

«Искусственный интеллект» Vs «человеческий разум»: новая этика технологического развития

Г. А. Щербаков

Финансовый университет при Правительстве Российской Федерации
gascherbakov@fa.ru

Аннотация. Текущая ситуация в сфере развития технологий отмечается стремительным наращиванием цифровой компоненты научно-технического прогресса, в частности, созданием интеллектуальных систем, способных обеспечивать осуществление ряда функций, ранее выполнявшихся исключительно человеком. Указанный тезис касается внедрения технологий искусственного интеллекта, ориентированных на замещение функций, связанных с когнитивной деятельностью людей. Представляется, что общество в развитии научно-технического прогресса приблизилось к опасной черте, за которой неконтролируемое внедрение когнитивных технологий способно генерировать потенциальные риски. Данное обстоятельство требует выработки новой этики технологического развития, базирующийся на безусловном приоритете общественных интересов и безопасности индивида.

Ключевые слова: технологическое развитие; цифровизация; искусственный интеллект; математическое моделирование; системный анализ

I. ВВЕДЕНИЕ

Научно-технический прогресс всегда был ориентирован на полное или частичное воспроизведение функций человека техническими средствами. Как правило, к ним относились функции, предполагавшие применение мышечных усилий в процессе выполнения каких-либо работ. Данное направление связано с развитием робототехники, в настоящее время активно применяемой в производстве общественного продукта, а также в сферах, представляющих общественный интерес, но связанных с физическим риском для человека. Технологии, обеспечивавшие функционирование робототехники, оставляли безусловный приоритет в управленческих решениях за человеком и выступали исключительно в качестве вспомогательного инструментария. Субъектность по отношению к сферам жизнедеятельности оставалась за человеком, и это являлось основополагающим принципом развития технологий в предыдущий период.

Вместе с тем, в настоящее время данный процесс претерпевает качественную трансформацию, обусловленную созданием технологий, имитирующих сложную когнитивную деятельность человека, такую как: распознавание речи, создание сложных вербальных конструкций, математические измерения и пр. Творческим ориентиром создателей новых технологий выступает воспроизведение способностей индивида в области анализа и интерпретации данных, а также принятие на основе поступающей информации тех или иных управленческих решений. Указанный процесс чреват постепенным превращением человека из субъекта в объект когнитивной деятельности, что переводит

перспективы развития технологической сферы в этическую плоскость.

II. ОБЗОР ЛИТЕРАТУРЫ И МЕТОДОВ

Еще в трудах Т. Гоббса «Человеческая природа» 1640 года [1] и Р. Декарта «Рассуждение о методе» 1647 года [2] можно обнаружить методологические основы построения «умных машин». «Концепция первой аналитической машины – сложного цифрового калькулятора британского математика Ч. Бэббиджа – появилась еще в 1830-е годы [3, С. 44]. И уже в 1914 году испанец Л. Торрес Кеведо разработал электромеханическое устройство, способное разыгрывать простейшие шахматные партии» [4, С. 172–175].

Спустя три десятилетия увидели свет труды британского математика А. Тьюринга, а также на практике была реализована идея создания т. н. «Baby Machine», машинное обучение которой производилось по методике обучения детей, но с использованием данных, предназначенных для использования взрослым человеком. В 1943 году американским нейропсихологом У. Мак-Каллоком и его учеником У. Питтсом была предложена математическая модель искусственной нейронной сети:

$$S = \sum_{i=1}^n X_i \cdot w_i,$$

прописанная через функцию нейрона $Y = F(C)$ как пороговое устройство, обладающее множеством входов и единственным (бинарным) выходом.

И, наконец, в 1956 году состоялся Дартмутский научный семинар, в ходе которого американский математик Дж. Маккарти ввел в публичный оборот термин «искусственный интеллект», охарактеризовав его как «науку и технику для создания интеллектуальных машин, подобных человеку, и особенно интеллектуальных компьютерных программ» [5].

Два десятилетия, последовавшие после триумфальной популяризации концепции искусственного интеллекта, отмечены некоторым падением интереса к данной тематике, обусловленным тем, что вычислительный потенциал имевшихся на тот момент вспомогательных средств обработки информации оказался не соответствующим уровню задач, выдвигаемых авторами концепции создания интеллектуальных систем.

С середины 1990-х годов – с началом активного наращивания вычислительных возможностей

компьютерной техники – интерес к развитию систем искусственного интеллекта начал восстанавливаться, а дополнительным мощным стимулом для форсированного проведения указанных работ явилось поражение в 1997 году в шахматной партии действующего чемпиона мира Г. Каспарова, нанесенное ему суперкомпьютером IBM Deep Blue.

Вместе с тем, реальный качественный прорыв в создании систем искусственного интеллекта был осуществлен в первом десятилетии XXI века. «Математический научный мир нашел новые теории и модели обучения многослойных нейронных сетей, ставших фундаментом развития теории глубокого машинного обучения, а ИТ-отрасль стала выпускать высокопроизводительные, и, что главное, недорогие и доступные вычислительные системы» [6].

К настоящему моменту «накоплены и систематизированы самые разнообразные подходы и математические алгоритмы для построения систем искусственного интеллекта, такие как байесовские методы, логистическая регрессия, метод опорных векторов, решающие деревья, ансамбли алгоритмов и т.д. Одна из наиболее современных и удачных реализаций – это решения, построенные на технологии глубоких нейронных сетей (deep neural networks) и глубокого машинного обучения (deep learning)» [6].

III. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ТЕХНОЛОГИЧЕСКАЯ СТРУКТУРА

Сам процесс построения систем искусственного интеллекта включает в настоящее время комплекс

предметных областей, имеющих к технологии искусственного интеллекта в большей мере практическое отношение.

Технологическую структуру современных систем искусственного интеллекта можно графически представить в виде пирамиды, включающей три взаимозависимых уровня (рисунок).

1. Уровень применения. Он представляет собой вершину пирамиды и отвечает за области применения системы искусственного интеллекта, то есть за вопросы практического использования указанной системы в реальной жизни. Такими областями применения на практике могут выступать: безопасность, контроль за производственными процессами, автономное управление транспортными средствами, сферы использования самоуправляемой робототехники и пр. [3, С. 31].

2. Уровень технологии. Данный уровень отвечает за непосредственное получение эффекта от вычислительных и сенсорных возможностей системы искусственного интеллекта.

3. Уровень инфраструктуры. Этот уровень отвечает за работоспособность всей совокупности электронных компонентов, обеспечивающих вычислительный процесс, а также за формирование информационного массива, являющегося рабочим материалом для вычислительных действий системы искусственного интеллекта.

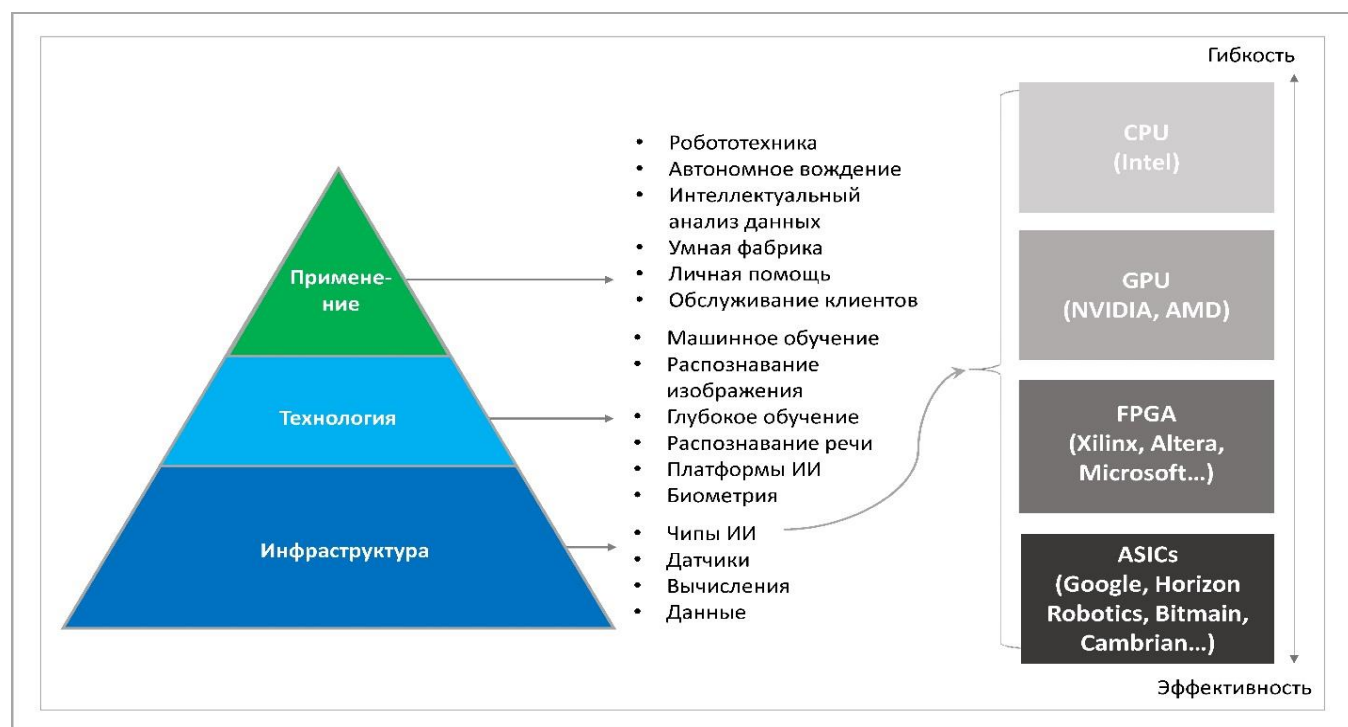


Рисунок. Технологическая структура современных систем искусственного интеллекта

В целом, искусственный интеллект понимается «как способность системы создавать в ходе самообучения программы (в первую очередь, эвристические) для решения задач определенного класса сложности и решать эти задачи» [7]. Вместе с тем, несмотря на множество подходов и методов, разработанных за прошедший период, ни один коллектив исследователей и

разработчиков не может претендовать на создание искусственного разума. Можно уверенно утверждать, что искусственного интеллекта как виртуального аналога человеческого разума в настоящее время не существует. Практически речь пока еще идет о создании алгоритмов глубокого машинного обучения с использованием нейронных сетей, в результате которого система

наделается новой способностью «делать выводы» на основе новых массивов информации. Процесс вывода информации может происходить в «облаке» или на «пограничных» устройствах. Данный процесс также как и процессы обучения, нуждается в сложной инфраструктуре, находящейся на третьем уровне пирамиды строения искусственного интеллекта.

Технологическая возможность обработки большого количества информации обуславливает эффективность алгоритмов глубокого обучения при обработке неструктурированных данных. Например, «ImageNet, общий эталон для глубокого обучения моделей для всестороннего распознавания изображений, имеет доступ к более чем 14 млн изображений» [8]. Вместе с тем, подобная возможность может оказаться излишней для решения менее сложных задач. Если данные неполные или слишком простые, возникает опасность «переобучения» системы и потери способности качественного обобщения новых массивов данных. «В результате, модели глубокого обучения становятся не такими эффективными, как другие методы (например, форсированные деревья решений или линейные модели), используемые для большинства практических бизнес-задач с меньшими наборами данных и меньшим количеством функций» [9].

IV. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: НОВАЯ ЭТИКА ТЕХНОЛОГИЧЕСКОГО РАЗВИТИЯ

Степень «зрелости» систем искусственного интеллекта можно оценить по уровню сочетания метода машинного обучения и метода рассуждений [10]. «На этом основании можно выделить следующие уровни развития искусственного интеллекта:

- первый уровень зрелости – включает только машинное обучение, поэтому его операционное влияние ограничено конкретными областями применения. Уроки, извлеченные из конкретных условий, нельзя применить к новым областям, а резкие изменения входных параметров снижают функциональность алгоритмов;
- второй уровень зрелости – расширяет операционное воздействие и включает несколько областей применения, а это означает, что обучение может быть применено к новым, неизведанным параметрам без вмешательства человека. Здесь вступают в действие простые рассуждения, позволяющие использовать расширенную область мощных алгоритмов машинного обучения;
- третий уровень зрелости – охватывает полностью самосознательную систему, способную связать свою мудрость и креативность с оперативными проблемами в любой области. На этом (еще гипотетическом) уровне искусственный интеллект способен превзойти человека в любой области» [9].

По сравнению с интеллектом человека система искусственного интеллекта представляет собой принципиально другой вид когнитивной деятельности. Современная «мыслящая машина» в упрощенном представлении – компьютер, обрабатывающий информацию и принимающий решения, которые внешне могут казаться осмысленными или разумными. Согласно

концепции А. Тьюринга, искусственный интеллект имитирует людей в процессе подготовки и принятия решений. Такой вид интеллекта весьма полезен в организационной деятельности. «Искусственный интеллект, обладающий способностью имитировать когнитивное поведение человека и автоматизировать процессы выявления и решения сложных проблем, предлагает возможности для повышения эффективности человеческой деятельности за счет извлечения необходимой информации из больших наборов данных и прогнозирования неожиданных событий» [11, 12]. Благодаря своим имитационным способностям он способен выявлять информационные шаблоны, которые оптимизируют тенденции, относящиеся к работе. Однако человек обладает когнитивными способностями, представляющими собой подлинный интеллект – разум. Находясь в открытой системе, человек, взаимодействуя с внешней средой, должен соответственно реагировать на экзогенные воздействия. Такой режим требует творческого подхода к формированию будущей стратегии, проявляющегося в способности правильно реагировать на внезапные изменения ситуации, предвидеть возможное развитие событий, а также корректно воспринимать искаженную информацию.

Реализованные разработки искусственного интеллекта пока еще представляют собой платформы для выполнения отдельных когнитивных задач, то есть проводимая исследовательская практика замыкается на создании алгоритмических моделей, способных выполнять, возможно весьма качественно, но всего лишь одну задачу.

Вместе с тем, развитие искусственных когнитивных систем не останавливается. В частности, сейчас по формуле AI + EI (Emotional intelligence) разрабатывается более продвинутая алгоритмическая модель, способная понимать эмоции собеседника с помощью внедренного модуля «Speech Emotion Recognition» [13]. Другая современная модель искусственного интеллекта, созданная компанией OpenAI (основатели: Илон Маск, Сэм Альтман) и получившая название GPT-3, способна создавать литературные тексты, писать стихи, делать письменные переводы, поддерживать общение, давать ответы на абстрактные вопросы. Создание данного продукта потребовало огромных инвестиций (более 1 млрд долларов США только в 2020 году) [5], и уже можно считать, что инвестиции принесли результат. Указанную модель, а также её продвинутый аналог GPT-4 можно к тому же использовать во многих сферах промышленного производства, дизайне, создании программных технологий.

Резонно предположить, что описанный выше разрыв между умственными способностями индивида и когнитивными возможностями систем искусственного интеллекта будет уменьшаться. Указанное предположение выдвигает в качестве первоочередной этической проблему, связанную с необходимостью установления пределов развития систем имитации когнитивных функций человека.

Проблемы этики развития современных технологий уже сегодня находятся в фокусе внимания целого ряда центральных направлений нового шестого технико-хозяйственного уклада. К примеру, в области биоинженерии, где в 1996 году был реализован первый случай клонирования живого млекопитающего

организма (овца Долли), не прекращаются жесткие споры относительно границ применения новых биотехнологий в практической сфере [14]. Все больше сторонников приобретает мнение, что и в работы по воспроизводству мыслительных функций индивида следует имплементировать элементы этического надзора. В частности, подобная позиция нашла отражение в недавнем (28.03.2023 г.) призыве более 1000 экспертов приостановить разработку и обучение нейросетей, как представляющих общественную опасность. В числе подписавших открытое письмо числятся такие авторитетные представители «цифрового» сектора мировой экономики, как И. Маск (Tesla, SpaceX, Twitter), Э. Шарп (Pinterest) и Ст. Возняк (Apple) [15].

Подобные требования получают весомые основания с учетом приоритетов двух основных групп непосредственных участников процесса создания когнитивной технологии.

Первая группа – коллектив разработчиков – концентрируется на возможностях, открывающихся по мере реализации прорывной идеи. Здесь доминирует эмоциональная компонента, определяемая такими нерациональными понятиями как «азарт» или «энтузиазм».

Другая группа участников, погруженных в процесс создания новой технологии – организаторы или лица, сопровождающие процесс (менеджеры, экономисты, финансисты), – склонны замечать, в первую очередь, рыночную компоненту новой технологии, то есть ее способность производить новую ценность. Уже к 2030 году рынок искусственного интеллекта, согласно прогнозам аналитиков, даст прибавку к мировому ВВП в размере 15,7 трлн долларов США [16]. Указанный факт определяет появление огромного и перспективного рынка.

Этические проблемы, связанные с внедрением когнитивной технологии, а также возможные сопутствующие риски не находятся в фокусе внимания разработчиков и управленцев, создающих новый продукт. К тому же эти люди могут просто не обладать необходимыми знаниями, чтобы полноценно учитывать этические риски, возникающие в результате внедрения перспективной идеи.

Указанные обстоятельства открывают проблемное поле, связанное с необходимостью уже на этапе стратегического планирования новой когнитивной технологии проводить мероприятия по выявлению сопутствующих непредвиденных последствий. Уже на этом этапе следует стремиться, чтобы эмоциональная составляющая созидательного процесса, а также эйфория от будущей рыночной востребованности новой технологии компенсировались рациональным понятием «ответственность». Необходим коренной пересмотр парадигмы создания передовых технологий, когда творческий порыв будет ограничиваться неангажированной, рискоориентированной логикой специалистов, обладающих специальными знаниями в тех сферах деятельности индивида и общества, которые могут стать объектом применения новой когнитивной технологии.

Одновременно необходимо учитывать временной и пространственный факторы, с учетом которых даже

наиболее оптимальные методы решения этических проблем не будут универсальными для всех случаев и не будут способны сохранять состоятельность в любом временном периоде. В связи с тем, что в процессе создания когнитивной технологии могут изменяться не только условия или задачи, но и цели, в организации-разработчике целесообразно иметь штатного сотрудника по контролю в области технологической этики, отслеживающего разработку технологии на всех этапах и располагающего необходимыми полномочиями для проведения ограничительных мероприятий. Указанные полномочия должны опираться на ответственную политику государства с обязательными для всех участников правилами. При этом соблюдение соответствующих «протоколов безопасности» должно контролироваться институтом независимых экспертов.

V. ЗАКЛЮЧЕНИЕ

Создание искусственных когнитивных систем «является идеей, длительное время находящейся в сфере научного интереса различных исследователей. Получив концептуальное оформление почти 70 лет назад на Дартмутской конференции, данная идея в течение нескольких десятилетий считалась некоей экзотикой, далекой от практического воплощения. И лишь в последние два десятилетия, когда соединились воедино две прогрессивные тенденции - цифровизация огромных массивов данных и революционный рост вычислительных способностей компьютерной техники, - концепция создания искусственного интеллекта приобрела не только возможности реального воплощения, но и перспективы практического применения в различных сферах общественной жизни. Вместе с тем, технологии искусственного интеллекта еще далеки от тех возможностей, которыми их наделяют оптимисты данного направления развития научно-технического прогресса» [9].

Тем не менее, активное развитие технологий, имитирующих мыслительные способности человека, подвело общество к опасной черте, за которой отсутствие контроля за их распространением, чревато формированием существенных рисков. Исключительно технико-внедренческая или рыночная логика в вопросах развития и применения когнитивных технологий способна принимать формат игры с нулевой суммой, где может быть только один победитель. Указанное обстоятельство свидетельствует о необходимости принятия уже на нынешнем этапе нового стиля «технологического» мышления, предусматривающего приоритет общественных интересов и безопасности индивида в общем комплексе возможностей, открывающихся в результате разработки когнитивной технологии.

СПИСОК ЛИТЕРАТУРЫ

- [1] Гоббс Т. Сочинения в 2 т. Т.1. М.: Мысль, 1989. 622 с. (Филос.насл. Т.107.). С.507-573.
- [2] Декарт Р. Рассуждение о методе, чтобы верно направлять свой разум и отыскивать истину в науках. М.: Академический проект, 2011. 335 с.
- [3] Гладких Б.А. Информатика от абак до интернета. Томск: Изд-во Науч.-техн. лит., 2005. 481 с.
- [4] Шилов В.В. Удивительная история информатики и автоматизации. М.: ЭНАС, 2013. 214 с.
- [5] Artificial Intelligence's First Message to Humanity: "Don't Be Afraid of Me" November 2020. URL: <https://arkasnews.com/artificial->

- intelligences-first-message-to-humanity-dont-be-afraid-of-me/ (дата обращения: 11.12.2021).
- [6] Гусев А.В., Добридюк С.Л. Искусственный интеллект в медицине и здравоохранении // Информационное общество, 2017. № 4-5. С. 78-93. URL: <https://webiomed.ai/blog/iskusstvennyi-intellekt-v-meditsine/> (дата обращения: 10.01.2023).
- [7] Искусственный интеллект. URL: <https://page.maple4.ru/inoc/stati/interesnoe/2495-iskusstvennyj-intellekt.html>. (дата обращения: 15.10.2021).
- [8] URL: <https://medium.com/botsupply/a-beginners-guide-to-deep-learning-5ee814cf7706> (дата обращения: 05.01.2022).
- [9] Щербаков Г. А. Искусственный интеллект: текущие проблемы развития технологии // Мягкие измерения и вычисления. 2022. Т. 51. № 2. С. 36-46.
- [10] Kaplan A, Haenlein M (2019) Siri, siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Bus Horizons* 62(1):15–25.
- [11] Lee J, Suh T, Roy D, Baucus M (2019) Emerging technology and business model innovation: the case of artificial intelligence. *J Open Innov Technol Market Complex* 5(3):44.
- [12] Zhuang Y-T, Wu F, Chen C, Pan Y-H (2017) Challenges and opportunities: from big data to knowledge in AI 2.0. *Front Inform Tech El* 18(1):3–14.
- [13] Теплицкий Д. EI+AI - формула будущего. URL: <https://bosfera.ru/dossier/dmitriy-teplickiy> (дата обращения: 10.10.2021).
- [14] Ammanath B. Thinking Through the Ethics of New Tech...Before There's a Problem / *Harvard Business Review*. 09.11.2021. URL: <https://hbr.org/2021/11/thinking-through-the-ethics-of-new-tech-before-theres-a-problem> (дата обращения: 10.11.2022).
- [15] Илон Маск призвал временно остановить улучшение нейросетей. URL: <https://ria.ru/20230329/razrabotka-1861573680.html> (дата обращения: 29.03.2023).
- [16] Cremer D., Kasparov G. AI Should Augment Human Intelligence, Not Replace It / *Harvard Business Review*. 18.03.2021. URL: <https://hbr.org/2021/03/ai-should-augment-human-intelligence-not-replace-it> (дата обращения: 11.01.2023).