CASE STUDIES

УДК 687.5.01
DOI: 10.32326/2618-9267-2024-7-2-108-120

КОГНИТИВНЫЕ БАРЬЕРЫ НА ПУТИ ДОВЕРИЯ ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ (НА ПРИМЕРАХ ИЗ ПОПУЛЯРНЫХ СЕРИАЛОВ)

Горман Анна Виктория Вячеславовна – соискатель на степень кандидата философских наук, философский факультет, Московский государственный университет им. М.В. Ломоносова; Российская Федерация, 119192 Москва, Ломоносовский проспект, 27, e-mail: anna.victoria.gorman@gmail.com, ORCID: 0009-0009-9626-7458

Автор предлагает разделить барьеры, связанные с доверием искусственному интеллекту (ИИ), на условно рациональные и условно иррациональные. Среди барьеров автор выделяет: страх перед технологической безработицей, страх потерять контроль, страх перед усилением дискриминации. В качестве объекта анализа выступают примеры презентации этих страхов в массовой культуре – в популярных сериалах «1899», «Мир Дикого Запада», «Разделение», «Чёрное Зеркало». Выбранные для анализа художественные примеры демонстрируют актуальность рациональных когнитивных барьеров, связанных со страхами бесконтрольного внедрения технологий. Также в художественных произведениях показаны возможные последствия несовершенного функционировании ИИ и ставится вопрос о принципиальной возможности создания идеальной системы ИИ. Художественное переосмысление классических философских вопросов о свободе воли, месте человека в обществе, справедливости показывает необходимость актуализации этих вопросов в контексте внедрения новых технологий. Анализ также показывает актуальность исследований в области целеполагания систем ИИ, а также влияния систем ИИ на психику индивида, устоявшиеся межличностные и общественные отношения. Художественное осмысление когнитивных аспектов может быть полезным для формирования успешной стратегии внедрения Доверенного ИИ, которая может быть реализована с привлечением философов и практиков.

Ключевые слова: искусственный интеллект, внедрение технологий, доверие к ИИ, доверенный ИИ, кино, искусство и технологии

Цитирование: Горман А.В.В. Когнитивные барьеры на пути доверия искусственному интеллекту (на примерах из популярных сериалов) // Цифровой учёный: лаборатория философа. 2024. Т. 7. № 2. С. 108-120. DOI: 10.32326/2618-9267-2024-7-2-108-120

Рукопись получена: 18 февраля 2024
Пересмотрена: 30 марта 2024
Принята: 8 апреля 2024

THE COGNITIVE BARRIERS ON THE WAY TO TRUSTING ARTIFICIAL INTELLIGENCE (USING EXAMPLES FROM POPULAR TV SERIES)

Anna Viktoriia V. Gorman – PhD candidate, Lomonosov Moscow State University, Faculty of Philosophy; 27 Lomonosovsky prospekt, Moscow 119192, Russian Federation; e-mail: anna.victoria.gorman@gmail.com, ORCID: 0009-0009-9626-7458

The author proposed to classify the cognitive barriers that stand in the way of trust towards AI as mostly rational and irrational. The author further classifies rational barriers into the following three types: fear of technical unemployment, fear of losing control, and fear of discrimination acceleration. In this research paper, the author analyzes these cognitive barriers by using examples from the most popular TV series centered around AI-related risks: 1899, Westworld, Severance, and Black Mirror. These examples demonstrate the relevance of rational cognitive barriers linked to the uncontrolled implementation of technologies. Artistic comprehension follows various scenarios with AI systems functioning consequences leaving the question of the fundamental possibility of an ideal system creation hanging. A fictional rendering of classic philosophical questions regarding the issue of free will, human place in society, social justice, proves the need to update these issues with new technologies emerging. The analysis also shows the need for additional research on the impact of the goal-setting systems on each other, society in general, and an individual. The creative comprehension of cognitive aspects serves as material that might help philosophers and practitioners derive strategies for successful AI implementations.

Keywords: artificial intelligence, trustworthy AI concept, series, art and technologies

How to cite: Gorman, A.V.V. (2024). The cognitive barriers on the way to trusting artificial intelligence (using examples from popular TV series). *The Digital Scholar: Philosopher's Lab, 7 (2)*: 108-120. DOI: 10.32326/2618-9267-2024-7-2-108-120 (In Russian)

Received: 18 February 2024
Revised: 30 March 2024
Accepted: 8 April 2024

В рамках концепции Доверенного искусственного интеллекта (ИИ) предлагается создать объективные критерии, при которых можно было бы допустить внедрение ИИ (European Commission, 2019). Однако на доверие будут оказывать влияние не только зало-

женные в конкретную систему алгоритмы, этические нормы или иные технологические особенности, но также непосредственно не относящиеся к каждой системе ИИ когнитивные аспекты. Именно они будут барьерами на пути эффективного использования ИИ, а иногда и диалога о разработке таких систем. Изучение этих барьеров необходимо для формирования соответствующих коммуникаций в рамках междисциплинарного диалога.

1. Типология когнитивных аспектов, стоящих на пути доверия к ИИ

По результатам анализа научной литературы о доверии к ИИ и художественных произведений, переосмысляющих эту тему, автор предлагает разделить когнитивные барьеры на условно рациональные и условно иррациональные. *Рациональными барьерами* являются те, которые могут быть валидированы или опровергнуты с помощью инструментов рациональной дискуссии – например, объяснены прошлым опытом. Под *иррациональными барьерами* подразумеваются те, валидность которых нельзя опровергнуть или доказать без обращения к теологии и метафизике. Такое разделение может быть полезным для стратегий повышения доверия к ИИ.

Также автор предлагает выделить корневые барьеры и отделить их от некорневых. Корневыми барьерами предлагается считать следующие: страх перед технологической безработицей, страх потерять контроль, страх перед усилением дискриминации, страх перед антропоморфностью и сингулярностью.

Примером барьера, не являющегося корневым, является страх за безопасность. Опасения на счёт безопасности не существуют сами по себе, но связаны с любым из названных выше корневых барьеров. Поэтому страх за безопасность и другие некорневые барьеры можно рассматривать в контексте обозначенных корневых.

Следует отметить, что в предложенной типологии не учитываются барьеры, связанные с дизайном систем ИИ. Так, например, предполагается, что защита от хакерских атак войдёт в перечень основных критериев в рамках концепции Доверенного ИИ и не требует отдельного рассмотрения.

Иррациональные страхи в данной статье не рассматриваются, а лишь упоминаются, чтобы обозначить границы исследования. Например, страх перед антропоморфностью (перед роботами, которые практически не отличаются от человека) не понят и не изучен до конца, поэтому относится автором к иррациональным страхам. Под страхом перед сингулярностью объединены экзистенциальные страхи с высокой степенью абстракции.

Таблица 1. Основные рациональные и иррациональные страхи, стоящие на пути доверия к системам общего ИИ и не относящиеся к дизайну данных систем

Категория	Примеры корневых барьеров
Рациональные барьеры	Страх перед технологической безработицей Страх перед потерей контроля Страх перед усилением дискриминации
Иррациональные барьеры	Страх перед антропоморфностью Страхи перед сингулярностью

2. Анализ когнитивных барьеров

2.1 Страх перед технологической безработицей выражается в беспокойстве, что ИИ (особенно общий ИИ, и чаще в форме роботов) заменит человека в производственных отношениях.

Страх перед технологической безработицей (страх быть ненужным) можно отнести к типу рациональных страхов. Д. М. Кейнс в своём эссе «Экономические возможности для наших внуков» определял технологическую безработицу как обусловленную временем ситуацию, когда «изобретение способов экономии на использование рабочей силы обгоняет создание новых способов её использования» (Keynes, 1932, p. 360). Этот страх сопутствовал всем волнам технологических революций, которые приводили к автоматизации: он не уникален для дискуссий об ИИ. Кейнс, рассуждая о временности этого явления, утверждал, что в долгосрочной перспективе, то есть через сто лет, в странах, которые он обозначал как прогрессивные, уровень жизни вырастет как минимум в четыре раза.

Метаобзор, посвящённый вопросу влияния технологий на уровень безработицы, показывает, что за последние четыре десятилетия технологии скорее создают для людей новые рабочие места, а не отнимают их (Hötte, Somers, Theodorakopoulos, 2022). Отчасти это связано с тем, что пока самым эффективными и безопасными остаются системы, которые совмещают работу человека и интеллектуальных систем. Изобретение калькулятора действительно не сделало математиков ненужными, однако само занятие математикой перешло на более абстрактный уровень, благодаря чему появилось множество открытий. Можно предположить, что в одном из возможных вариантов будущего ИИ станет ещё одним инструментом, который человек будет использовать для улучшения результатов своей интеллектуальной деятельности.

Однако исследователи прогнозируют и возможные негативные последствия внедрения технологий. Отчёт *McKinsey & Company* показывает, что блага от технологических изменений всё чаще распределяются неравномерно (McKinsey Global Institute, 2016). Таким образом, усиливается поляризация общества, а технологический прогресс сказывается негативно на наиболее уязвимых социальных группах. Так, люди без образования, женщины и люди старшего возраста имеют гораздо больше шансов столкнуться с потерей работы в связи с технологическим прогрессом, чем люди с высшим образованием, мужчины, люди молодого возраста. Естественно, уязвимым группам сложнее адаптироваться к изменениям как финансово, так и психологически, что порождает страх перед ИИ и роботами.

На страх перед технологиями оказывает влияние уровень удовлетворенностью жизнью. Тим Хинкс представил корреляцию между низким уровнем удовлетворенности жизнью и страхом перед роботами, а также предположил, что опыт безработицы был для некоторых исследуемых настолько травмирующим, что усиливает страх (Hinks, 2021).

2.2. Страх перед потерей контроля над той или иной сферой человеческой жизни при внедрении ИИ относится к числу рациональных страхов. К этой группе автор относит различные страхи: от утечек персональных данных и контроля над личностью до размытия границ между реальным и нереальным. Воплощениями этого страха могут быть как индивидуальные риски (например, при краже персональных данных), так и риски для общества в целом, если люди потеряют мотивацию и волю к жизни в силу экзистенциального кризиса.

Часто при обсуждении этого страха не проводится различие между узким и общим ИИ, тогда как сама технология ИИ является лишь одной из многих, вызывающих такой страх.

В рамках даже самых простых систем ИИ может реализовываться более общий страх потери контроля, существующий в двух плоскостях: с одной стороны, финансовая и социальная зависимость, с другой –деградация (когнитивная или физическая зависимость).

Страх потери контроля над той или иной сферой человеческой жизни также часто сосуществует со страхом размытия границ между реальным и нереальным миром. Так, О.Н. Гуров говорит в контексте метавселенных о растущей зависимости государства, общества и индивида от технологий и выдвигает гипотезу, что «развитие технологий, связанных с метавселенными, подвергает человека риску потерять автономию», лишиться «физической и моральной свободы…, возможности и необходимости существовать в измерениях нравственности» (Гуров, 2022). Также Гуров высказывает опасение, что развитие таких технологий приведёт

к «деградации в виртуальном мире ценностей и морали», а также к «распаду идентичности индивида». Из этого философ делает вывод о том, что такие цифровые технологии, внедрённые в социальную жизнь бесконтрольно, соединяя технологические и человеческие физические миры, представляют угрозу «цивилизационному развитию» и человечеству в целом. Метавселенные Гуров, как и многие другие алармисты, сравнивает с «современными версиями Паноптикона» (по концепции идеальной тюрьмы Бентама, родоначальника утилитаризма).

Концепция Паноптикона имела невероятный успех у философов, в том числе у М. Фуко. Современные алармисты, высказываясь против современных систем контроля, используют данную концепцию для характеристики «общества наблюдения», когда наблюдение ведётся за каждым индивидом через видеокамеры наружного наблюдения, в социальных сетях и в поисковых системах. Также алармисты подчёркивают, что дальнейшее развитие технологий усилит этот тренд до точки невозврата.

При этом в сердце Паноптикона, по мнению алармистов, лежат такие утилитарные принципы, как эффективность и экономия. Единственной целью компаний, внедряющих технологии, О.Н. Гуров видит получение сверхприбыли: «Для этики, морали и нравственности в такой системе не найдётся места – регулирование и контроль обеспечиваются совершенно иными, "нечеловеческими" системами» (Гуров, 2022).

В.А. Кутырёв смотрит на внедрение новых технологий и вовсе как на угрозу стремлениям человека к духовному развитию, угрозу субъектности человека и, наконец, угрозу существованию человечества: «Человек умирает, активно, интенсивно и прогрессивно» (Кутырёв, 2010, с. 10, 15).

Большинство алармистов, рассуждая о риске потерять контроль, исходят из нескольких предпосылок или гипотез. Например, что мир нетехнологический (физический, человеческий) лучше технологического. О.Н. Гуров описывает офлайновый мир как «естественный» и именно «по своей природе более человеческий, глубокий и истинный, скреплённый этико-морально-нравственным комплексом» (Гуров, 2022). В.А. Кутырёв условием существования человека и высшей ценностью культуры называет «пол и любовь»; связывая с новыми технологиями упадок гуманизма, философ предполагает, что естественный мир руководствуется в какой-то мере гуманистическими ценностями (Кутырёв, 2010, с. 10).

Эти концепции можно свести к трём основным тезисам, которые сопровождаются аргументами за и против: (1) человек обладает контролем, автономией, свободой воли в текущей системе; (2) человек превосходит животных и другие формы жизни – этот тезис возник ещё в христианскую эпоху; (3) текущий офлайновый мир скреплён «морально-нравственным комплексом» и руковод-

ствуется гуманизмом, любовью, духовными ценностями и т. д. Таким образом, теоретическая возможность рациональной дискуссии с алармистами существует.

2.3 Страх перед усилением дискриминации стоит также рассматривать как рациональный, учитывая анализ прошлого опыта. Этот страх во многом совпадает со страхом потерять контроль. Однако в этом случае предполагается, что контроль будет утерян более уязвимой группой, в то время как с доминирующей группой этого не произойдёт. В качестве агента контроля может выступать социальный класс или различные объединения людей (корпорации и т. п.). Одной из самых частых форм презентации данного страха является концепция надзорного капитализма, описанная, например, Шошаной Зубофф. Она рассматривает надзорный капитализм как угрозу автономии «угнетённого» человека (Zuboff, 2019).

В области образования или других сферах применения ИИ не предусмотрено норм, предоставляющих равный доступ к качественным технологиям. При этом сама система обучения ИИ основана на имеющихся данных, которые практически гарантированно будут дискриминировать отдельные группы, так как собраны в заведомо дискриминирующих средах. При этом эта дискриминации может усилиться благодаря утилитарной логике построения систем.

В области образования один из сильнейших барьеров заключается в страхе деградации отдельных групп. Автор рассматривает этот страх как частный случай страха потери контроля. На данный момент часть педагогического сообщества утверждает, что оно напрямую влияет на воспитание и образование детей, но в случае передачи отдельных педагогических функций ИИ (то есть при потере контроля) произойдут катастрофические изменения, в том числе деградация будущих поколений. В свою очередь, это может привести к экзистенциальным барьерам.

Авторы аналитической записки ЮНЕСКО подчёркивают, что делегирование интеллектуальным системам некоторых когнитивных функций приведёт не только к зависимости от технологий, но и к ослаблению когнитивных функций обучающихся. Если не применять навык, есть риск его утратить, – так утверждают авторы, приводя в пример навыки устного счета и ориентации в пространстве без навигатора (Аналитическая записка…, 2020).

Некоторые исследователи среди основных рисков отмечают потерю способности к критическому мышлению (Гуров, Хвесюк, 2022), называя причинами избыток информации (когнитивную перегрузку), ускорение времени (изменений), а также совокупность иных внешних факторов, таких как культура потребления и интенсивное развитие технологий. При этом педагоги и родители всех времён высказывали такие опасения. Исследователи на основе метаанализа обнаружили положительную корреляцию между интел-

лектом и выживаемостью в старшем возрасте (Macarena, Rocío, Valeriano-Lorenzo et al., 2023). Однако не ясно, обусловлено ли это генетически или связано с системой образования.

ИИ может способствовать внедрению и развитию индивидуальных образовательных траекторий, что само по себе вызывает у некоторых педагогов опасения из-за риска социального неравенства. Другие исследователи также видят риски в других областях: применение непроверенных методик, эффект от которых в долгосрочной перспективе недостаточно изучен; дополнительный источник стресса для учеников; неполнота законодательный базы, неспособной на данном этапе гарантировать равный доступ к технологиям (Дегтева, Удалова, 2023).

3. Художественное осмысление когнитивных барьеров на пути доверия к ИИ

В качестве объекта исследований выбраны элементы массовой культуры, а именно – сериалы, представляющие те или иные катастрофические сценарии внедрения ИИ в общественную жизнь: стираются грани между реальным и виртуальным миром; государство, иные институты, злой гений или ИИ могут контролировать жизнь людей.

В рамках исследования были рассмотрены сериалы, отвечающие следующим критериям. Центральной темой сериала (или серии) является доверие ИИ; сериал выходил в 2022-2023 гг.; получил международные премии; относится к наиболее популярным по просмотрам в глобальных рейтингах (как минимум неделю был в топ-10 самых популярных сериалов платформы, на которой проходил показ). Заданным критериям соответствуют четыре сериала: «1899», «Мир Дикого Запада», «Разделение», «Чёрное Зеркало».

В сериале «1899» Барана Бо Одара и Янтье Фризе пассажиры лайнера Керберос, отправляющегося из Старого в Новый Свет, оказываются в искусственной реальности, состоящей из трёх слоёв симуляций и созданной злым гением – «создателем» (остаётся загадкой то, кем именно он является). В художественном произведении используется известная аллегория Платона – миф о пещере. Главная героиня фильма, невролог Маура, упоминает, что мозг человека имеет ограничения и не может понять, что его стимулирует – реальность или её конструкт. Даниэль, вырвавшийся из пещеры Платона, говорит о необходимости просыпаться от вечного сна: отсутствие стимулов из реальности грозит сумасшествием, так как мозг начинает воспринимать симуляцию в качестве реальности. То, что вызывает максимальный комфорт, может оказаться иллюзией. Так как нет чёткой грани между реальностью и её симуляцией, «создателю» удаётся манипулировать сознанием пассажиров. В сериале также поднимается вопрос о том, может ли человек

загнать себя самого в сконструированную реальность, выбрав стратегию эскапизма, и может ли вся эта борьба с целью выживания быть направлена против самого же себя. В этой картине исследуются риски потери контроля.

Другая работа исследует возможность потери человеком контроля в рамках созданной вселенной. Это сериал-антиутопия «Мир Дикого Запада» (Westworld) Джонатана Нолана и Лизы Джой. Он создан по мотивам одноимённой работы 1970-х гг., в которой герои наслаждаются иммерсионным парком развлечений, виртуальной реальностью до того момента, пока не утрачивают контроль перед более сильным антропоморфным ИИ.

В работе поднимаются вопросы свободы воли (может ли человек выйти за рамки «своего программного кода» и переписать свой основной нарратив; справедлива ли концепция «двухпалатного ума») и фиксации человека на собственной ценности (есть ли у человека что-то, что ставит его выше искусственных личностей, которые также чувствуют мир и осознают себя; при каких условиях человечество достойно спасения). Однако центральным в сериале остаётся страх человека перед очередным обновлением ИИ в этой метавселенной; в частности, ИИ дал «свободу воли» искусственным личностям (андроидам), которые получили возможность нанести человеку вред (ранее программа не допускала этого). Таким образом, владельцы корпорации *Delos Inc.* потеряли контроль над своим изобретением. Однако выход андроида из своего нарратива сопровождается экзистенциальным кризисом. В результате андроиды разрабатывают вирус, который вмешивается в реальность человека. Таким образом, андроиды и люди меняются местами в борьбе за право отличать реальность от симуляции.

Следующей работой о страхе потерять контроль над ИИ, на этот раз разработанным корпорациями, является метамодернистский сериал-антиутопия «Разделение» (Severance). Корпорация *Lumon Industries* разделяет сознание своих работников с помощью микрочипа так, что субличности не помнят, что происходит на работе или вне работы, а следовательно, теряют контроль над своим мироощущением: не могут отличить реальность от вымышленного мира. Сюжет строится на попытках героев осознать грань между реальным и дополненным миром, выбраться в реальный мир. В некотором смысле это аналогично тому, что происходит в «Матрице», в «1899» и других кинокартинах, однако в сериале «Разделение» люди идут на это добровольно и осознанно.

При этом корпорация предлагает процедуру разделения под предлогом благих намерений: вживление чипа нужно для того, чтобы дать человеку чувство облегчения на фоне его рабского положения в капиталистической системе. Для некоторых героев это единственный вариант выживания – например, для Марка, который

страдал депрессией после потери близкого человека и поэтому не мог зарабатывать обычным способом.

В сериале подчёркивается тема связи человека и его воспоминаний в контексте концепции *tabula rasa* Джона Локка, согласно которой у человека нет врождённых идей, а Я формируется на основе эмпирического опыта, состоящего из чувств и воспоминаний. При этом герои сериала всё же показывают превалирующие черты своего характера в корпоративном мире.

Разделение приводит к экзистенциальному кризису личности. Само название корпорации отсылает к понятию лиминальности Арнольда ван Геннепа и Виктора Тёрнера, которое означает переходный период между двумя состояниями изменения ценностей, норм, социального статуса и ощущения самости. Переходный период в большинстве культур проходит через три стадии: разделения, грани (порога, лимена) и восстановления. Таким образом, нахождение в состоянии лиминальности – двойственности, перехода, подвешенного и растерянного состояния – рассматривается как неестественное и уязвимое для человека, что отражается в мифах и сказаниях (например, миф об Амуре и Психее). Несмотря на то, что работники корпорации в сериале, оставшиеся в условно свободном мире, не догадываются о том, что будет происходить в рабочее время, они испытывают, как реакцию на эмоциональное насилие, ту же «воскресную хандру», что и обычные работники корпораций, а также необъяснимую тоску и другие симптомы неустойчивой психики.

Примером художественной реализации страха утратить контроль перед государственными и иными общественными системами может служить серия «Чёрного зеркала» под названием «Враг народа» (2016 г., сценарист – Чарли Брукер). Исследуется тема кибербуллинга в ситуации наличия таких инструментов, как автономные дроны-насекомые, подчиняющиеся общественному решению. Хотя в серии и присутствует злой гений, основная часть сюжета построена на последствиях общественного консенсуса. По сюжету всё началось с «игры» в социальной сети, когда для убийства дроном выбирался человек, наиболее часто упоминаемый в качестве жертвы. При этом после того, как люди поняли, что смерти реальны, «игра» набрала популярность.

Таким образом, если имеется инструмент, обладающий силой поражения, будь то насекомое-дрон или ИИ, способный поместить человека между реальным и вымышленным мирами, то нерешённым становится вопрос о том, кому можно доверить управление такой системой. Индивид может оказаться злым гением, общественный консенсус (и государство как его частный случай) – приводить к жестоким решениям в отношении индивида, тогда как условные корпорации ставят целью максимизацию прибыли и не учитывают человеческие факторы, если те противоречат их целям.

Выводы

Выбранные для анализа художественные примеры демонстрируют рациональные когнитивные барьеры, которые связаны со страхами бесконтрольного внедрения технологий, оказывающих сильное влияние на любую социальную сферу, деятельность или развлечение. Анализ подобных художественных примеров высвечивает актуальность проведения отдельных исследований в области систем целеполагания для ИИ, а также важность исследований о том, как внедрение технологий скажется на психике индивида, устоявшихся межличностных и общественных отношениях. Очевидно и то, что влияние технологий на эти системы должно быть регламентировано и контролироваться, но кем? Эта дилемма о том, в чьих руках должна быть сосредоточена абсолютная власть, остаётся нерешённой с древнейших времён. Предложенные ранее решения имеют риски, которые могут привести к серьёзным последствиям для индивида и противоречить гуманистическим принципам. По крайней мере, создатели художественных произведений в своих мысленных экспериментах не смогли решить вопрос об идеальном контроле над общим ИИ и иными автономными системами, потенциально угрожающими здоровью и жизни человека.

Для снятия барьера в массовой культуре демонстрируются ситуации, в которых предлагаются определённые условия функционирования систем ИИ, предотвращающие потерю контроля. Однако все сюжеты строятся на несовершенстве таких систем и ставят вопрос о принципиальной возможности их создания в условиях ограниченности человеческого интеллекта.

С философской точки зрения, помимо актуализации базовых философских проблем, таких как свобода воли, справедливость устройства общественной жизни, требуется приведение логики и базовых понятий в соответствие с новыми технологическими явлениями, ведущими к слиянию техно- и антропосферы.

Таким образом, популярные сериалы продемонстрировали актуальность выделенных в типологии страхов. Автор предлагает рассматривать вышеуказанные рациональные барьеры в качестве инструмента для уточнения и разработки критериев Доверенного ИИ, а также для коммуникации при внедрении связанных с ИИ систем. Рассмотренные художественные работы относятся к наиболее популярным, а значит не теряют актуальности. На взгляд автора, для философов и тех, кто внедряет системы ИИ, значимым является обращение к корневым барьерам, описанным в данной статье, а именно – к страху перед потерей контроля, технологической безработицей и усилением дискриминации.

Информация о конфликте интересов

Автор заявляет об отсутствии конфликта интересов.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Список литературы

Аналитическая записка…, 2020, web – Искусственный интеллект в образовании: изменение темпов обучения. Аналитическая записка ИИТО ЮНЕСКО, 2020. URL: https://unesdoc.unesco.org/ark:/48223/pf0000374947 (дата обращения: 20.02.2024).

Гуров, 2022 – *Гуров О.Н.* Метавселенные – из сумерек во тьму перелётов? // Наука и телевидение, АНОВО "Институт кино и телевидения" (ГИТР). 2022. Т. 18. № 1. С. 11-48.

Гуров, Хвесюк, 2022 – *Гуров О.Н., Хвесюк Л.М.* Угрозы эпидемической (смысловой) безопасности: перспективы образования // Морозовские чтения: международный научно-практический семинар. Иваново: Иванский государственный университет, 2023. С. 38-43.

Дегтева, Удалова, 2023 – *Дегтева Е., Удалова Т.* Возможности и риски индивидуальных образовательных траекторий // Искусственные общества. 2023. Т. 18. № 2. https://doi.org/10.18254/S207751800025887-4

Кутырёв, 2010 – *Кутырёв В.А.* Философия трансгуманизма: учебно-методическое пособие. Нижний Новгород: Нижегородский университет, 2010. 85 с.

European Commission, 2019, web – European Commission. Ethics Guidelines for Trustworthy AI. 2019. URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения: 18.02.2024).

Hinks, 2021 – *Hinks T.* Fear of robots and life satisfaction // International Journal of Social Robotics. 2021. Vol. 13. Pp. 327-340. https://doi.org/10.1007/s12369-020-00640-1

Hötte, Somers, Theodorakopoulos, 2022, web – *Hötte K., Somers M., Theodorakopoulos A.* The fear of technology-driven unemployment and its empirical base // Web publication/site, ROA. 2022. URL: https://voxeu.org/article/fear-technology-driven-unemployment-and-its-empirical-base (дата обращения: 18.02.2024).

Keynes, 1932 – *Keynes J. M.* Economic possibilities for our grandchildren (1930) // Essays in Persuasion. N. Y.: Harcourt Brace, 1932. Pp. 358-373.

Macarena, Rocío, Valeriano-Lorenzo et al., 2023 – *Macarena S.-I., Rocío F.-B., Valeriano-Lorenzo E.L., Botella J.* Intelligence and life expectancy in late adulthood: a meta-analysis // Intelligence. 2023. Vol. 98 (C). https://doi.org/10.1016/j.intell.2023.101738

McKinsey Global Institute, 2016, web – McKinsey Global Institute, "Technology, Jobs, and the Future of Work". Briefing Note for the Fortune Vatican Forum, December 2016, updated February 2017. URL: https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Employment%20and%20Growth/Technology%20jobs%20and%20the%20future%20of%20work/MGI-Future-of-Work-Briefing-note-May-2017.pdf (дата обращения: 18.02.2024).

Zuboff, 2019 – *Zuboff S.* The Age of Surveillance Capitalism: the Fight for a Human Future at the New Frontier of Power. L.: Profile Books, 2019. 704 p.

References

AI in Education: Change at the Speed of Learning. UNESCO Institute for Information Technology in Education Policy Brief (2020). Retrieved February 20, 2024 from https://unesdoc.unesco.org/ark:/48223/pf0000374947

Degteva, E., Udalova, T. (2023). Opportunities and risks of individual educational trajectories. *Artificial Societies*, 2023, 18 (2). https://doi.org/10.18254/S207751800025887-4 (In Russian)

European Commission (2019). *Ethics Guidelines for Trustworthy AI*. Retrieved February 18, 2024 from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Gurov, O.N. (2022). Flight from dusk to darkness. *The Art and Science of Television*, 18 (1), 11-48. (In Russian)

Gurov, O.N, Khsevuk, L.M. (2023). Epidemic (cognitive) safety threats: educational perspectives. *Morozov Readings: International Seminar*. Ivanovo State University Publ. (In Russian), pp. 38-43.

Hinks, T. (2021). Fear of robots and life satisfaction. *International Journal of Social Robotics*, 13, 327-340. https://doi.org/10.1007/s12369-020-00640-1

Hötte, K., Somers, M., Theodorakopoulos, A. (2022). The fear of technology-driven unemployment and its empirical base. *Web publication/site, ROA*. Retrieved February 18, 2024 from https://voxeu.org/article/fear-technology-driven-unemployment-and-its-empirical-base

Keynes, J. M. (1932). Economic possibilities for our grandchildren (1930). In *Essays in Persuasion* (pp. 358-373). Harcourt Brace.

Macarena, S.-I., Rocío, F.-B., Valeriano-Lorenzo, E.L., Botella, J. (2023). Intelligence and life expectancy in late adulthood: a meta-analysis. *Intelligence*, 98 (C). https://doi.org/10.1016/j.intell.2023.101738

McKinsey Global Institute, "Technology, Jobs, and the Future of Work". Briefing Note for the Fortune Vatican Forum, December 2016, updated February 2017. Retrieved February 18, 2024 from https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Employment%20and%20Growth/Technology%20jobs%20and%20the%20future%20of%20work/MGI-Future-of-Work-Briefing-note-May-2017.pdf

Zuboff, S. (2019). *The Age of Surveillance Capitalism: the Fight for a Human Future at the New Frontier of Power*. Profile Books.