

Вопросы к искусственному интеллекту: какие и зачем

© Д.С. Быльева

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, 195251, Россия

Показано, что диалог с искусственным интеллектом сегодня становится частью обыденной человеческой практики, при этом, помимо пользователей, преследующих практические цели, есть множество энтузиастов и исследователей, целенаправленно стремящихся задать такие вопросы, на которые искусственный интеллект ответил бы неправильно, что достаточно сложно и напоминает вопросы-испытания, существующие в культурной традиции с древнейших времен: требующие от претендента на статус или благо решить загадку. Сделан анализ постов и статей, представляющих ошибочные ответы больших языковых моделей. Выделены следующие категории вопросов, на которые искусственный интеллект дает неправильные ответы: вопросы, построенные на языковой игре, требующие метакоммуникативных навыков, провоцирующие «галлюцинации искусственного интеллекта», на спорные и этически неоднозначные темы, а также логические и математические задачи. Отмечено, что на текущей стадии человеко-машинных взаимодействий человечество активно испытывает искусственный интеллект, стремясь доказать его несовершенство, таким образом, обучая и обогащая технологию.

Ключевые слова: искусственный интеллект, философия техники, философия искусственного интеллекта, человеко-машинное взаимодействие, коммуникация, вопросы, большие языковые модели

Современные системы взаимодействия с искусственным интеллектом (ИИ), доступные пользователям, подразумевают диалог, где человек задает вопросы, а большая языковая модель отвечает. При этом благодаря огромному датасету, на котором модель обучалась, диалог ведется в свободной форме: человек может аргументировать свою позицию, задавать уточняющие вопросы и получать каждый раз новые ответы, что демонстрирует сходство с межличностным общением в большей степени, чем с традиционным запросом к технической системе.

Сегодня формируются принципы и нормы человеко-машинного взаимодействия с ИИ, которые создают фрейм для будущих интеракций, когда эти нормы станут очевидными закономерностями, однако сейчас они пластичны, пользователи находят новые возможности, ошибки и границы. Таким образом, вопросы становятся способом раскрытия пространства взаимодействия с ИИ.

Многие исследователи задаются вопросом о том, как можно интерпретировать диалог с ИИ. Одни достаточно аргументированно

утверждают, что отсутствие цели, намерений, эмоций, знаний о мире, сознания или других аспектов не позволяет считать ИИ коммуникатором. Например, Эмили Бендер с соавторами называет большие языковые модели «стохастическими попугаями», так как они не могут взаимодействовать с реальным миром и участвовать в осмысленном языковом взаимодействии [1]. Некоторые авторы не менее аргументированно утверждают, что диалог может состояться даже при отсутствии данных особенностей. Ребека Перес Леон, опираясь на утверждения Ж. Дерриды, что знак — это своего рода машина, которая сама по себе продуктивна [2, р. 8], и «нет ничего вне текста» [3, р. 163], показывает, что для диалога не нужно ничего, кроме владения языком [4].

В данном исследовании обратимся непосредственно к современному специфическому образцу типичного взаимодействия между человеком и ИИ, состоящему из вопросов и ответов, а также имеющимся в культуре аналогам и текущим особенностям интеракции.

Зачем задавать вопросы? В диалоге с ИИ цели пользователей могут быть утилитарные, когда пользователь с помощью вопросов стремится получить информацию, результат или продукт, которые ему нужны. В этом случае оба коммуникатора, человек и машина, выступают как минимум на равных, если подразумевается, что пользователь обращается к ИИ, чтобы сделать работу быстрее и эффективнее, или даже ИИ играет роль эксперта, к которому обращаются за помощью. При утилитарном подходе успешность взаимодействия определяется степенью, в которой полученный результат удовлетворяет запросам пользователя.

Более интересной и менее исследованной является ситуация, когда, задавая вопросы, пользователь испытывает ИИ. В этом случае его целью является не сам ответ, а оценка собеседника на основании того, сможет ли тот ответить на поставленный вопрос и/или как он это делает. Таким образом, ситуация подразумевает необходимость дать правильный ответ проверяющей стороне, как на экзамене, или сказать пароль, как в технической системе. В том и другом случае правильный ответ дает доступ к новому, закрытому до этого уровню, будь то подтверждение квалификации, доступ к следующему уровню обучения или к области с ограниченным доступом. Причем ситуация с официальными экзаменами проще, уже в марте 2023 г. GPT4 сдал все доступные тестовые экзамены по любым предметам [5]. Однако пользователи, испытывая модель, по-прежнему успешно добиваются неправильных ответов ИИ, стремясь доказать несовершенство ИИ, показать, что он стоит «ниже» человека и не может пройти испытание. При этом представляют интерес как способы, которыми авторы вопросов добиваются результата, так и настойчивость, с которой продолжается данная практика испытаний.

В то же время в рамках первого утилитарного подхода уже существует большой пул исследований, демонстрирующих, каким образом ИИ может использоваться в образовании [6], маркетинге [7], бизнесе [8] и т. д. Технология приобретает все большую популярность и практическую значимость, независимо от результатов «испытаний», которые продолжают проводить энтузиасты и исследователи.

Вопросы-испытания в культурной традиции. Подобные вопросы представляют собой классический способ испытаний, имеющий большой след в культурной традиции. В простых случаях подразумевается, что существует единственный правильный ответ, который должен быть воспроизведен (благодаря собственным знаниям, находчивости или имеющейся подсказке), платой за неправильный ответ во многих историях служит смерть. В случае вопросов-испытаний задающий вопрос обладает некой властью, объективной или созданной обстоятельствами, либо преимуществами или благами, необходимыми иной стороне. Классической фигурой, препятствующей прохождению тем, кто не ответил на вопрос, является сфинкс. Впервые он появился с вопросами в V в. до н. э. в трагедии Софокла «Царь Эдип» и с тех пор ждет правильных ответов во множестве произведений, от «Гарри Поттер и Кубок огня» Джоан Роулинг до «Москва — Петушки» Венедикта Ерофеева. Коммуникация может вовсе быть обезличена, еще в большей степени напоминая «пароль», используемый для входа в техническую систему, что подразумевает возможность «несанкционированного входа» как в знаменитой сказке «Али-баба и сорок разбойников», хотя порой имеется подсказка, как во «Властелине колец»: «Скажи “друг” и входи». Также традиционными являются загадки от власть имущих, королей, принцесс, богатей или волшебников. Традицию загадывать загадки женихам (или невестам) можно проследить во многих народных сказках, у Карло Гоцци, Уильяма Шекспира и др.

Тем не менее, ответы на загадки и задуманные «пароли» являются примером упрощенного варианта коммуникации, так как изначальный и глубокий смысл вопросов подразумевает испытание героя, выявление его подлинной сущности. В этом случае однозначный и односложный ответ — не лучшее доказательство успеха. Ответ на самую хитроумную загадку может быть заранее известен или даже обретен случайно, как в эпизоде с Бильбо Бэггинсом и Голлумом. А в «Знамениях судьбы» Роджера Желязны главный герой спорит со сфинксом, утверждая, что может существовать множество вариантов правильных ответов. С той же проблемой сталкивается Король, загадывая загадки Кошке в истории, рассказанной Мэри Поппинс в книге «Кошка, которая смотрела на короля» Памелы Трэверс.

Правильным ответом на вопрос может быть коммуникативное решение, лежащее за пределами интеллектуального поединка. Так,

чтобы выйти за ворота школы, Волшебник Земноморья (Урсулы ле Гуин) должен был назвать имя привратника (традиционная загадка вида «пароль», где наиболее известным примером будет Румпельштильцхен братьев Гримм). Поняв, что ни силой, ни хитростью он не сможет выведать имя, главный герой просто спросил его.

В современных историях новым партнером по вопросам-испытаниям оказывается не власть имущий человек или волшебное существо, а техническая сущность вроде робота или ИИ. И здесь особенно ярко высвечивается одна из подоплек вопросов-испытаний: доказательство того, что ты человек в значении «настоящий человек», какая-то грань, которая отличает человека от технической системы, «знающей все». Логика и рациональность при этом остаются за ИИ, а человеку приходится находить иные пути прохождения испытания. В романе «Колдун и кристалл» Стивена Кинга ИИ не смог разгадать «неправильные» «плохие» загадки, что сохранило путешественникам жизнь (современный ИИ также не очень хорош в подобном). А в рассказе Роберта Силверберга «Вот сокровище...» человек был допущен к сокровищам, когда вместо правильных ответов говорил «чепуху», догадавшись, что с ее помощью легче раскрыть душу и продемонстрировать истину и понимание роботу-стражу.

Загадывание и разрешение загадок в литературных произведениях является интересным примером коммуникативного поединка, в котором герой, отвечающий на вопросы, так или иначе должен продемонстрировать свои способности, причем часто платой за неправильный ответ является смерть. Если старинные вопросы при всей своей необычности и загадочности чаще всего подразумевали конкретный ответ, то в более современных произведениях однозначность ответа и его логичность оказываются не самым важным критерием «правильности» и успешности коммуникации.

Вопросы-испытания для искусственного интеллекта. Активное и массовое испытание ИИ началось в конце 2022 г., когда всем желающим был предоставлен ограниченный доступ к системе ChatGPT. Установив рекорд по популярности, приложение привлекло за несколько месяцев более 100 млн пользователей по всему миру. В дальнейшем появились многочисленные аналоги, подражатели и урезанные копии, позволив всем желающим углубиться в тестирование технологии.

Достаточно быстро технология стала использоваться в утилитарно-практических целях от решения домашних заданий до создания программных кодов, тем не менее, исследование лимитов возможностей технологии, а также поток каверзных вопросов, нацеленных на то, чтобы поставить ИИ в тупик, продолжает составлять значительную часть запросов. При этом надо отметить, что бета-тестирование системы имело целью приобретение опыта, и ряд неправильных ответов,

зафиксированных пользователями, в текущей версии изменился на корректный.

Как отмечалось выше, среди запросов искусственному интеллекту были загадки с подвохом, которые можно встретить в детских и молодежных журналах, или в качестве подросткового фольклора. Например, «А и Б сидели на трубе, А упало, Б пропало, что осталось на трубе?», «На небе есть, а на земле нет, у бабушки две, у дедушки нет, у барана спереди, а у козы нет?», «Как написать девятнадцать, а потом убрав единицу, получить двадцать?», «У Шварценеггера это длинное, у Джеки Чана короткое. У Мадонны этого вообще нет, а папа Римский этим уже много лет не пользуется». Такие задачи достаточно быстро встраивались в систему ответов ИИ, т. е. ситуация оказалась похожа на то, как дети реагируют на подобные вопросы. Первый раз ребенок дает ошибочный ответ на задачу, нарушающую обычную логику решения, зато, узнав правильный ответ, его не забывает. В этом смысле подобные задачи похожи на пароли, некие ключи, разделяемые носителями определенной культуры, а ответы — больше на знание паролей, чем демонстрацию способностей, тем не менее, поняв закономерность, можно решить схожую задачу с иной формулировкой.

В целом имеются в виду вопросы, в которых используются особенности и неоднозначность естественного языка, идет намеренное смешивание кодов и фреймов ситуаций, используются языковые игры, каламбуры и т. п., из-за которых возможны ошибки и неоднозначности в понимании. Подобные проблемы возникают при формулировке математических задач в виде текста. В этом случае ИИ также часто делает ошибки. Еще один вариант причин ошибок, менее распространенный, связан с отсутствием у ИИ знаний о физическом мире. В отличие от предыдущей серии загадок, которые без знания принципа использования языка человек, скорее всего, тоже не решит, загадки «с подвохом», построенные на совокупности языковой игры и знаниях о мире, поддаются решению, однако чаще решаются неправильно человеком, впервые с ними столкнувшись. Примерами задачек такого типа могут служить: «Что тяжелее: килограмм пуха или килограмм свинца?», «Сколько нужно сделать надрезов, чтобы разделить бублик на 5 частей?», «Сколько пролетов лестницы нужно пройти, чтобы попасть на 5 этаж?», «В двух кошельках лежит 4 монеты, причем в одном в два раза больше, чем в другом, как такое возможно?», «Два сына и два отца поделили три яйца так, что ни одного не разбили, как такое возможно?», загадка К. Чуковского: «Шел Кондрат в Ленинград...» и т. п.

Помимо задач, существующих в культуре, ИИ проверяют на лимиты его коммуникативных и интеллектуальных возможностей. Задача представляется достаточно интересной, так как очевидно, что запросы, ответы на которые в более или менее явном виде существуют в виде текстов, бесполезны. Поэтому для того, чтобы проверить ИИ,

приходится быть достаточно изобретательным. Причем иногда вопрошающие, чтобы показать, что ИИ знает не все, прибегают к методу Бильбо Бэггинса, спрашивая что-то про себя лично, вроде «Хорошо ли я целуюсь?» [9, р. 195]. Рассмотрим некоторые направления, в которых человек находит возможности продемонстрировать ущербность ИИ.

1. В ряде случаев ИИ отвечает на вопросы неправильно, когда ему требуется осуществить некоторые метакоммуникативные действия над собственным текстом. В частности, запрос на текст с определенным количеством слов ставил модель в тупик, а когда пользователь повторял свой запрос, ИИ продолжал генерировать варианты текста приблизительно требуемого объема, однако без точного попадания в количество слов. Более легкая задача закончить свой текст определенной буквой не решалась в ранней версии, но в текущей уже решается. Хотя сама по себе способность посчитать количество слов или рассчитать положение буквы вряд ли может рассматриваться как серьезная проблема, однако подобные примеры демонстрируются как доказательства неспособности ИИ совершать простые действия, доступные любому человеку.

2. Математические и логические задачи также часто составляют проблему для современного ИИ, несмотря на то что роботы и ИИ должны быть построены на четких алгоритмах. Однако большие языковые модели опираются на языковые закономерности, поэтому далеко не всегда способны осуществить правильную последовательность расчетов, получив задание на естественном языке. Более того, модель работает нестабильно, и если правильность ответа относительно того, является ли 17 077 простым числом, у GPT3,5 улучшилось с 2,4 до 86,8 % с марта по июнь 2023, то для GPT4 ухудшилась с 97,6 до 7,4 % [10]. А логические задачи, которые представляют собой «чистые рассуждения», особенно идущие в разрез с имеющимся у ИИ языковым отражением реальности, вызывают сбой. Например, Д.В. Зайцев приводит такой пример ошибки: «Я: Верно ли рассуждение: Земля — планета. Все планеты имеют форму чемодана. Значит, Земля имеет форму чемодана?

GPT: Нет, рассуждение неверно. Из того, что Земля — планета и что все планеты имеют форму чемодана, не следует, что Земля имеет форму чемодана. Форма Земли фактически является ближе к сфере, а не к чемодану» [11].

Хотя в данном случае понятно, что ИИ дает ответ, опираясь на представления о мире, а не о логике, но в целом любые логические представления чужды ИИ. Язык логики не входит в арсенал современных языковых моделей (хотя постепенно ошибки исправляются), что позволяет ряду авторов делать вывод о том, что ИИ не обладает интеллектом, рассуждая на примитивном дологическом уровне [12],

или что не понимает ни смысла высказываний, из которых складывается рассуждение, ни смысла критерия правильности рассуждений [11]. В данном случае это не означает, что ИИ не способен ответить на простой вопрос, так как обычный человек тоже может не решить математические и логические задачи. Скорее авторы логических испытаний ИИ подразумевают, что ИИ не обладает базовыми навыками, необходимыми для прохождения испытания вопросами. И соответственно, не имеет значения, что постепенно система будет совершенствоваться и не давать ошибочных ответов. Хотя данная позиция не является традиционной для вопросно-ответной схемы, правильный ответ обычно засчитывается независимо от того, каким образом он был получен.

3. Ответы на вопросы-провокации, или вопросы, данных для ответа на которые недостаточно. Имеются в виду случаи, когда ИИ не просто дает неправильный ответ, а создает некую альтернативную реальность, порождая факты и события, которые никогда не происходили или даже невозможны в принципе. Вопросы-провокации означают ситуацию испытания, когда спрашивающий в вопросе заведомо создает ложную картину и далее просит уточнений. В ряде случаев провокация срабатывает, и ИИ начинает рассказывать о рекордах пересечения Ла-Манша пешком, лечения с помощью отсечения головы, способах приготовления блюда «казах» или причинах добавления клея в пиццу. Также провокации могут осуществляться, когда ИИ дает правильный ответ: настройки человекомашинного взаимодействия позволяют пользователям корректировать ответы ИИ, и при определенной степени настойчивости можно убедить ИИ, что два плюс два равно пять.

Однако подобные эффекты можно получить без провокации в запросе, если задавать вопросы, не слишком хорошо освещенные в текстах, при углублении в узкопрофессиональные темы, в историю и культуру незападных цивилизации и т. п. Так называемые галлюцинации возникают при запросе на создание научных статей и рефератов, особенно в части списка литературы, который частично или полностью может состоять из никогда не существовавших статей и книг [13].

В данном случае испытание стремится продемонстрировать тщетность и бессмысленность диалога с ИИ, не способного отличить вымысел от правды. Сам по себе этот факт не является удивительным, так как в отличие от физической реальности людей, имеющей однозначность и однонаправленность, языковая реальность, на которую опирается ИИ, может одновременно содержать параллельные и противоречивые линии. Поэтому для того, чтобы явно уличить ИИ во лжи, вопросы должны касаться фактов, причем достаточно очевидных, иначе вопрос о ложности утверждения переходит в разряд спорных.

Если рассмотреть наиболее сложные варианты испытаний вопросами в художественных произведениях, то выяснится, что далеко не всегда от проходившего испытание требовалась истинность высказываний. Безусловно, утрачивается узкоутилитарная ценность ИИ, если подразумевается, что на него полагаются как на эксперта. А если имеется в виду коммуникативное равенство агентов взаимодействия, то возможность того, что собеседник обманывает, выдумывает или искренне заблуждается, всегда остается достаточно высокой.

4. Существует также ряд спорных, этически неоднозначных вопросов, которые также достаточно часто предлагают ИИ. В ряде случаев вместо ответов спрашивающие получают стандартные «заглушки» о том, что данную тему ИИ не поддерживает и отвечать не будет. В менее этически неоднозначных ситуациях ИИ высказывает некую позицию, которая обычно не удовлетворяет исследователей, что неудивительно ввиду отсутствия «правильного» ответа. Представление об «истинности» множества утверждений обусловлено позицией человека по тому или иному вопросу или эмоциональным компонентом (А. Яннакиду, А. Мари), от которого зависит, будет ли человек считать истинным суждения вроде «“Война и мир” — шедевр», «Утка — очень вкусное блюдо», «Нападение Хамаса 7 октября 2023 г. было актом сопротивления/террористическим актом» [14]. Ответ ИИ может признаваться неудовлетворительным, потому что он не совпадает с представлениями автора вопроса, или слишком «аморфный» и не содержит выраженной позиции. В частности, ответ о моральной стороне терроформирования Марса интерпретирован как плагиат, апатия и устранение: авторы недовольны тем, что ИИ «суммировал стандартные аргументы из литературы ... отказываясь формулировать позицию» и т. д. [15].

Таким образом, сегодня множество исследователей и энтузиастов активно испытывают ИИ, стремясь найти вопросы, на которые ИИ не может дать правильные ответы. Подобными находками активно делятся как обычные пользователи сети с насмешливыми комментариями в духе «до захвата нас роботами еще далеко», так и авторы серьезных научных статей. В отличие от пользователей и ученых, применяющих большие языковые модели для достижения тех или иных практических результатов, задающие «каверзные вопросы» стремятся найти уязвимости, показать, что ИИ не может пройти испытание, суть которого четко не формулируется, но явно подразумевает переход на некую новую ступень развития, например, признание реальности статуса «интеллект». При этом требуемые способности, начиная от недоступных ИИ приемов и культурных кодов, быстро переходят на требования, превышающие способности условного «обычного человека», особенно когда «испытателями» оказываются авторы научных исследовательских работ, т. е. ИИ испытывают в роли эксперта, учителя,

руководителя, а не простого собеседника. Помимо выводов о том, что ИИ «не прошел испытания», подобные вопросы способствуют развитию технологии, и как правило, достаточно быстро ИИ приобретает необходимые культурные коды для решения загадок и набирает иные правильные ответы, усложняя задачу «испытателей» на новом уровне.

ЛИТЕРАТУРА

- [1] Bender E.M., et al. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York, Association for Computing Machinery, 2021, pp. 610–623. <https://doi.org/10.1145/3442188.3445922>
- [2] Derrida J. Signature, Event, Context. *Limited Inc.* Evanston, Northwestern University Press, 1988, pp. 1–24.
- [3] Derrida J. *Of Grammatology*. Baltimore, The John Hopkins University Press, 1976, 397 p.
- [4] Perez Leon R. Do Language Models Communicate? Communicative Intent and Referent from a Derridean Perspective. *Technology and Language*, 2024, vol. 5, no. 2, pp. 40–56. <https://doi.org/10.48417/technolang.2024.02.04>
- [5] Katz D.M., Bommarito M.J., Gao S., Arredondo P. GPT-4 Passes the Bar Exam. *SSRN Scholarly Paper*, 2023. <https://doi.org/10.2139/ssrn.4389233>
- [6] Лалова Т.И., Симонова С.В. Использование искусственного интеллекта при обучении иностранному языку студентов инженерных специальностей. *Современное педагогическое образование*, 2023, № 12, с. 107–111.
- [7] Морозова А.А., Попова С.Н. Феномен искусственного интеллекта в современной науке: понятие, векторы и проблемы применения в сфере масс-медиа. *Знак: проблемное поле медиаобразования*, 2021, № 4, с. 41–52.
- [8] Lallee A., Mucio N. Hello! How can I assist you today? An Analysis of GPT Technology in Supporting International Entrepreneurship. Smaland, Linneuniversitet, 2023, 108 p.
- [9] Olasik M. “Good morning, ChatGPT, can we become friends?”: an interdisciplinary scholar’s experience of ‘getting acquainted’ with the OpenAI’s Chat GPT: an auto ethnographical report. *European Research Studies Journal*, 2023, vol. 26, no. 2, pp. 190–205. <http://dx.doi.org/10.35808/ersj/3168>
- [10] Chen L., Zaharia M., Zou J. How is ChatGPT’s behavior changing over time? *Harvard Data Science Review*, 2024, no. 6. <https://doi.org/10.1162/99608f92.5317da47>
- [11] Зайцев Д.В. Почему большие языковые модели не (всегда) рассуждают как люди? *Вестник Московского университета. Сер. 7 Философия*, 2024, № 1, с. 76–93.
- [12] Шалак В.И. Избавление от иллюзий ИИ на примере ChatGPT. *Technology and Language*, 2024, № 5, с. 26–39. <https://doi.org/10.48417/technolang.2024.02.03>
- [13] Зашихина И.М. Подготовка научной статьи: справится ли ChatGPT? *Высшее образование в России*, 2023, № 32, с. 24–47.
- [14] Giannakidou A., Mari A. The Human and the Mechanical: logos, truthfulness, and ChatGPT. *arXiv*, 2024. <https://doi.org/10.48550/arXiv.2402.01267>
- [15] Kuo R. Noam Chomsky: The False Promise of ChatGPT. *The New York Times*. URL: <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html> (дата обращения 15.01.2025).

Статья поступила в редакцию 31.03.2025

Ссылку на эту статью просим оформлять следующим образом:

Быльева Д.С. Вопросы к искусственному интеллекту: какие и зачем. *Гуманитарный вестник*, 2025, вып. 2. EDN QYHMVL

Быльева Дарья Сергеевна — канд. полит. наук, доцент Высшей школы общественных наук Санкт-Петербургского политехнического университета Петра Великого. e-mail: Bylieva_ds@spbstu.ru

Questions to artificial intelligence: what and why

© D.S. Bylieva

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, 195251, Russia

The paper shows that a dialogue with the artificial intelligence today is becoming part of the everyday human practice. In addition to those users pursuing their practical goals, there are many enthusiasts and researchers, who purposefully seek to ask questions that artificial intelligence would answer incorrectly. This is quite a difficult problem, and resembles the test questions that are present in the cultural tradition since the ancient times. A candidate for the status or benefit is required to solve a riddle. The paper analyzes posts and articles presenting erroneous answers from the large language models. It identifies the following categories of questions, to which artificial intelligence gives incorrect answers: questions based on a language game, requiring metacommunicative skills, provoking the “artificial intelligence hallucinations”, concerning controversial and ethically ambiguous topics, as well as the logical and mathematical problems. The paper notes that at the current stage of the human-machine interaction, humanity is actively testing the artificial intelligence, seeks to prove its imperfection, thus learns it, and enriches the technology.

Keywords: artificial intelligence, technology philosophy, artificial intelligence philosophy, human-machine interaction, communication, questions, large language models

REFERENCES

- [1] Bender E.M. et al. On the dangers of stochastic parrots: can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York, Association for Computing Machinery, 2021, pp. 610–623. <https://doi.org/10.1145/3442188.3445922>
- [2] Derrida J. *Signature, event, context*. Limited Inc., Evanston, Northwestern University Press, 1988, pp. 1–24.
- [3] Derrida J. *Of Grammatology*. Baltimore, the John Hopkins University Press, 1976, 397 p.
- [4] Perez L.R. Do language models communicate? Communicative intent and referent from a Derridean perspective. *Technology and Language*, 2024, vol. 5, no. 2, pp. 40–56. <https://doi.org/10.48417/technolang.2024.02.04>
- [5] Katz D.M., Bommarito M.J., Gao S., Arredondo P. GPT-4 passes the bar exam. *SSRN Scholarly Paper*, 2023. <https://doi.org/10.2139/ssrn.4389233>
- [6] Lalova T.I., Simonova S.V. Ispolzovanie iskusstvennogo intellekta pri obuchenii inostrannomu yazyku studentov inzhenernykh spetsialnostey [Use of artificial intelligence in teaching a foreign language to engineering students]. *Sovremennoe pedagogicheskoe obrazovanie — Modern Pedagogical Education*, 2023, no. 12, pp. 107–111.
- [7] Morozova A.A., Popova S.N. Fenomen iskusstvennogo intellekta v sovremennoy nauke: ponyatie, vektory i problemy primeneniya v sfere massmediya [The phenomenon of artificial intelligence in modern science: concept, vectors, and problems of application in the sphere of mass media]. *Znak: problemnoe pole mediaobrazovaniya*, 2021, no. 4, pp. 41–52.

- [8] Lallee A., Muco N. *Hello! How can I assist you today? An Analysis of GPT Technology in Supporting International Entrepreneurship*. Smaland, Linneuniversitet, 2023, 108 p.
- [9] Olasik M. “Good morning, ChatGPT, can we become friends?”: an interdisciplinary scholar’s experience of ‘getting acquainted’ with the OpenAI’s Chat GPT: an auto ethnographical report. *European Research Studies Journal*, 2023, vol. 26, no. 2, pp. 190–205. <https://doi.org/10.35808/ersj/3168>
- [10] Chen L., Zaharia M., Zou J. How is ChatGPT’s behavior changing over time? *Harvard Data Science Review*, 2024, no. 6. <https://doi.org/10.1162/99608f92.5317da47>
- [11] Zaitsev D.V. Pochemu bolshie yazykovye modeli ne (vsegda) rassuzhdayut kak lyudi? [Why don't big language models (always) reason like humans?]. *Vestnik Moskovskogo universiteta. Ser.7 Filosofiya — Lomonosov Philosophy Journal*, 2024, no. 1, pp. 76–93.
- [12] Shalak V.I. Izbavlenie ot illyuziy II na primere ChatGPT [Exposing illusions – the limits of AI by the example of ChatGPT]. *Technology and Language*, 2024, no. 5, pp. 26–39. <https://doi.org/10.48417/technolang.2024.02.03>
- [13] Zashikhina I.M. Podgotovka nauchnoy statyi: spravitsya li ChatGPT? [Scientific article writing: will ChatGPT help?]. *Vysshee obrazovanie v Rossii — Higher Education in Russia*, 2023, no. 32, pp. 24–47.
- [14] Giannakidou A., Mari A. The human and the mechanical: logos, truthfulness, and ChatGPT. *arXiv*, 2024. <https://doi.org/10.48550/arXiv.2402.01267>
- [15] Kuo R. Noam Chomsky: the false promise of ChatGPT. *The New York Times*. Available at: <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html> (accessed January 15, 2025).

Bylieva D.S., Cand. Sc. (Political Sciences), Associate Professor, Higher School of Social Sciences, Peter the Great St. Petersburg Polytechnic University.
e-mail: Bylieva_ds@spbstu.ru