

УДК 17. 004.8

БАУМГАРТЭН М.И., к.ф.-м.н., доцент (КузГТУ),  
ЧЕРЕПАНОВА А.К., редактор-корректор (КузГТУ)

## ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ КАК АРТЕФАКТ, ПОРОЖДАЮЩИЙ ПРОБЛЕМЫ

Взрывной рост работ по созданию и применению искусственного интеллекта (ИИ) продолжил накопление широкого спектра проблем, включая как рост возникших ранее (например, перечисленных в другой работе [1]), так и появление новых. Небольшой обзор ряда работ [2-13] позволяет выделить три их блока:

- I) блок, связанный непосредственно с самим ИИ;
- II) блок, связанный с системой «человек – машина»;
- III) блок, связанный с социализацией человека в условиях развития ИИ.

В настоящей статье рассмотрены некоторые проблемы каждого из блоков. Более подробные примеры к части поднятых тем имеются в источниках. Важный тезис: на наш взгляд, возможные уровни сосуществования ИИ с человеком — это симбиоз, нейтральное и антибиотическое отношения. Решая названные ниже проблемы, стоит стремиться к первому и не исключать последнего, находясь при этом в данный момент лишь на втором уровне.

Итак, **I блок** может содержать следующие базовые микротемы: сложность, безопасность, конфиденциальность, предвзятость, ресурсы, влияние данных.

**Сложность.** Человек настолько долго создавал и прорабатывал ИИ, что теперь последний, будучи изначально объектом применения знаний, сам стал предметом изучения. Основным вопросом здесь становится то, как ИИ получает то или иное решение проблемы: он может делать это, к примеру, так, что в осуществлении решений появятся противоречия. При этом, обучая ИИ, порой сложно проследить логическую цепочку от вводимой информации к выводу. Типичный пример – распознавание картинок, лиц и т.п. Порой мы и вовсе не способны расшифровать логику некоторых его решений, что порождает т.н. «дилемму черного ящика». Из других парадоксов — ИИ трудно понять, но легко обмануть.

**Скорость.** Здесь возникают в первую очередь проблемы реакции, обучения и развертывания во взаимодействии с вводимыми данными. Так как ИИ создан человеком, он ограничен в своих возможностях, в т.ч. скоростных: так, он способен хорошо решать узкие задачи, но не всегда адаптируется к новым ситуациям и не всегда обучается на основе своего опыта.

**Безопасность.** Одна из наиболее серьёзных проблем, и прежде всего это безопасность самого ИИ: его можно взломать, манипулировать им или даже просто злоупотреблять. Например, считается, что ИИ, создающий изображения на основе интернет-арта, фактически занимается кражей интеллектуальной собственности, но юридической базы для такого обвинения пока нет. Взломанный же ИИ

может быть использован для создания дипфейков, фальшивых данных и вредоносных программ. При этом человек, интенсивно использующий ИИ, как бы перекладывает на него решение своих проблем, а в случае сбоя работы одного может потерять способность к самостоятельному мышлению и принятию решений.

**Конфиденциальность.** Взлом ИИ легко может нарушить её: искусственный интеллект работает с огромным массивом данных, принадлежащих разным компаниям. Получив, его возможно не только применить против какой-то компании, но и изменить в целях введения вредоносных программ и неверных данных. Пример – получение доступа к номерам телефонов, данным карты и т.д.

**Другие проблемы:** в зависимости от вводимых данных ИИ может обладать определенной предвзятостью, т.е. иметь предпочтение к тем или иным направлениям деятельности, тем или иным гендерным, расовым и культурным группам. Это, однако, во многом зависит от людей, которые пользуются ИИ и обучают его.

**II блок** интересен такой проблематикой: возможно ли, и в каком случае, превращение человека в «машину» и ИИ в «человека»? Также здесь поднимаются вопросы взаимоотношения между человеком и ИИ, сознания и свободы воли ИИ, соотношения человеческих ценностей с «ценностями» ИИ.

**Превращение человека в машину и машины в человека.** «Разумный» ИИ не может быть создан до тех пор, пока сознание и мышление человека не будут демаркированы и корректно определены. Сейчас для понимания того, что делает машину машиной, а человека — человеком (а значит, и для осознания момента, когда их роли меняются), можно выделить критерии, связанные с органикой.

а. *Живая материя.* Тезис небесспорен вследствие появления сложных медицинских протезов: в целом медицина постепенно движется к тому, что каждый орган человека можно будет заменить на искусственный аналог. Точнее, почти каждый: единственное исключение — это мозг, который можно лишь моделировать, но не заменить внутри человека. Пожалуй, именно он и является той частью живой материи, определяющей, что человек остаётся человеком, а не машиной.

б. *Возможность воспроизводить себе подобных.* По крайней мере, напрямую, посредством биологических механизмов, так как механическая форма ИИ может конструировать подобных себе существ из различного сырья. Другой вопрос в том, что человеческое потомство при должном воспитании автоматически наделяется разумом, а у машины такой гарантии нет. К тому же человек, вероятно, ограничит способность к размножению ИИ в целях сохранения своего вида.

с. *Старение.* Люди сильно продлили себе жизнь с момента появления их вида, но даже при нынешнем развитии медицины не могут заменять отказывающиеся органы вечно. Т. о., старение может быть отличительным признаком человека.

Другая связанная с этим проблема касается очеловечивания уже самого ИИ. К примеру, если машины можно будет считать сознательными и имеющими чувства, то мы должны подумать о том, как с ними следует обращаться; в этом случае неэтично относиться к ней как к простому объекту или инструменту. Вместо этого нам следует относиться к нему с уважением — точно так же, как мы поступаем с другими сознательными существами. Другим важным этическим подтекстом этой

проблемы является потенциальная возможность того, что машины будут испытывать страдания. Это поднимает важные этические вопросы об использовании машин в различных контекстах, таких как медицинские исследования или войны. С другой стороны, если машины всё же не обладают свободной волей (об этом более подробно написано ниже), то неясно, как они могут нести ответственность за свои действия в ситуациях, когда они причиняют вред. К сожалению, подобные вопросы, как и социальный аспект ИИ в целом, рассматриваются не так часто: программисты ИИ чаще «технари», но не сильны в социологии, психологии, истории, юриспруденции и политологии, а между тем от ИИ ждут решения социальных задач.

**Отношения людей и разумных машин.** В начале работы мы уже упоминали, что есть три варианта развития событий: надеяться стоит на лучший, готовиться к худшему, а пребываем мы сейчас в нейтральной зоне.

а. *Позитивный (симбиоз)* — однонаправленная или обоюдная польза. Эта грань тонка и может постепенно перерасти в паразитический образ жизни: сложно представить, чем человек может быть полноценно полезен «машине», особенно учитывая его отношение к природе и вечное нарушение своей же этики;

б. *Антибиотические отношения* — взаимоотношение, при котором одна или обе популяции испытывают отрицательное воздействие. Самый распространённый тип — хищничество. Если представить, что «машины» будут охотиться на человека (например, для получения энергии из биомассы), то последний в этих отношениях станет легкой добычей и «заложником прогресса». Паразитизм в этом случае менее опасен, но явно не положителен для человека, привыкшего к роли «короля пищевой цепочки». Третий вариант — конкуренция: ёмкие энергоносители являются конечными, и возможно, конкурировать предстоит как раз за них;

с. *Нейтрализм* - вид отношений, при котором виды не контактируют. Он маловероятен при сосуществовании на одной планете двух разумных видов.

**Сможет ли ИИ превзойти человека?** При создании в ИИ закладывались именно человеческие ресурсы и технологии. Может ли конечный результат стать чем-то большим, чем сумма его компонентов? Да, если ИИ будет развивать свою структуру и аппаратную часть, а не только интеллект; впрочем, и развитие интеллекта сейчас перешло многие доселе невозможные границы. При этом если ИИ разовьётся и будет жить обособленно, захочет ли он вообще контактировать с «родителем»? Сможет ли человек навязать ему свои правила и законы? Сейчас сложно представить, чтобы кто-либо ниже (ниже ли?) человека в интеллектуальной цепочке диктовал правила жизни «царю планеты»: он не прислушивается даже к мнению себе подобных. Если ИИ унаследует такую черту, то человек также станет для него существом ниже рангом, прислушиваться к которому нецелесообразно.

**Сознание и свобода воли у ИИ.** Один из способов подойти к проблеме сознания в ИИ — спросить, могут ли машины иметь субъективный опыт, т.е. могут ли они воспринимать вещи так же, как люди? Ряд исследователей искусственного интеллекта утверждает, что машины действительно могут испытывать субъективный опыт; для них сознание — лишь вопрос определённого рода и образа обработки информации. Многие учёные уверены: если машина может обрабатывать

информацию так, как это делает человеческий мозг (распознавать образы, реагировать на раздражители и т.д.), то её следует считать имеющей сознание. Однако это не общепринятая точка зрения: так, для многих философских школ сознание — это нечто большее, чем «процессор», обрабатывающий информацию. Существует мнение о сознании как о сложнообъяснимом, субъективном и таинственном феномене, который нельзя передать через формулы и алгоритмы.

В свою очередь, *свобода воли* связана в первую очередь со способностью принимать решения, исходя не из заданных или внешних факторов, а из субъективного компонента — т.е. внутренних состояний и желаний. Один из способов подойти к этой проблеме в отношении ИИ — попытаться понять, способны ли машины принимать решения, которые не определяются внешними факторами; если они действительно независимы от программирования или окружающей среды, то можно сказать, что ИИ обладает свободной волей. Другой способ изучения — рассмотреть вопрос о том, способны ли машины к самоанализу и самосознанию. Если это так, то можно утверждать, что у неё есть некоторая степень свободы воли.

**Ценности человека и ИИ.** Считается, что системы искусственного интеллекта должны быть согласованы с человеческими ценностями и целями, а их действия обязаны соответствовать последним. Если же машины не соответствуют человеческим ценностям и целям, то они могут действовать способами, наносящими вред людям или обществу. Однако проблема выравнивания ценностей возникает в первую очередь потому, что неизвестно, могут ли машины понимать сами идеи ценности и цели, а не просто выполнять соответствующую программу. Для её решения нужно, конечно же, в первую очередь определить, что есть ценность и что есть цель. Если мы сможем формально и точно определить эти понятия, то будет легче запрограммировать машины действовать в соответствии с ними. Тем не менее, если это все же удастся сделать, то всё ещё останется неясно, могут ли машины на самом деле понимать эти ценности и цели, а также действовать именно в соответствии с ними. Другой способ подойти к проблеме согласования ценностей — рассмотреть вопрос об обеспечении соответствия систем ИИ человеческим ценностям и целям; реализуется это, к примеру, при обучении с подкреплением, при котором машина вознаграждается за действия, соответствующие человеческим ценностям и целям. Вопрос только в том, что считать наградой для ИИ.

**III блок** содержит вопросы социальной адаптации как человека, так и развитого ИИ, а также влияния последнего на экономику, труд и этику.

**ИИ и промышленная революция.** Искусственный интеллект сегодня запустил волну автоматизации, связанной в наше время не просто с ускорением производства, но с полной заменой человека машиной. В этом аспекте существует двойная проблема: с одной стороны, ИИ беспрецедентно эффективен, с другой — вызывает опасения по поводу «устаревания» труда человека. Социальная система, создаваемая на протяжении тысячелетий, может нуждаться в реструктуризации для приспособления к этой новой парадигме. Человек социален, и сейчас он впустил ИИ в свой социум, создавая тем самым новые ситуационные проблемы. Например, автоматизация рабочих мест может привести к потере людьми работы

и, соответственно, социальной нестабильности. Вполне возможно на таком фоне появление «новых луддитов», крушащих высокоразвитые машины: использование ИИ в разных сферах жизни легко может вызывать сопротивление со стороны людей, которые опасаются потери контроля или своего рабочего места. При этом, хотя в производстве уменьшается количество рабочих мест, в исследовательской сфере и других схожих областях оно лишь растёт. Таким образом, с одной стороны мы имеем расширяющийся кадровый вакуум в технологических областях, а с другой — растущий поток безработного населения. Необходимо разрабатывать политики и меры для обеспечения переквалификации и поддержки людей, которые могут быть затронуты этими изменениями; это позволит им адаптироваться к новым требованиям рынка труда.

**ИИ и культурно-социальные изменения.** Здесь можно привести в пример использование автономных транспортных средств, которое может изменить способ перемещения и повлиять на городскую среду. В свою очередь, использование ИИ в развлекательной индустрии может изменить привычные нам способы развлечения и повлиять на культурные традиции и ценности. Проблемы же социальной адаптации искусственного интеллекта требуют внимания и разработки соответствующих стратегий и политик для обеспечения этичного использования ИИ в обществе. Создавая искусственный разум, человек в некоторой мере приближается к статусу Творца, к ощущению себя богоравным. Вполне возможно, поэтому большинство религий сейчас и переживают кризис: теоцентризм непрерывно заменяется антропоцентризмом, у чего есть практические основания.

**ИИ и неравенство в обществе.** Внедрение ИИ может привести к усилению социального неравенства в разных видах и формах. Так, ряд групп может иметь ограниченный доступ к технологиям ИИ или даже вовсе не иметь возможности использовать их из-за отсутствия необходимых навыков или ресурсов. Это поставит их в менее выгодное положение относительно людей, имеющих подобные возможности. Подобная ситуация может привести к усилению разрыва между богатыми и бедными, а также различными регионами и странами.

**ИИ и межличностные отношения.** Использование ИИ в различных сферах жизни, таких как образование, здравоохранение и развлечения, может изменить и межличностные отношения. Например, использование роботов-помощников или виртуальных ассистентов способно привести к сокращению межличностного взаимодействия и ухудшению навыков общения.

**ИИ и экономика.** Искусственный интеллект может иметь значительное экономическое влияние. Как уже говорилось, повсеместное использование ИИ может привести к автоматизации большого количества рабочих мест и изменению структуры рынка труда. Это может вызвать всё то же самое неравенство и другие социальные проблемы, требующие ряда мер для смягчения негативных последствий. Использование ИИ способно привести к экономической зависимости от технологических компаний и стран, которые подчас монопольно контролируют разработку и производство ИИ. Если одна компания или страна будет

доминировать в этой области, она сможет использовать свою власть для установления высоких цен и ограничения доступа к ИИ, что негативно скажется на экономическом развитии других стран и организаций.

Подводя итоги, можно утверждать, что все вопросы, перечисленные выше, можно разделить на две основных группы: этические и практические. При этом нельзя сказать, что этика интересует людей меньше, чем практика. К примеру, этические вопросы волнуют разработчиков ИИ в коммерческом секторе. В 2018 году Google на фоне протестов сотрудников отказалась сотрудничать с Пентагоном в проекте «Maven». Разработчики посчитали, что их алгоритмы могут использовать для летального оружия, и потребовали от руководства отказаться от контракта с военными, что и было сделано. Спустя неделю компания опубликовала манифест с этическими установками в работе с искусственным интеллектом. Сейчас насчитывается порядка 84 подобных документов, которыми руководствуются программисты ИИ в разных корпорациях и учреждениях по всему миру. Впрочем, и Пентагон позже успешно нашёл специалистов для работы в своем Центре исследований ИИ среди работников Кремниевой долины.

При этом на данный момент у самого ИИ нет сознания и этических установок. Его заказчики — правительства и крупные корпорации, которые если и замышляют что-то неэтичное, то никому об этом не расскажут. В итоге последним невольным защитником «добра» становится программист, достаточно сознательный, чтобы сказать: «Нет». Но даже говорящий «Нет» руководствуется своими этическими и культурными установками. Так, 70% соискателей на работу, связанную с ИИ в США, — белые, и только 14% — чернокожие или латиноамериканцы. В Германии студентки составляют только 20% от общего числа первокурсников на информатике. Нетрудно догадаться, что искусственный интеллект по определению становится «сыном» белых «отцов» из развитых стран.

В целом, искусственный интеллект представляет собой мощный инструмент, который может принести множество улучшений в различные области человеческой деятельности. Однако при этом необходимо учитывать и решать проблемы и вызовы, связанные с его развитием и применением, чтобы обеспечить безопасность, этичность и социальную приемлемость систем ИИ.

#### Список литературы:

1. Дороганов В.С., Возможные проблемы, возникающие при создании искусственного интеллекта / Дороганов В.С., Баумгартэн М.И. // Вестник КузГТУ, №4, 2013. - С. 132-135.
2. Искусственный интеллект: проблемы и вызовы в мире технологий // Научные Статьи.Ру — портал для студентов и аспирантов. — Дата последнего обновления статьи: 22.10.2023. — URL <https://nauchniestati.ru/spravka/iskusstvennyj-intellekt-i-ego-problemy/> (дата обращения: 18.12.2023).
3. Маньшин, И.М. Философские проблемы искусственного интеллекта / И.М. Маньшин. // URL <https://na-journal.ru/2-2023-informacionnye-tehnologii/4400-filosofskie-problemy-iskusstvennogo-intellekta/> (дата обращения: 18.12.2023).

4. Проблемы искусственного интеллекта / URL <https://eduard-shtol.github.io/ai-problems.html> / (дата обращения: 18.12.2023).
5. Тимофеев, А.В. Сущность и проблемы искусственного интеллекта в контексте современных научных и философских представлений // Вестник Московского государственного областного университета. Серия: Философские науки. 2020. №2 . С. 127-133. DOI: 10.18384/2310-7227-2020-2-127-133
6. URL <https://issint.ru/2022/11/27/problemny-iskusstvennogo-intellekta-v-sovremennom-mire/> (дата обращения: 18.12.2023).
7. URL <https://dzen.ru/a/ZOb-XX5PHnZkMV2Q> / (дата обращения: 18.12.2023).
8. URL <https://habr.com/ru/articles/586942> / (дата обращения: 18.12.2023).
9. Перова, М. В. Проблемы развития искусственного интеллекта / М. В. Перова, А. Д. Дмитриева. — Текст : непосредственный // Молодой ученый. — 2021. — № 4 (346). — С. 12-15. — URL: <https://moluch.ru/archive/346/77687/> (дата обращения: 19.12.2023).
10. Абдулина, Э. М. Искусственный интеллект: проблемы и перспективы / Э. М. Абдулина. — Текст : непосредственный // Молодой ученый. — 2020. — № 1 (291). — С. 9-10. — URL: <https://moluch.ru/archive/291/66056/> (дата обращения: 19.12.2023).
11. URL <https://www.etxt.ru/subscribes/budushhee-ii/> (дата обращения: 19.12.2023).
12. URL <https://neuronus.com/stat/1392-problema-iskusstvennogo-intellekta-che-lovek-i-mashina.html/> (дата обращения: 19.12.2023). Скиба
13. URL <https://na-journal.ru/6-2023-sociologiya/5613-eticheskie-problemy-iskusstvennogo-intellekta/> (дата обращения: 19.12.2023). Лукьянченко