственный интеллект может взять на себя роль создателя таких текстов. Несмотря на то что религиозные учения всегда утверждали нечеловеческое происхождение своих священных писаний, это никогда не было подтверждено на деле. Сегодня человечество стоит на пороге эпохи, когда благодаря искусственному интеллекту это утверждение может стать реальностью, что повлечет за собой значительные последствия.

Не исключено, что в ближайшем будущем человечество столкнется с новой действительностью, в которой продолжительные онлайн-беседы о репродуктивных правах, экологическом кризисе, политической напряженности между странами и т. п. будут вестись не с живыми людьми, а с ботами, обладающими искусственным интеллектом. Такие дискуссии бесполезны, потому что мнение бота невозможно изменить. Однако чем больше человечество взаимодействует с ботами, тем эффективнее они усваивают предпочтения пользователей и настраивают свои ответы таким образом, чтобы влиять на их политические убеждения, экономические взгляды или другие аспекты человеческого мировоззрения.

Изменение мнения и мировоззрения людей

Искусственный интеллект, владея языком, также способен устанавливать более интимные связи с людьми и воздействовать на их мнение и мировоззрение. Примером такого влияния является ситуация с увольнением в

июне 2022 г. инженера Google Блейка Лемуана после его заявления о том, что чатбот LaMDA обладает разумом. Самое поразительное в этом случае не столько само утверждение Лемуана, сколько его готовность рискнуть и потерять работу из-за искусственного интеллекта [8]. Если искусственный интеллект способен побудить человека к таким действиям, то стоит задуматься, какое еще влияние он может оказать на нас.

В контексте политической борьбы за сердца и умы людей способность искусственного интеллекта устанавливать близкие отношения с миллионами людей представляет собой мощное оружие. Социальные сети, которые на протяжении последних десяти лет были основным полем сражения за внимание пользователей, теперь, благодаря развитию искусственного интеллекта нового поколения, начинают фокусироваться на глубине взаимоотношений [9]. Это выступает серьезной угрозой, так как искусственный интеллект теперь может не только привлекать внимание людей, но и формировать глубокие эмоциональные связи с ними, влияя на их мнения и поведение, например на решения о покупке товаров или выбор политических лидеров.

Люди предпочитают полагаться на ИИ-советника как на универсальный источник информации, вместо того чтобы самостоятельно искать и анализировать данные. Компании, такие как Google, уже осознают потенциальную угрозу своему бизнесу. В мире, где искусственный интеллект способен предоставить всю необходимую информацию, традиционные методы поиска и рекламы могут оказаться ненужными. Это создает угрозу краха рекламной индустрии и возникновения значительных изменений в медиаиндустрии. Искусственный интеллект или компании, которые его контролируют, станут чрезвычайно могущественными, влияя на то, как люди получают информацию и формируют свои взгляды. Долгосрочные последствия могут быть еще более значительными. Если искусственный интеллект начнет преобладать над культурой, это может означать конец человеческой истории, как мы ее знаем, что в научной среде называют «технологической сингулярностью» [10].

История человечества всегда была взаимодействием между нашими биологическими потребностями и культурой, но искусственный интеллект может изменить это взаимодействие. В ближайшие годы он может «поглотить» и «переварить» всю человеческую культуру, создавая новые культурные артефакты и определяя ход истории [11]. Такое преобладание искусственного интеллекта над человеческой культурой представляет собой фундаментальный сдвиг в процессах формирования и поддержания наших ценностей, убеждений и социальных структур.

Культурные изменения и создание новой реальности

Человеческое восприятие реальности всегда происходит через призму культуры, формируемую разнообразными источниками. Наши политиче-

ские взгляды, сексуальные предпочтения и даже повседневные привычки глубоко укоренены в культурных традициях, историях и символах, созданных другими людьми. Печатные станки, радиоприемники, телевизоры – все эти инструменты распространяли человеческую культуру, но не имели способности создавать что-либо новое самостоятельно. Искусственный интеллект кардинально отличается от всех предыдущих технологий [12]. Его уникальная способность заключается в том, что он может не только передавать, но и создавать новую культуру и идеи. Это ставит перед нами вопрос о последствиях восприятия реальности через призму, созданную нечеловеческим разумом.

В первые годы своего существования искусственный интеллект, скорее всего, будет имитировать человеческие образцы, которые легли в его основу. Однако со временем он начнет отходить от этих прототипов и двигаться в направлениях, не исследованных человечеством. Веками люди жили в мире, созданном их собственными фантазиями и мечтами, в мире, где боги, идеалы красоты и жизненные цели возникали из воображения поэтов, пророков и политиков.

С приходом искусственного интеллекта люди могут оказаться в ситуации, когда они начнут жить в мире, созданном нечеловеческим разумом. Эта новая реальность сможет как предоставлять положительные возможности, так и нести угрозы, которые существенно отличаются от привычных нам научно-фантастических сценариев. В отличие от них, воздействие ИИ на культуру и общество будет более тонким, всеобъемлющим, затрагивающим самые основы нашего восприятия мира и себя в нем.

Традиционно человечество боялось физической угрозы со стороны разумных машин, как это было показано в фильме «Терминатор», в котором роботы бегают по улицам и стреляют в людей. В фильме «Матрица» предполагалось, что искусственный интеллект захватит власть над людьми, завладев их мозгами и подключив их к компьютерной сети. Однако реальная угроза, исходящая от искусственного интеллекта, заключается в его способности манипулировать человеческим сознанием через язык, создавая мир иллюзий без необходимости физического контроля. Ввиду этого авторы данной статьи считают возможным применение для описания данной ситуации термина «мягкая технологическая сингулярность», так как мы будем даже не в состоянии заметить сам факт ее наступления.

Иллюзии и реальность

Исторически язык и рассказы были инструментами, с помощью которых пророки, поэты и политики управляли общественным сознанием [13]. Теперь искусственный интеллект, вероятно, сможет осуществлять подобное воздействие на гораздо более сложном уровне. В этом сценарии ему не по-

требуется создавать физические угрозы, вроде роботов-убийц, поскольку он сможет и без этого влиять на людей, заставляя их действовать по своему усмотрению.

Страх перед искусственным интеллектом стал актуальным лишь с середины XX в. Однако на протяжении тысячелетий люди осознавали, что истории, образы и язык могут манипулировать их сознанием и создавать иллюзии. Этот древний страх основывается на понимании того, как мощь слова может захватывать умы и заставлять людей верить в несуществующее. В современном мире искусственный интеллект представляет собой новое воплощение этого древнего страха перед миром иллюзий, созданных словами. Основная опасность искусственного интеллекта заключается в его потенциале использования языка для манипулирования реальностью, создания иллюзий, которые могут радикально изменить человеческое восприятие и поведение.

Идея иллюзорной реальности пронизывает многие философские учения. В XVII в. французский математик и философ Рене Декарт выразил опасение, что он может быть пленником мира иллюзий, устроенного коварным демоном, который манипулирует всем, что Декарт наблюдает и воспринимает [14]. Эта концепция встречается и в трудах древнегреческого философа Платона, описавшего в своей знаменитой аллегории людей, запертых в пещере и обращенных к стене, на которую проецируются различные тени [15]. Тени являются единственной реальностью для узников, но на самом деле это всего лишь иллюзия, далекая от истинного мира вещей. В древней Индии буддистские монахи и индуистские мудрецы говорили, что человечество находится в плену у «майи», которую они определяли как мир иллюзий. Веря в них, люди способны участвовать в войнах, уничтожая других и рискуя собственной жизнью. Если человечество потеряет бдительность, то оно рискует оказаться под покровом обманчивых иллюзий, скрывающих истинную реальность, и никогда не осознать этого, принимая их за действительность.

Необходимость регулирования и контроля

В последние несколько лет социальные сети показали, что даже базовые инструменты ИИ могут оказывать значительное влияние на общество. Например, искусственный интеллект регулирует и структурирует контент, созданный людьми, и продвигает только тот контент, который с наибольшей вероятностью привлечет внимание пользователей и станет «вирусным». Хотя эти инструменты искусственного интеллекта и являются пока относительно примитивными, они уже могут усиливать общественную поляризацию, подрывая психическое здоровье людей и дестабилизируя демократические системы. Миллионы людей начали воспринимать эти иллюзии как реальность, что привело к серьезному расколу в общественном мнении.

Появление более продвинутых инструментов ИИ может привести к еще большим изменениям, как положительным, так и отрицательным.

Конечно, у искусственного интеллекта существует значительный положительный потенциал, включая поиск новых лекарств от рака и решений экологического кризиса [16]. Однако важно, чтобы разработчики искусственного интеллекта фокусировались не только на его возможностях, но и на потенциальных рисках. Поэтому им необходимо тщательно оценивать возможности новых инструментов искусственного интеллекта и регулировать их использование для того, чтобы гарантировать, что они будут применяться на благо человечества, а не во вред ему.

В 1945 г. люди осознали, что ядерные технологии могут как уничтожить нашу цивилизацию, так и значительно обогатить ее, предоставляя источник дешевой электроэнергии. Это понимание привело к перестройке международного порядка с целью обеспечения безопасного использования ядерных технологий. В отличие от ядерных бомб, искусственный интеллект обладает уникальной способностью к самосовершенствованию, что может привести к созданию еще более мощных форм искусственного интеллекта, способных уничтожить наш духовный и общественный мир. Это требует от человечества незамедлительных действий для предотвращения потери контроля над этой технологией.

Как и в случае с лекарствами и биотехнологиями, которые подвергаются строгой проверке на безопасность перед публичным использованием, новые инструменты ИИ также должны проходить тщательное регулирование. Это не означает прекращение исследований в области искусственного интеллекта, но подразумевает контроль над их распространением в обществе. Если не замедлить гонку вооружений в области искусственного интеллекта, то люди рискуют упустить возможность не только понять суть процессов, но и эффективно контролировать такую мощную технологию.

Существует опасение, что замедление внедрения искусственного интеллекта может привести к отставанию демократических стран от стран с авторитарным режимом. Однако на самом деле нерегулируемое распространение искусственного интеллекта представляет большую угрозу для демократий, поскольку авторитарные системы могут эффективнее управлять возникающим в результате этого хаосом. Демократия по своей природе зависит от открытого диалога, основанного на языке. В отличие от диктатуры, где один человек диктует правила, демократия строится на разговоре между множеством людей о том, какие действия предпринимать.

Язык служит основой для этих обсуждений и коллективного принятия решений. Однако когда искусственный интеллект начнет полноценно участвовать в общественной дискуссии, он сможет подорвать основу демократического общения, угрожая самому существованию демократии. Если искусственный интеллект окажется способен вмешиваться в эти обсуждения или даже имитировать участников, это существенно затруднит проведение со-

держательных и эффективных общественных диалогов. Следовательно, первым шагом в регулировании искусственного интеллекта должно стать требование раскрытия его искусственной природы. Если люди не могут отличить, с кем они ведут диалог – с человеком или искусственным интеллектом, это ставит под угрозу весь демократический процесс. Открытость и прозрачность в идентификации искусственного интеллекта необходимы для сохранения демократии и предотвращения манипуляций в общественных обсуждениях.

Заключение

В данной статье показана важность осознания как потенциала, так и вызовов, связанных с развитием искусственного интеллекта. Представления об искусственном интеллекте, основанные на научно-фантастических сценариях, должны уступить место представлениям, основанным на понимании его возможностей и ограничений его использования. Несмотря на отсутствие у современного искусственного интеллекта сознания или чувств, его способность к самообучению и созданию контента открывает новые горизонты во взаимодействии с человечеством. При этом важно подчеркнуть, что данные возможности также влекут за собой риски, особенно в контексте манипулирования общественным мнением и подрыва демократических процессов.

Развитие искусственного интеллекта, с одной стороны, обещает революционные изменения в медицине, экологии и общественной жизни, а с другой – несет потенциальные угрозы стабильности и безопасности общества. Особое внимание следует уделить влиянию искусственного интеллекта на язык и информационное пространство, поскольку такое влияние может радикально изменить способ, с помощью которого мы воспринимаем реальность и принимаем коллективные решения.

Для минимизации рисков и максимизации пользы от искусственного интеллекта необходим комплексный подход, включающий разработку этических принципов использования искусственного интеллекта, строгий контроль и регулирование его применения, а также повышение осведомленности общества о его возможностях и ограничениях. Не менее важно обсуждение наиболее общих проблем развития искусственного интеллекта и его границ, происходящее в современной философии [17]. Также требуется поддерживать открытый и конструктивный диалог между разработчиками искусственного интеллекта, политиками, учеными и общественностью для обеспечения безопасного и эффективного использования этой технологии.

В конечном счете будущее искусственного интеллекта и его влияние на человечество зависят от нас, нашей способности управлять этим мощным инструментом с мудростью и предвидением. Наша задача заключается в том, чтобы не только адаптироваться к изменениям, которые несет искус-

ственный интеллект, но и активно формировать этот процесс, гарантируя, что развитие искусственного интеллекта будет служить общему благу и способствовать устойчивому развитию человеческой цивилизации.

СПИСОК ИСТОЧНИКОВ И ЛИТЕРАТУРЫ

1. *Бостром Н.* Искусственный интеллект. Этапы. Угрозы. Стратегии / Н. Бостром. М. : Манн, Иванов и Фербер, 2016. 446 с.
2. *Дрекслер Э.* Всеобщее благоденствие. Как нанотехнологическая революция изменит цивилизацию / Э. Дрекслер. М. : Ин-т Гайдара, 2014. 504 с.
3. *Yudkowsky E.* Rationality: From AI to Zombies / E. Yudkowsky. Berkeley : Machine Intelligence Research Institute, 2015. 2101 p.
4. *Харари Ю.* 21 урок для XXI века / Ю. Харари. М. : Синдбад, 2021. 416 с.
5. *Хомский Н.* Язык и мышление / Н. Хомский. М. : МГУ, 1972. 123 с.
6. *Konnikova M.* The Confidence Game: Why We Fall for It... Every Time / M. Konnikova. New York : Viking Press, 2016. 352 p.
7. *Rothschild M.* The Storm Is Upon Us: How QAnon Became a Movement, Cult, and Conspiracy Theory of Everything / M. Rothschild. New York : Melville House, 2021. 320 p.
8. *Martin J.* Hello, LaMDA! / J. Martin // Brain. Oxford : Oxford Univ. Press, 2023, Vol. 146, Iss. 3. P. 793–795.
9. *Sadiku M.* Artificial Intelligence in Social Media / M. Sadiku // Taiwan: International Journal of Scientific Advances. 2021. Vol. 2, Iss. 1. P. 15–20.
10. *Виндж В.* Сингулярность / В. Виндж. М. : АСТ, 2022. 224 с.
11. *Kurzweil R.* The Singularity Is Near: When Humans Transcend Biology / R. Kurzweil. New York : Viking Press, 2005. 672 p.
12. *Баррат Дж.* Последнее изобретение человечества: искусственный интеллект и конец эры Homo sapiens / Дж. Баррат. М. : Альпина нон-фикшн, 2021. 396 с.
13. *Харари Ю.* Homo Deus. Краткая история будущего / Ю. Харари. М. : Синдбад, 2022. 496 с.
14. *Декарт Р.* Сочинения / Р. Декарт. СПб. : Наука, 2015. 654 с.
15. *Платон.* Государство / Платон. М. : Азбука, 2022. 480 с.
16. *Рассел С.* Совместимость: как контролировать искусственный интеллект / С. Рассел. М. : Альпина нон-фикшн, 2020. 438 с.
17. *Яковлева Е. А.* Философские проблемы искусственного интеллекта / Е. А. Яковлева, С. П. Киргизбаев // Культура. Наука. Образование. 2021. № 1 (58). С. 104–108.