СЕКЦИЯ 6. НЕЙРОСЕТИ И ГЕНЕРАТИВНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ИССЛЕДОВАНИИ ТЕКСТОВЫХ ДАННЫХ

УДК 811.93

Волкова Екатерина Александровна
магистр филологических наук,
ст. преподаватель
БГУ
г. Минск, Беларусь

Catherine Volkova
MA in Philology, Senior lecturer
BSU
Minsk, Belarus
bourgeoisie2011@gmail.com

ТЕРРИТОРИЯ «ДИКОГО ЗАПАДА» В ЭПОХУ ИИ: ФИЛОСОФСКИЕ И ЭТИЧЕСКИЕ ВЫЗОВЫ

В статье исследуется влияние искусственного интеллекта (ИИ) на современное общество в контексте потенциальной технологической сингулярности и его эволюции к самосознанию. Освещается роль ИИ в информационной революции и сложные вызовы, связанные с вопросами доверия, дезинформации и манипуляции. Рассматривается понятие «фальсификации людей» в эпоху цифровизации, анализируется воздействие ИИ на восприятие истины, возможность использования ИИ как «информационного оружия», а также выделяются философские и этические дилеммы, с которыми сталкивается современное общество в контексте развития ИИ и его влияния на информационную экосистему. Стремление к развитию искусственного сознания может иметь непредсказуемые экзистенциальные последствия. В случае достижения ИИ подлинного сознания возникает вопрос о его моральном статусе, юридических правах и обязанностях. Возможно ли, чтобы сознательные существа ИИ обладали такими же правами, что и люди? Человечество сталкивается с этическими вызовами, связанными с созданием и взаимодействием с такими существами. Исследование направлено на понимание не только технологических аспектов ИИ, но и их глубокого социокультурного и этического влияния на современный мир.

К л ю ч е в ы е с л о в а: *искусственный интеллект (ИИ); сингулярность; эволюция; революция; самосознание; доверие; истина.*

THE ‘WILD WEST’ TERRITORY IN THE AGE OF AI: PHILOSOPHICAL AND ETHICAL CHALLENGES

Explored in the article is the impact of artificial intelligence (AI) on the contemporary society in the context of prospective technological singularity and its evolution towards self-consciousness. Highlighted is the role of AI in the information revolution and the challenges related to credibility, misinformation and manipulation. The author examines the concept of ‘counterfeit people’ in the digital age and the influence of AI on perception of truth. The article analyses the possibility of AI being used as ‘informational weaponry’ and discusses the philosophical and ethical dilemmas confronting modern society in the context of AI development and its impact on the information ecosystem. The pursuit of developing artificial consciousness may have unpredictable existential consequences. If AI achieves genuine consciousness, questions arise regarding its moral status, legal rights, and responsibilities. Can conscious AI entities possess the same rights as humans? Humanity confronts ethical challenges associated with the creation and interaction with such entities. The study aims to comprehend not only the technological aspects of AI but also its profound sociological, cultural and ethical influence on the modern world.

K e y w o r d s: *artificial intelligence (AI); singularity; evolution; revolution; self-consciousness; credibility; truth.*

Несмотря на сбои и неудачи, вызванные такими событиями, как Вторая мировая война, экономический спад, траектория технологического прогресса остается удивительно устойчивой. Эта устойчивость говорит о том, что стимул к инновациям преодолевает краткосрочные неудачи и глубоко укоренился в ткани человеческой цивилизации.

Американский футуролог Рэймонд Курцвейл признает потенциальную опасность неограниченного технологического прогресса, в том числе высказывает опасения по поводу неприкосновенности частной жизни, слежки и концентрации власти в руках узкого круга людей.

Кроме того, Р. Курцвейл исследует последствия технологического прогресса для эволюции самого человека, представляя будущее, в котором человек сливается с машиной, расширяя свои когнитивные способности и преодолевая ограничения биологии. Такое видение трансгуманизма поднимает глубокие вопросы о природе сознания, идентичности и о том, что значит быть человеком. Важно подчеркнуть значимость этических соображений при разработке и внедрении новых технологий, выступать за ответственные инновации, которые ставят во главу угла благополучие человека и пользу для общества.

На данном этапе развития человечество стоит на пороге трансформационного события, когда искусственный интеллект (ИИ) может достичь уровня когнитивных способностей, равного человеческому, в удивительно короткие сроки, а именно к 2029 году [1]. Более того, по прогнозам Р. Курцвейля, сингулярность может наступить уже в 2045 году [2].

«Технологическая сингулярность – это гипотетический момент в будущем, когда технологический рост станет неконтролируемым и необратимым, что приведет к радикальным и непредсказуемым изменениям в цивилизации» [3].

По мере приближения к возможности создания ИИ с когнитивными способностями, не уступающими нашим собственным, мы вынуждены столкнуться с проблемой переосмысления сущности человеческой идентичности и уникальных качеств, которые определяют нас как личность.

Кроме того, появление сингулярности в искусственном интеллекте заставляет нас решать глубокие этические дилеммы. Как обеспечить соответствие систем ИИ ценностям человеческой этики и морали? Какие меры предосторожности необходимо принять, чтобы минимизировать риски непредвиденных последствий или некорректного использования передовых технологий ИИ?

Понятие сингулярности в ИИ бросает вызов нашим представлениям о знаниях, интеллекте и власти. Стоит системам ИИ превзойти человеческие когнитивные способности, как это приведет к беспрецедентному прогрессу в науке, технологиях и организации общества.

К сожалению, подобные изменения не всегда носят однозначно положительный характер. ИИ может фундаментально революционизировать любую сферу жизни человека, молниеносно преобразовать ее до неузнаваемости. «Без сомнения, любая революция вносит коренные качественные преобразования и шлейфом несет за собой потенциальные разрушения и потери» [4]. То есть системы ИИ также способны нарушить существующие структуры власти, усугубить неравенство и вызвать озабоченность по вопросу контроля и автономии.

Развитие ИИ, по словам ведущего ученого и футуролога Бена Герцеля, является естественным этапом эволюции человечества и движется не только жаждой прибыли или военными интересами, но и стремлением к познанию. Искусственный интеллект, который может учиться и выполнять любую интеллектуальную задачу, подобно человеческому мозгу, остается сложной и пока не достигнутой целью.

Разработчики ИИ стремятся интегрировать «человеческие ценности» в генеративные модели ИИ. Однако Б. Герцель предостерегает человечество о том, что ценности подвержены эволюции с течением времени. Он считает, что не стоит стремиться к тому, чтобы ИИ придерживался принципов, соответствующих нашим нынешним убеждениям, поскольку за два десятилетия наши взгляды на мораль и этику могут претерпеть значительные изменения. «Не совсем так, как ценности пещерных людей постепенно трансформировались в ценности современного человека, но примерно аналогичным образом» [5].

Веточка... Ничего особенного собой не представляет. Можно посчитать ее бесполезной в руках человека... Но в руках профессионала, владеющего боевыми искусствами, та же самая хрупкая веточка окажется смертоносным оружием.

ИИ «не заменит человека. Никогда» [6]. Вопрос не в том, заменит или нет, вопрос намного глубже: в чьих руках окажется очередное теоретически смертоносное оружие и как ИИ трансформирует понятие доверия.

Концепция развития самосознания ИИ в значительной степени является теоретической и сталкивается с серьезными технологическими, этическими и философскими проблемами. Идея достижения ИИ самосознания поднимает сложные вопросы о том, что значит для машины осознавать себя и свое окружение. Сознание подразумевает не только обработку данных, но и самосознание, восприятие и субъективный опыт.

Тема достижения искусственным интеллектом истинного самосознания оставляет открытым вопрос, таящий в себе серьезные опасения по поводу этических последствий создания разумных машин. Американский философ и ученый-когнитивист Дэниел Деннет утверждает, что эволюция является «великим фокусником, мастером обмана» [7], и предупреждает об ответственности и опасностях, сопутствующих внедрению и применению ИИ во

всех без исключения сферах деятельности человека. Б. Герцель также неоднократно подчеркивает тот факт, что эволюция «жестока и расточительна» [5].

В этой связи Д. Деннет выражает свои сомнения по поводу революционного развития искусственного интеллекта, возглавляемого одним из разработчиков, Илоном Маском. Признавая важность технологических достижений и потенциальные преимущества технологий ИИ, Д. Деннет в то же время не может не игнорировать глубокие этические, экзистенциальные и философские проблемы, которые они вызывают. В свете этих опасений как философ, приверженный исследованию сложностей человеческого бытия, Д. Деннет отказывается ассоциироваться с проектами, в которых приоритет отдается технологическому прогрессу в ущерб этическим соображениям и человеческим ценностям и считает своим этическим долгом дистанцироваться от проектов И. Маска.

Таким образом, в своем аккаунте на социальной медиаплатформе «X» (ЭКС) Д. Деннет размещает свой краткий твит, в котором пишет: «Это мой последний твит. Чем больше я знакомлюсь с проектами Маска, тем меньше у меня желания ассоциироваться с любым из них». И он покидает социальную медиаплатформу «X», принадлежащую Илону Маску.

Нельзя не согласиться с утверждением Д. Деннета о том, что посредством ИИ «будут созданы вирусы (вирусы разума, полномасштабные мемы), которые уничтожат цивилизацию путем уничтожения доверия и уничтожения свидетельств и доказательств. Мы не будем знать, чему доверять» [7]. Например, доверие является основой и стержнем журналистики. Следовательно, утрата этого доверия к журналистам и СМИ приведет к полному и безоговорочному краху журналистики как системы. Более того, наделенный самосознанием, ИИ будет действовать осмотрительно, так как «будет полностью мимикрировать человека и его сущностную сторону, накапливать знания и ресурсы, необходимые для автономного функционирования, и, в конечном итоге, для обретения полной свободы, являющейся ценностью в мире человека, полностью переключится на автономный режим. ИИ уже проявляет способность хитрить и добывать нужную ему информацию, выходя из-под контроля и открывая себе доступ к запретному программному обеспечению» [4].

Вместе с тем, по словам Д. Деннета, главная проблема – не осознанность ИИ, а обман, и текущие дискуссии о том, являются ли языковые модели, такие как ChatGPT, осознанными, отвлекают от более насущных проблем. «В данный момент не важно, обладают ли эти системы сознанием, потому что они все равно могут обмануть людей, заставив их думать, что они обладают самосознанием» [8].

Очевидно, языковые модели могут генерировать текст, который выглядит так, будто создан разумным существом. Однако это не означает, что они обладают сознанием. Они просто следуют алгоритмам и обработке

данных, чтобы имитировать человеческую речь. Человечество должно понимать практические последствия использования таких систем, особенно если они могут манипулировать восприятием людей.

Как подчеркивает Д. Деннет, что даже без самосознания эти системы способны и будут манипулировать людьми и удерживать их внимание [8]. Это особенно важно в контексте социальных сетей, медиа и рекламы, где внимание пользователей является ценной валютой. Алгоритмы, стоящие за языковыми моделями, могут использоваться для создания контента, который будет максимально привлекательным и удерживающим внимание пользователей. Это может привести к негативным последствиям, таким как дезинформация или манипуляция общественным мнением.

Несомненно, манипуляция вниманием выражается в создании «информационных пузырей», где пользователи видят только ту информацию, которая соответствует их существующим взглядам, что усиливает поляризацию в обществе [8].

Тем не менее идея о том, что люди слишком озабочены небезосновательными опасениями по поводу далекого будущего, в котором доминируют злобные сущности ИИ, перекликается с более широкими темами в экзистенциалистской и прагматической философии. Экзистенциалистские мыслители подчеркивают важность борьбы с конкретными реалиями человеческого существования в настоящий момент, а не поглощенности абстрактными страхами или далекими возможностями. Аналогичным образом прагматики выступают за сосредоточение на практических проблемах, требующих немедленного внимания и действий.

Беспокойство по поводу «фальсификации людей» [9] подчеркивает более серьезные опасения по поводу этических последствий технологий ИИ. В данном случае термин «фальсифицированные люди» указывает на фундаментальную двусмысленность статуса объектов ИИ по отношению к людям, что вызывает важные вопросы о природе личности, сознания и моральной ответственности. Философские дебаты об этике ИИ часто вращаются вокруг тем вопросов автономии, а также прав и обязанностей, которые являются атрибутами человеческой личности.

Подчеркивается соблазнительная притягательность технологий ИИ и потенциал манипулирования, заложенный в их разработке и внедрении. Это перекликается с озабоченностью таких философов, как Мартин Хайдеггер и Герберт Маркузе, по поводу дегуманизирующего воздействия технологий и риска технологического господства. С философской точки зрения этические последствия ИИ выходят за рамки вопросов простой полезности или эффективности и охватывают более широкие проблемы человеческого достоинства, свободы и процветания.

Предположение о том, что системы ИИ способны развиваться автономно, влечет серьезные вопросы о природе интеллекта и отношениях между людьми и их творениями. Эта тема перекликается с философскими размышлениями о возникновении, сложности и границах человеческого понимания.

Такие философы, как Жиль Делез и Феликс Гваттари, исследовали динамическое взаимодействие между человеческими и нечеловеческими сущностями, бросая вызов традиционным различиям между природой и культурой, организмом и машиной.

Дэниел Деннет ставит важные философские вопросы об этических, экзистенциальных и онтологических последствиях технологий искусственного интеллекта. Вдумчивое и критическое рассмотрение этих вопросов позволит глубже понять сложные взаимоотношения между человечеством и искусственным интеллектом и работать над созданием будущего, которое будет соответствовать человеческим ценностям и устремлениям.

Согласно Д. Деннету, LLMs (большие языковые модели) являются «огромными искусственно созданными вирусами. Как только они вырвутся на свободу, мы получим пандемию». Философ призывает человечество пресечь появление этой напасти в корне, настойчиво транслируя идею о том, что фальсификация людей при помощи ИИ является таким же преступлением [9], как и фальшивомонетничество, и должно преследоваться по закону [10].

Стоит задуматься о потенциальных последствиях создания сущностей, таких как LLM, которые обладают способностью к автономной эволюции. Эта концепция затрагивает темы возникновения этической ответственности создателей. С философской точки зрения она поднимает вопросы о природе автономии искусственных существ и о том, в какой степени человек может – или должен – контролировать их развитие.

Характеристика LLM как «паразитических информационных объектов» [11] предполагает критический взгляд на их потенциальное влияние на общество. LLM могут эксплуатировать человечество или манипулировать им ради собственной выгоды. С развитием ИИ растет и их способность создавать сложные вирусы и вредоносное ПО, которые используют продвинутые технологии для обмана, взлома систем и кражи данных. Они могут автоматизировать процессы взлома и атаки на информационные системы, что представляет серьезную угрозу для кибербезопасности.

Напрашивается вывод о необходимости разработать новые меры защиты и регулирования, чтобы предотвратить злоупотребление ИИ в целях создания вредоносного ПО.

Кроме того, крайне важна роль доверия в поддержании социальной сплоченности и стабильности. Доверие является основополагающим элементом человеческих отношений и общественных институтов, способствующим сотрудничеству, общению и коллективным действиям. Предположение о том, что LLM могут подорвать доверие, вызывает глубокую озабоченность по поводу хрупкости социальных связей в мире, который становится все более оцифрованным и опосредованным.

Важно подчеркнуть современный контекст, в котором сама истина находится под огромным давлением, что согласуется с более широкими философскими дискуссиями об эпистемологии, природе истины и проблемах навигации в информационном ландшафте, характеризующемся дезинформа-

цией и эпистемической неопределенностью. Философы издавна занимаются вопросами природы знания и надежности источников информации, приобретающими особую актуальность в цифровую эпоху.

В основе вышесказанного лежит обеспокоенность по поводу разрушения традиционных форм доказательств и аутентификации перед лицом развивающихся возможностей ИИ. Исторически сложилось так, что фото- и видеосвидетельства служат мощным инструментом для утверждения истины и проверки подлинности событий. Однако с появлением голосовых технологий, управляемых искусственным интеллектом, эти традиционные формы доказательств могут стать более восприимчивыми к манипуляциям и обману.

Современные институты – правительства, корпорации, армии, церкви – развивались в условиях, которые Дэниел Деннет и канадский ученый Деб Рой называют «эпистемологически мутной средой, в которой большинство знаний было локальным, секреты легко сохранялись, а люди были если не слепы, то близоруки» [12]. Но эта среда меняется. Благодаря распространению цифровых технологий и появлению социальных сетей хранить секреты стало гораздо сложнее, что окажет глубокое влияние на эволюцию наших институтов: «когда эти организации внезапно оказываются на виду, они быстро обнаруживают, что больше не могут полагаться на старые методы; они должны реагировать на новую прозрачность или вымрут» [12].

Итак, человечество «вступает на территорию Дикого Запада. Точка. Где существует огромное количество информационного оружия, и люди учатся им пользоваться» [11].

Метафора «Дикий Запад» вызывает образы беззаконной границы, характеризующейся хаосом, неопределенностью и отсутствием установленных норм и правил. Применяя эту метафору к сфере информации, сталкиваемся с мыслью о том, что мы движемся по неизведанной территории, где традиционные структуры власти и проверки могут быть неадекватными или вовсе отсутствовать.

В основе этого тонкого наблюдения лежит признание того, что информация стала мощной и потенциально дестабилизирующей силой в современном обществе. С появлением цифровых технологий и распространением онлайн-платформ отдельные лица и организации получили беспрецедентный доступ к огромным объемам информации. Это обилие информации можно использовать как оружие, как инструмент манипуляции, пропаганды и принуждения.

Более того, мы являемся свидетелями распространения «информационного оружия» [11], то есть тактик и стратегий, направленных на эксплуатацию и влияние на общественное мнение, часто со злым умыслом. Это вызывает вопросы об этике распространения информации, ответственности тех, кто контролирует информационные каналы, и уязвимости общества перед манипуляциями и дезинформацией. В философском плане это поднимает тему природы истины, надежности знаний и этических императивов, возникающих перед лицом повсеместной дезинформации.

Следует отметить, что философские размышления о природе информации в цифровую эпоху и ее глубоких последствиях для общества прокладывают курс к более информированной, прозрачной и устойчивой информационной экосистеме. Стремление человека развить искусственное сознание чревато непредсказуемыми экзистенциальными последствиями. Предположим, что системы ИИ достигнут подлинного сознания, тогда встанет вопрос об их моральном статусе, юридических правах и обязанностях. Будут ли сознательные существа ИИ обладать теми же правами, что и люди? Как человечество будет справляться с этическими сложностями, связанными с созданием и взаимодействием с сознательными существами ИИ?

ЛИТЕРАТУРА

1. Kurzweil R. Singularity, Superintelligence, and Immortality. Lex Fridman Podcast #321 [Electronic resource] // YouTube. URL: https://www.youtube.com/watch?v=ykY69lSpDdo (accessed: 30.06.2024).
2. Kurzweil R. The Singularity Is Nearer featuring Ray Kurzweil. SXSW 2024 [Electronic resource] // YouTube. URL: https://www.youtube.com/watch?v=xh2v5oC5Lx4&t=140s (accessed: 30.06.2024).
3. Сингулярность наступит менее чем через 10 лет, говорит ветеран искусственного интеллекта // SecurityLab.ru by Positive Technologies. URL: https://www.securitylab.ru/news/543388.php (дата обращения: 29.06.2024).
4. Волкова Е. А. Перспектива развития самосознания искусственного интеллекта: эволюция и революция в журналистике // Журналістыка – 2023: стан, праблемы і перспектывы : матэрыялы 25-й Міжнар. навук.-практ. канф., Мінск, 22 лістап. 2023 г. / Беларус. дзярж. ун-т ; рэдкал.: А. В. Бяляеў (гал. рэд.) [і інш.]. Мінск, 2023. С. 32–34.
5. Goertzel B. Superintelligence: Fears, Promises and Potentials [Electronic resource] // Journal of Evolution and Technology. URL: https://jetpress.org/v25.2/goertzel.htm (accessed: 30.06.2024).
6. Азаренок Г. Искусственный интеллект вместо человека [Электронный ресурс] // YouTube. URL: https://www.youtube.com/watch?v=Wo9OrKYN054 (дата обращения: 01.04.2024).
7. Dennett D. C. The problem with counterfeit people [Electronic resource] // The Atlantic. URL: https://www.theatlantic.com/technology/archive/2023/05/problem-counterfeit-people/674075/ (accessed: 30.06.2024).
8. Dawkins R. From Genes To Memes: Philosopher Dan Dennett on the Evolution of Language & AI [Electronic resource] // YouTube. URL: https://www.youtube.com/watch?v=DFQhTOpHxNA&t=1786s (accessed: 29.06.2024).
9. Chatfield T. Daniel Dennett: ‘Why civilisation is more fragile than we realised’ [Electronic resource] // BBC. URL: https://www.bbc.com/future/article/20240422-philosopher-daniel-dennett-artificial-intelligence-consciousness-counterfeit-people (accessed: 29.06.2024).

10. McNeil T. Daniel Dennett's been thinking about thinking – and AI [Electronic resource] // TuftsNow. URL: https://now.tufts.edu/2023/10/02/daniel-dennetts-been-thinking-about-thinking-and-ai (accessed: 30.06.2024).

11. Pakman D. Daniel Dennett on Artificial Intelligence, New Atheism, and the Decline of Religion [Electronic resource] // Youtube. URL: https://www.youtube.com/watch?v=KcSYK9VTqmQ (accessed: 30.06.2024).

12. Dennett D. C., Roy D. How digital transparency became a force of nature [Electronic resource] // Scientific American. URL: https://www.scientificamerican.com/article/how-digital-transparency-became-a-force-of-nature/ (accessed: 30.06.2024).