DOI: 10.30727/0235-1188-2022-65-1-44-71
Оригинальная исследовательская статья
Original research article

Фетиш искусственного интеллекта*

Д.И. Дубровский
Институт философии РАН, Москва, Россия

А.Р. Ефимов
ПАО «Сбербанк», Москва, Россия,
Национальный исследовательский технологический университет «МИСиС», Москва, Россия

В.Е. Лепский
Институт философии РАН, Москва, Россия

Б.Б. Славин
Финансовый университет при Правительстве РФ, Москва, Россия

Аннотация

В статье представлены основания, позволяющие констатировать фетиш искусственного интеллекта (ИИ). Выделяются принципиальные отличия ИИ от всех предшествующих технологических инноваций, связанные прежде всего с внедрением в когнитивную сферу человека и принципиально новыми неконтролируемыми последствиями для общества. Представлены убедительные аргументы того, что лидеры глобалистского проекта являются главными интересантами и заказчиками фетиша ИИ. Это отчетливо проявляется в работах философов, приближенных к гигантским ИТ-корпорациям, и в мегапроектах этих корпораций. Предлагается к рассмотрению проблема использования возможности ИИ для преодоления нарастающих международных конфликтов и в целом мирового кризиса. В центре внимания оказывается вопрос субъектности, решение которого с позиций антропоморфного подхода к ИИ чревато серьезными негативными последствиями. При наделении субъектностью ИИ неявно снимается ответственность с человека, который применяет эту технологию, а также разрушается сложившаяся законодательная практика. Предлагается представление ИИ как агента, наделенного набором инвари-

* Работа поддержана Российским научным фондом (РНФ), грант № 21-18-00184 «Социогуманитарные основания критериев оценки инноваций, использующих цифровые технологии и искусственный интеллект».

антных упрощенных качеств, которыми обладают естественные субъекты. Среди этих качеств – способность к целеустремленности, своего рода рефлексивность, коммуникативность и упрощенные элементы социальности. Такое представление ИИ как агента (псевдосубъекта) согласуется с принципом распределенного управления в биологии и психологии, который был назван принципом двойного субъекта. В сочетании с системами принципов и онтологий, задаваемых в концепции постнеклассической кибернетики саморазвивающихся сред, это позволит использовать ИИ как средство социальных инноваций при сохранении контроля над технологиями ИИ, а также ставить и решать проблему интеграции образований искусственного и естественного интеллекта при сохранении базовых качеств носителей естественного интеллекта.

Ключевые слова: философия искусственного интеллекта, естественный интеллект, глобалистский проект, антропоморфный подход, субъект, псевдосубъект, постнеклассическая кибернетика.

Дубровский Давид Израилевич – доктор философских наук, профессор, главный научный сотрудник Института философии Российской академии наук.
ddi29@mail.ru
https://orcid.org/ 0000-0003-4392-2526

Ефимов Альберт Рувимович – кандидат философских наук, директор Управления исследований и инноваций ПАО «Сбербанк», заведующий кафедрой инженерной кибернетики Национального исследовательского технологического университета (НИТУ) «МИСиС».
makkawity@gmail.com
https://orcid.org/0000-0001-6857-8659

Лепский Владимир Евгеньевич – доктор психологических наук, главный научный сотрудник сектора междисциплинарных проблем научно-технического развития Института философии РАН.
VELepskiy@mail.ru
https://orcid.org/0000-0002-6893-0234

Славин Борис Борисович – доктор экономических наук, профессор департамента бизнес-информатики Финансового университета при Правительстве РФ.
bbslavin@gmail.com
https://orcid.org/0000-0003-3465-0311

Для цитирования: *Дубровский Д.И., Ефимов А.Р., Лепский В.Е., Славин Б.Б.* Фетиш искусственного интеллекта // Философские науки. 2022. Т. 65. № 1. С. 44–71. DOI: 10.30727/0235-1188-2022-65-1-44-71

The Fetish of Artificial Intelligence*

D I. Dubrovsky
Institute of Philosophy, Russian Academy of Sciences, Moscow, Russia

A.R. Efimov
PJSC Sberbank, Moscow, Russia,
National University of Science and Technology MISiS, Moscow, Russia.

V.E. Lepskiy
Institute of Philosophy, Russian Academy of Sciences, Moscow, Russia

B.B. Slavin
Financial University under the Government of the Russian Federation, Moscow, Russia

Abstract

The article presents grounds for defining the fetish of artificial intelligence (AI). We highlight the fundamental differences of AI from all earlier technological advances, as they are primarily related to its introduction into the human cognitive sphere and generating fundamentally new uncontrollable consequences for society. We provide solid evidence that the leaders of the globalist project are the main beneficiaries of the AI fetish. This is clearly manifested in the works of philosophers who are close to major technology corporations and their mega-projects. We suggest considering the problem of how to use the capabilities of AI to overcome the growing international conflicts and the global crisis. The focus is on the problem of agency, which solution from the standpoint of an anthropomorphic approach to AI is fraught with serious negative consequences. Endowing AI with agency, responsibility is implicitly removed from the person who uses the technology, and the established legislative practice is also destroyed. We present AI as an agent endowed with a set of invariant generalized qualities that is similar to natural subjects. These qualities include: the ability to deliberation, reflexivity, communication and elements of sociability. Such a representation of AI as an

* The work was supported by the Russian Science Foundation, grant no. 21-18-00184 "Social and humanitarian foundations for evaluation criteria for innovations based on digital technologies and artificial intelligence."

agent (pseudo-subject) is consistent with the principle of distributed control in biology and psychology, which was called the principle of a dual subject. In combination with the systems of principles and ontologies specified in the concept of post-nonclassical cybernetics of self-developing environments, this will allow the use of AI as a means of social innovation, while maintaining control over AI technologies. This will also help to pose and solve the problem of integrating formations of artificial and natural intelligence while maintaining the basic qualities of carriers of natural intelligence.

Keywords: philosophy of artificial intelligence, globalist project, anthropomorphic approach, subject, pseudo-subject, post-nonclassical cybernetics.

David I. Dubrovsky – D.Sc. in Philosophy, Professor, Chief Research Fellow, Department of Theory of Knowledge, Institute of Philosophy, Russian Academy of Science.
ddi29@mail.ru
https://orcid.org/ 0000-0003-4392-2526

Albert R. Efimov – Ph.D. in Philosophy, Vice-President of Innovation and Research, PJSC Sberbank; Head of the Department of Engineering Cybernetics, National University of Science and Technology MISiS.
makkawity@gmail.com
https://orcid.org/0000-0001-6857-8659

Vladimir E. Lepskiy – D.Sc. in Psychology, Chief Research Fellow, Department of Interdisciplinary Problems in the Advance of Science and Technology, Institute of Philosophy, Russian Academy of Science.
VELepskiy@mail.ru
https://orcid.org/0000-0002-0590-4020

Boris B. Slavin – D.Sc. in Economics, Professor of the Department of Business Informatics, Financial University under the Government of the Russian Federation.
bbslavin@gmail.com
https://orcid.org/0000-0003-3465-0311

For citation: Dubrovsky D.I., Efimov A.R., Lepskiy V.E., & Slavin B.B. (2022) The Fetish of Artificial Intelligence. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 65, no. 1, pp. 44–71.
DOI: 10.30727/0235-1188-2022-65-1-44-71

Введение

По мере того, как экономика и социальные коммуникации оцифровываются, возрастает и роль цифровых технологий. Тех-

нологические инновации всегда привлекали к себе внимание: так было и во времена появления паровых машин, и в период электрификации, и в эпоху расцвета электроники. Появление новых технологий приводило к трансформации экономических и общественных отношений, и поэтому часто технологиям приписывали субъектные возможности по переустройству мира. Однако время все расставляет на свои места, и предсказания наподобие «Со временем, телевидение перевернет жизнь всего человечества. Ничего не будет: ни кино, ни театра, ни книг, ни газет – одно сплошное телевидение», как правило, не оправдываются. Технологии находят свою нишу в общей экосистеме научного прогресса и не ломают привычных устоев. Поэтому всегда необходим элемент скепсиса при появлении новых революционных технологий.

Изложенный подход можно было бы применить и к современным цифровым технологиям, долго не обсуждая вопрос о возможных трансформациях, к которым они приводят. Однако цифровые технологии имеют особенность, которая их существенно отличает от остальных технологий. Все новшества, появившиеся в доцифровую эпоху, были призваны повысить производительность труда человека либо облегчить и сделать более комфортной его жизнь. Они преобразовывали окружающую среду, делая ее более удобной для человека, но при этом не затрагивали его сущность и когнитивные возможности. Конечно, совершенствование орудий труда вело к тому, что человечество накапливало новые знания, но эти знания по-прежнему находились в головах людей. Книги были лишь инструментом передачи знаний от человека к человеку, но без человека они не имеют смысла.

Цифровая эпоха дала возможность накапливать знания в цифровом виде. В начале это, как и любое нововведение, лишь облегчало использование знаний человеком, поскольку стало возможным читать книги и статьи с электронных носителей, искать в них информацию. Но по мере развития технологий обработки данных оказалось, что аналитические программы способны настолько глубоко анализировать данные (т.н. *data mining*), что можно найти в них то, чего не обнаружишь непосредственно в тексте или в данных. Впервые технологии покусились на самое святое – на когнитивные возможности человека. Если технологии смогут хотя бы частично замещать мыслительные способности

человека, не означает ли это, что технологии могут получить ту самую субъектность по трансформации общества, которой они ранее никогда не обладали? Этот вопрос сегодня в той или иной мере будоражит умы многих ученых.

Нарастающая волна интереса к ИИ и предвестники шторма

Благодаря увеличению производительности вычислительной техники, успехам в области разработки алгоритмов использования искусственных нейронных сетей (в первую очередь машинного обучения, в особенности глубокого) и появлению инструментов для работы с большими данными технологии искусственного интеллекта (ИИ) перешли из разряда перспективных в разряд самых востребованных технологий. Стоит обратить внимание на то, что существующие технологии ИИ в основном направлены на выполнение задач распознавания, предсказаний, имитации человеческой деятельности и не претендуют на создание реального аналога человеческого интеллекта.

Текущие успехи ограничены решением проблем определенных классов: ИИ все еще нельзя доверить самостоятельное принятие сложных решений, от которых зависит жизнь человека. Однако даже в таком виде ИИ стал широко использоваться в предиктивной аналитике, скоринге, распознавании лиц и игровых приложениях. В 2022 году по прогнозу *IDC* крупнейшие мировые компании вложат более 430 млрд долларов США в исследования и разработки в области ИИ[1]. При этом мировой рынок ИИ технологий составит 554,3 млрд долл. США к 2024 году при среднегодовом темпе роста 17,5 % [Прорывные инновации… 2022, 44].

Одновременно с ростом успеха ИИ началась череда принятия стратегий его развития на национальных уровнях. В 2016 году в США Национальным советом по науке и технологиям подготовлены доклады «Подготовка к будущему искусственного интеллекта», «Национальный стратегический план исследований и разработок в области искусственного интеллекта» и «Искусственный интеллект, автоматизация и экономика», определившие стратегию развития ИИ в стране. В июле 2017 года Государствен-

[1] IDC Forecasts Companies to Increase Spend on AI Solutions by 19.6% in 2022 // IDC. 15 February 2022. – URL: https://www.idc.com/getdoc.jsp?containerId=prUS48881422

ный совет Китая опубликовал «План развития искусственного интеллекта нового поколения», рассчитанный на создание индустрии ИИ и превращение Китая в ведущую державу в области ИИ к 2030 году. В 2018 году, на ежегодной встрече в рамках Всемирного экономического форума, премьер-министр Великобритании Тереза Мэй объявила, что собирается сделать «Великобританию мировым лидером в области искусственного интеллекта». В 2019 году в России Президентом В.В. Путиным была утверждена «Национальная стратегия развития искусственного интеллекта на период до 2030 года», а позднее, в рамках программы «Цифровая экономика», выделен проект по развитию ИИ. В последние годы набирает оборот новый мировой тренд – разработка общего искусственного интеллекта, который по своим функциям должен приблизиться к решению задач, специфичных для естественного интеллекта, о чем пойдет речь далее в статье.

Блицкриг ИИ в бизнесе и на государственной ниве совпал с протестами и требованиями об ограничении использования инструментов распознавания, которые нарушали права личности и несли риски принятия ошибочных решений. Так, европейские страны уже при подписании в апреле 2018 года Декларации о сотрудничестве в области искусственного интеллекта делали акцент на необходимости учета социальных, экономических, этических и юридических вопросов. В частности, Европейская комиссия учредила Группу экспертов, которые опубликовали руководящие принципы этики ИИ в апреле 2019 года. В сентябре 2021 года Совет ОНН по правам человека выпустил доклад с рекомендациями «для государств и предприятий относительно разработки и внедрения гарантий для предотвращения и сведения к минимуму вредных последствий и содействия полному использованию преимуществ, которые может предоставить искусственный интеллект» [The Right to Privacy… 2021]. Руководитель Совета (верховный комиссар ООН по правам человека) Мишель Бачелет, комментируя доклад, призвала ввести мораторий на системы ИИ, которые угрожают правам человека, до тех пор, пока правительства не смогут установить гарантии[2].

[2] Мишель Бачелет призвала ввести мораторий на использование систем искусственного интеллекта // Новости ООН. 15 сентября 2021. – URL: https://news.un.org/ru/tags/iskusstvennyy-intellekt/date/2021-09

В докладе Совета ООН по правам человека, в частности, говорится о том, что «процессы принятия решений во многих системах искусственного интеллекта непрозрачны. Сложность информационной среды, алгоритмов и моделей, лежащих в основе разработки и функционирования систем искусственного интеллекта, а также преднамеренная секретность государственных и частных субъектов являются факторами, которые подрывают значимые способы понимания общественностью вопросов влияния систем искусственного интеллекта на права человека» [The Right to Privacy… 2021]. Интересен тот факт, что именно машинное обучение вызывает у защитников прав человека наибольшие опасения, поскольку результат таких вычислений непредсказуем: «Системы машинного обучения добавляют важный элемент непрозрачности; они могут быть способны выявлять закономерности и разрабатывать рецепты, которые трудно или невозможно объяснить. Это часто называют проблемой “черного ящика”. Непрозрачность затрудняет тщательное изучение системы искусственного интеллекта и может стать препятствием для эффективной подотчетности в тех случаях, когда системы искусственного интеллекта наносят ущерб» [The Right to Privacy… 2021]. Кроме того, вызывает большую озабоченность то обстоятельство, что любое машинное обучение очень сильно зависит от тех данных, на которых происходит само обучение. Если эти данные не проверять на соблюдение элементарных этических норм или, что еще хуже, специально подобрать их, чтобы они им противоречили, то системы машинного обучения неизбежно будут выдавать заведомо неэтичные рекомендации.

Главные интересанты и заказчики фетиша ИИ

В этом году в одном из зарубежных изданий опубликована статья «К теории справедливости для искусственного интеллекта» Ясона Габриэля [Gabriel 2022]. Автор работает штатным философом в компании *DeepMind*, принадлежащей *Google*. В статье речь идет о разработке гуманистических принципов использования технологий ИИ. Проанализируем философско-методологические основания позиции автора, их влияние на представления о потенциальных последствиях развития ИИ, а также, что особенно важно, о главных интересантах предложенного подхода.

В первую очередь обратим внимание на идеологическую установку автора статьи, опирающегося на труды Джона Ролза, посвященные его теории справедливости [Rawls 1999]. Суть данной установки отражена в следующем: люди заинтересованы в увеличении своей и уменьшении общей доли выгоды. Это – ярко выраженная позиция идеологии социального либерализма. Для того, чтобы сформировать представление о роли и месте ИИ, автор предлагает понимать основную структуру общества как совокупность социотехнических систем, функционирование которых складывается под все возрастающим влиянием ИИ. Такое представление об ИИ может быть охарактеризовано как яркое проявление технократического редукционизма в организации социальных процессов. В качестве следствия неявно предлагается нарастающее под влиянием ИИ снижение роли государства и общества. Интересантами таких результатов являются лидеры глобалистского проекта. Но кто именно? Это следует конкретизировать.

Лидерами глобалистского подхода и яркими выразителями, защитниками идеологии западного либерализма выступают в первых рядах именно владельцы гигантских ИТ-корпораций, которые, подобно спруту, охватывают мировое коммуникативное пространство. Владельцы таких компаний, как *Facebook*, *Twitter*, *Amazon*, *Google* и др., их наемные теоретики, приписывая системам ИИ *качество субъектности*, пытаются создать впечатление о том, что они наделены некими «божественными» функциями. Они же не просто разрабатывают новые программные продукты, а создают инструменты, которые способны якобы устанавливать справедливость и защищать права человека. Такие инструменты – утопия. Но прокламирование веры в них, ее имитация служит благодатной почвой для защиты глобалистских идей, а в то же время и сокровенных интересов крупнейших мировых производителей ИТ-технологий. Манипулируя массовым сознанием, используя все формы социальных коммуникаций, все средства ИИ, они заявляют о том, что будут для всех нас «сеять разумное, доброе, вечное». Неслучайно основатель *Facebook* Марк Цукерберг решил на базе своих ИТ-продуктов создавать т.н. метавселенную, в которой будет построена новая, более правильная и благополучная жизнь. Однако крайне трудно допустить мысль о том, что Марк

Цукерберг действительно озадачен настолько высокими гуманистическими устремлениями. Судя по его бизнес-деятельности и его опыту манипуляции массовым сознанием в целях достижения максимальной прибыли, сказки о «метавселенной» служат лишь маскировкой для осуществления этих же целей.

Глобалистский подход, опирающийся на идеологию западного либерализма с его псевдогуманистическими клише, убедительно показал свою несостоятельность в условиях пандемии COVID-19 [Лепский 2020] и тем самым продемонстрировал свою негативную роль в деле оценки и разработки способов преодоления нарастающих угроз для человечества, в том числе угроз, связанных с цифровыми трансформациями и развитием ИИ.

Перспективы ИИ в контексте глобального кризиса мировой цивилизации

Именно этот контекст рассмотрения процесса развития ИИ и его социальной значимости, отодвигаемый часто на дальний план (особенно в публикациях, подобных упомянутой выше), приобретает сегодня первостепенное, судьбоносное значение для человечества. Неуклонное нарастание глобального кризиса нашей потребительской цивилизации ведет к разжиганию все более масштабных экономических, политических, всевозможных социальных конфликтов, бескомпромиссной борьбы за ресурсы, за сферы влияния и в конечном итоге – за передел мировой структуры социально-экономической и политической самоорганизации в целом. В этой связи ИИ становится инструментом борьбы в противостоянии различных сил.

Особое внимание привлекает задача создания общего искусственного интеллекта (ОИИ). Ее решение стало в последние годы предметом конкуренции между крупнейшими бигтехами, а в более широком понимании – между государствами-лидерами в области ИИ, в числе которых находятся наши стратегические противники. Это обязывает нас максимально сконцентрировать усилия в данном направлении и добиваться опережения конкурентов. В прошлом году в России вышла первая книга, посвященная специфической проблематике ОИИ [Сильный искусственный интеллект… 2021]. В ней подробно проанализированы главные теоретические и методологические вопросы в области разработки

ОИИ и необходимые для этого научные подходы, обозначены его специфические функции, которые должны быть созданы и, что особенно интересно, предполагаемые масштабные перемены, которые он способен произвести в нарождающемся мироустройстве.

Авторы выделяют две главные способности ОИИ, которые характерны для естественного интеллекта. В отличие от т.н. узкого ИИ, решающего одну определенную задачу, он должен быть интегральным, т.е. способным решать много разных задач. И он должен обрести качество автономности, т.е. способность самостоятельно и эффективно действовать в широком диапазоне сред. Все это качественно повышает деятельные возможности систем ИИ, их использование для военной техники и военных действий, для решения задач производства, управления, планирования, организации экономических процессов, оптимизации самых разнообразных сфер общественной жизни и научных исследований, что чрезвычайно важно для нашей страны в нынешних условиях. Это не менее значимо для осмысления и осуществления грядущих исторических изменений в развитии земной цивилизации, связанных с крушением принципа и практики монополярного мира.

Вместе с тем развитие ОИИ ставит новые сложные теоретические вопросы о его взаимодействии с естественным интеллектом, создании и перспективах гибридного интеллекта, возможном состязании с естественным интеллектом, возможных рисках и угрозах для человека и общества. Достижение высокой степени автономности общим интеллектом создает вероятность появления таких видов и способов его «самодеятельности», которые могут представлять опасность для человека и общества, потребуют разработки новых методов обеспечения безопасности. Перед нами окажется новый аспект все той же проблемы *субъектности* систем ИИ, сохраняющей свою высокую актуальность.

ИИ – это субъект или агент, контролируемый субъектами естественного интеллекта?

Обратимся еще раз к истолкованиям теоретических вопросов о субъектности систем ИИ по отношению к реальным человеческим субъектам. Подход Ясона Габриэля, автора упомянутой выше статьи [Gabriel 2022], заключается в том, чтобы перенести на

ИИ такие же принципы, которые установлены для людей. Автор пишет о том, что «ИИ все больше формирует элементы базовой структуры общества», и, «следовательно, его проектирование, разработка и развертывание потенциально взаимодействуют с принципами правосудия». По мнению Габриэля, «ИИ взаимодействует с поведением людей, принимающих решения, и формирует характер этих практик, включая распределение выгод и бремени среди населения» [Gabriel 2022].

Таким образом, у автора технологии будто «оживают». Он пишет: «Для нашей цели важно учесть, что в современных обществах фоновая справедливость все больше осуществляется алгоритмически» [Gabriel 2022]. Вводя понятие фоновой (видимо, массовой) справедливости, неявно предполагается, что эту справедливость осуществляет алгоритм. Автор продолжает: «Делая оценки или прогнозы на основе прошлого выбора человека и предоставляя решения или рекомендации, которые затем формируют набор возможностей, доступных этому человеку в будущем, эти системы сильно влияют на разворачивающуюся взаимосвязь между индивидуальным выбором и коллективными результатами» [Gabriel 2022]. Снова неявно предполагается, как системы что-то «разумно» делают и что-то предоставляют. Автор придает субъектность технологиям ИИ и требует распространить законы социальных отношений на использование ИИ, которое должно «поддерживать основные свободы граждан, способствовать справедливому равенству возможностей и приносить наибольшую пользу тем, кто находится в наихудшем положении» [Gabriel 2022].

Такого рода антропоморфный подход к ИИ чреват существенными последствиями. Перенося субъектность на ИИ, неявно снимается ответственность с человека, который применяет эту технологию, что нивелирует законодательную практику. Любая технология несовершенна, и она не может сама по себе принимать решение. Ошибка в работе детектора лжи никогда не будет равна нулю, как и при использовании ИИ. Задача людей состоит в том, чтобы учесть ограничения технологий, а не пытаться просто их поставить в какие-то заданные рамки. Неслучайно в судебной практике при всех возможностях криминалистики окончательное решение принимают люди. Системы ИИ ничем не отличаются

от других технологий. Поэтому технологии не должны снимать ответственность с человека, а следовательно, не должны обладать человеческими характеристиками, т.е. быть справедливыми, гуманными и т.д.

Успехи ИИ оказались настолько значительными, что многие решили, что эта технология может заменить человека, только нужно ее поставить в определенные рамки. Фактически ИИ стал фетишем XXI века, который одни стали превозносить как наше будущее, а другие начали вести с ним борьбу. В действительности ИИ, хотя и может распознать то, что не удается человеческому взгляду, или выявить корреляцию, которую не может найти человек, вместе с тем остается далеким от реальных когнитивных возможностей человека. Не исключено, что в будущем удастся создать ИИ, интегрированный в социальную среду, т.н. сильный ИИ, но пока мы далеки от этого, и придавать ему субъектность не только неправильно, но опасно.

Этические вопросы развития ИИ

Вопросы этики использования ИИ широко обсуждаются сегодня как общественными деятелями, политиками, так и учеными. Большое внимание привлек скандал, возникший в связи со статьей об этике ИИ, подготовленной к публикации сотрудником компании *Google* Тимнит Гебру, одним из ведущих мировых экспертов по проблемам необъективности алгоритмов и извлечения данных (*data mining*) [Hao 2020]. Научные и коммерческие позиции принципиально разошлись, и, как следствие, исследовательница покинула компанию *Google*.

Дискуссии ведутся и среди российских ученых. Наряду с конструктивными предложениями отдельные авторитетные ученые предлагают сомнительные идеи, призывают к политизации науки и технологий, наделяя технологии свойствами «патриотизма». Так, 23 ноября 2021 года состоялось заседание Президиума Российской академии наук, на которой академики обсуждали в том числе и этическую сторону ИИ, и возможности для ИИ быть доверенным. Один из выступающих заявил: «С моей точки зрения, ИИ должен быть не только доверенным, о чем сегодня говорилось, но и патриотичным, т.е. он должен в первую очередь работать на интересы страны, а не против нее» [Славин 2021, 34].

Сегодня многие протестуют против технологий распознавания лиц (в некоторых городах США такое распознавание запрещено, Европарламент тоже предложил запретить технологии распознавания лиц). Однако вредно не распознавание как таковое, а использование его в противоправных целях. Человек не скрывает свои болезни перед врачом потому, что доверяет ему и рассчитывает на помощь. Необходимо регулировать законы применения технологий, в том числе и ИИ, серьезно наказывать, если они были использованы во вред человеку или незаконно. Россия сегодня оказалась во многом слабо защищенной перед мошенниками, которые используют средства коммуникаций для обмана доверчивых граждан. Все, что сегодня власть может сделать, – это предупреждать население о новых способах мошенничества. При этом власть вполне эффективно, в том числе и с использованием ИИ, борется с политическими противниками. Этические проблемы ИИ должны решать не путем ограничения технологий, а в первую очередь путем ограничения действий людей, которые их неправомерно используют.

Решение этических вопросов применения ИИ – очень сложное, комплексное дело, которое невозможно выполнить лишь посредством принятия декларативных кодексов этики ИИ. Необходимо учитывать, что даже наши этические принципы – от библейских заповедей до кодекса строителя коммунизма – есть лишь точки в бесконечном пространстве морально-этических решений, в котором мы движемся ежедневно. Дискуссии об этике ИИ только начинают разворачиваться, и сообщества философов, инженеров должны тесно сотрудничать для выработки ответов на вызовы времени.

Сегодня крайне важно рассматривать этические проблемы применительно к развитию ИИ в более широком концептуальном плане: под углом реальных особенностей функционирования этических норм в социальной жизнедеятельности, реального состояния нравственности массового сознания, индивидуальных, групповых, институциональных субъектов. Сплошь и рядом, всегда, на всех этапах истории человечества ясно наблюдался разрыв между знанием этических норм и их исполнением. Вспомним древнеримскую поговорку: «Вижу лучшее и одобряю, но следую худшему». Слишком часто интерес оказывался выше нравственных установлений, а обман подавлял правду и вил себе уютные

гнезда в самых высоких этических наставлениях. Говорить о нравственном прогрессе в развитии человечества нет достаточных оснований (обстоятельные материалы, посвященные данной теме, представлены во множестве философских исследований [Дубровский 2007]). Все эти обстоятельства следует учитывать, если мы рассуждаем на тему «Этика ИИ», причем как в отношении создателей систем ИИ, так и в отношении пользователей.

При попытках моделирования принципов этики и воплощения их в работе ИИ ситуация осложняется тем, что совокупность этических норм не может быть упорядочена в виде четкой иерархической структуры, допускающей альтернативный выбор. Выбор практически всегда может быть сделан лишь при рассмотрении и оценке конкретных условий. Поэтому указанное моделирование представляется возможным только в специально определенных частных случаях.

Вместе с тем проблема субъектности в области разработок ИИ по-прежнему заслуживает пристального внимания. Так или иначе способность системы ИИ решать сложные задачи мы связываем с описаниями некоторых функций естественного интеллекта. Если ИИ нецелесообразно представлять в качестве субъекта, аналогичного человеку, то как понимать и определять ИИ, которому передаются возможности принятия решений в определенных ситуациях и который побеждает чемпиона мира по шахматам или игры в го? Наиболее адекватным подходом, на наш взгляд, может быть представление ИИ как агента, наделенного набором инвариантных упрощенных качеств, которыми обладают естественные субъекты. К таким качествам можно отнести подобие целеустремленности, своего рода рефлексивность, коммуникативность и упрощенные элементы социальности. Представление ИИ как агента (псевдосубъекта) согласуется с принципом распределенного управления в биологии и психологии, который был назван принципом двойного субъекта [Лепский 1998]. Это представление ИИ в сочетании с системами принципов и онтологий, задаваемых в концепции постнеклассической кибернетики саморазвивающихся сред, позволяет использовать ИИ как средство социальных инноваций, при сохранении контроля над технологиями ИИ, а также ставить и решать проблему интеграции образований искусственного и естественного интеллекта при

сохранении базовых качеств носителей естественного интеллекта [Lepskiy 2018; Лепский 2021].

Заключение

ИИ переживает фазу бурного роста. Масштабные цифры эффектов от внедрения не должны вводить нас в заблуждение: нынешний период – это лишь начало тотального проникновения ИИ в нашу жизнь. Именно поэтому нам следует очень внимательно относиться к возможным когнитивным искажениям при исследованиях возникающих феноменов. К примеру, антропоморфизм в применении к ИИ может заставить нас легко поверить в ложную субъектность машины. Настоящая работа авторов, специалистов в философии и методологии, служит призывом к более широкому диалогу и переходу от создания кодексов поведения ИИ к созданию следующего поколения ИИ, действующего вместе с человеком и для человека.

The Fetish of Artificial Intelligence

Introduction

As currently economy and social communications are undergoing digitalization, the role of digital technologies also increases. Technological innovations have always attracted attention. This was the case during the advent of steam engines, later during electrification, and in the heyday of electronics. The emergence of new technologies led to a transformation of economic and social relations. Due to this, technology was considered to possess capabilities of restructuring the world. However, time puts everything in its place, and the prediction from the classical Soviet movie script has not come true: “Over time, television will change the life of all mankind. There will be nothing else: no cinema, no theater, no books, no newspapers. Only television.” Technologies fill their niches in the general ecosystem of scientific progress but do not destroy the traditional bases. Therefore, we should use some skepticism to new revolutionary technologies.

Such an approach could be applied to modern digital technologies, thus postponing any lengthy discussions they may lead to. However,

digital technologies have one feature that significantly distinguishes them from all others. All the innovations that appeared in the pre-digital era were designed to increase human labor productivity or to make our life easier and more comfortable. They transformed the social environment, making it more convenient for humans, but at the same time they did not affect the very essence of human beings or their cognitive capabilities. Of course, the improvement of tools led to accumulation of new knowledge, but this knowledge still remained in the people's minds. Books were just tools for transferring knowledge from person to person.

Our epoch has made it possible to accumulate knowledge in digital form. At first this only facilitated personal use of knowledge: it became possible to read books and articles in electronic formats and to search for information. But with the development of data processing technologies, it turned out that analytical software could analyze data in such a way (the so-called data mining) that made it possible to find what was not directly presented in the text or in the data. For the first time in the history of mankind, technology encroached on the most sacred thing: man's cognitive capabilities. If technology can replace (even partially) the ability of a human to think, does this not mean that technology can get the very agency for the transformation of society, which they have never possessed before? This question today stirs the minds of many scientists in various degrees.

Increasing interest in AI and harbingers of a storm

Thanks to the increasing performance rates in modern computing, advances in the development of algorithms for artificial neural networks (primarily machine learning and especially deep learning) and the emergence of tools for working with big data, artificial intelligence (AI) technologies have outgrown the category of promising technologies, to enter the category of the most popular ones. The existing AI technologies mainly focus on tasks of recognition, prediction and imitation of human activity, and do not claim to createa real analogue of human intelligence.

The current successes are limited to solving problems of certain classes, and AI still cannot be trusted to independently make complex decisions on which life depends. However, even in this form, AI is now widely used in predictive analytics, scoring, face recognition

and game applications. IDC predicts that in 2022, the world's largest companies will invest more than 430 billion US dollars in AI research and development[3]. Also, the global AI technology market will amount to 554.3 billion US dollars by 2024 with an average annual growth rate of 17.5% [Gokhberg, Efimov, & Milshina 2022, 44].

Simultaneously with the growing success of AI, adoption of strategies for its development at national levels began. In 2016, the U.S. National Science and Technology Council prepared the reports "Preparing for the future of Artificial Intelligence," "National Artificial Intelligence Research and Development Strategic Plan," and "Artificial Intelligence, Automation, and the Economy," which determined the strategy for AI development in the nation. In July 2017, China's State Council issued a document entitled "A New Generation Artificial Intelligence Development Plan," designed to create an artificial intelligence industry and to turn China into a leading power in the field by 2030. At the 2018 annual meeting of the World Economic Forum, British Prime Minister Theresa May announced that she was going to make the UK a world leader in artificial intelligence. In 2019, Russian President V.V. Putin approved the "National Strategy for the Development of Artificial Intelligence for the period up to 2030," and then a project for the development of AI was drafted within the framework of the Digital Economy program. In recent years, a new global trend has been gaining popularity – the development of Artificial General Intelligence (AGI), whose functions approach the solution of tasks that are specific to natural intelligence. We discuss these issues in more detail below.

Simultaneously with the breakthrough of AI in business and government programs, there started protests and demands to limit the use of recognition tools that violated individual rights and posed risks of making wrong decisions. Thus, already at the signing of the Declaration of cooperation on Artificial Intelligence in April 2018, European states emphasized the need to take into account various social, economic, ethical and legal issues. In particular, the European Commission established a Group of Experts who published guidelines on AI ethics in April 2019. In September 2021, the UN Human Rights Council released a report with recommendations "for states and busi-

[3] IDC Forecasts Companies to Increase Spend on AI Solutions by 19.6% in 2022. *IDC*. 2022, February 15. Retrieved from https://www.idc.com/getdoc.jsp?containerId=prUS48881422

nesses to develop and implement safeguards to prevent and minimize harmful effects and promote the full use of the benefits that artificial intelligence can provide" [OHCHR 2021]. The head of the Council (UN High Commissioner for Human Rights) Michelle Bachelet, commenting on the report, called for a moratorium on artificial intelligence systems that threaten human rights, until governments can provide guarantees[4].

The report of the UN Human Rights Council states that decision-making processes in many AI systems are not transparent. The complexity of the informational environment, algorithms, and models underlying the development and operation of artificial intelligence systems, as well as deliberate secrecy of public and private actors are factors that undermine ways for the public to understand the impact of AI systems on human rights. Interestingly, it is machine learning that causes the greatest concern among defenders of human rights, since the result of such computations is unpredictable: machine learning systems will inevitably reduce essential transparency; they may be able to identify patterns and develop recipes that are difficult or impossible to account for. This is often referred to as the "black box" problem. Insufficient transparency makes an AI system hard for examination and may hinder effective accountability in cases where AI systems cause damage [OHCHR 2021]. Besides, it is a very sensitive issue that all machine learning depends very much on the data fed during the training itself. If such data are not checked for compliance with basic ethical values, or even worse, if they are specially selected so that they contradict the values, then machine learning systems will inevitably issue deliberately unethical recommendations.

The main stakeholders and beneficiaries of the AI fetish

Recently there was a paper by Iason Gabriel, "Toward a Theory of Justice for Artificial Intelligence" (the author is a Staff Research Scientist and an expert in philosophy at DeepMind, a company owned by Google) [Gabriel 2022]. The publication is devoted to the development of humanistic principles of using AItechnologies. Let us analyze the

[4] Urgent Action Needed over Artificial Intelligence Risks to Human Rights. *UN News.* 2021, September 21. Retreved from https://news.un.org/en/story/2021/09/1099972

philosophical and methodological foundations of the author's position, their influence on the ideas about the potential consequences of AI development, as well as the principal stakeholders of the proposed approach.

First of all, let us look at the ideological stance of the author, based on John Rawls's theory of justice [Rawls 1999]. The essence of this approach is depicted as follows: people are interested in increasing their own profits and decreasing the common share of benefits. This is a pronounced position of the ideology of social liberalism. To present the role of AI, the author treats the basic structure of society as a set of sociotechnical systems, whose functioning develops under the increasing influence of AI. This presentation of AI is a vivid manifestation of technocratic reductionism in the organization of social processes. Here, as a consequence, the author proposes a growing decline in the role of the state and society under the influence of AI. The leaders of the globalist project are interested in such results. But who are they, exactly? This needs specifying.

It is easy to see that the leaders of the globalist approach and the active proponents and defenders of the ideology of Western liberalism are the owners of huge IT corporations who, like an octopus, command the global informational space. The owners of such companies as Facebook, Twitter, Amazon, Google and their hired theorists, attributing the quality of agency to AI systems, seek to create the impression that they are endowed with certain "god-like" functions. After all, they do not just develop new software products but create tools that are supposedly able to establish justice and protect human rights. These tools are utopic. But proclaiming faith in them or imitation of such faith provides a good reason for protecting globalist ideas, and at the same time, the innermost interests of the world's largest IT manufacturers. Manipulating public consciousness, using all forms of social communication, and all AI tools, they declare that they will strive for common benefit. It is no coincidence that Facebook's founder Mark Zuckerberg decided to create a Metaverse based on his IT products, in which a new, "better arranged and prosperous" life will be built. However, it is extremely difficult to assume that Zuckerberg is really concerned about such high humanistic aspirations. Judging by all his business activities and his experience of manipulating mass consciousness in order to achieve a maximum profit, the tales of the "Metaverse"

serve only as a disguise for the implementation of the same mercenary goals.

The globalist approach, based on the ideology of Western liberalism with its pseudo-humanistic cliches, convincingly showed its inconsistency in the COVID-19 pandemic [Lepskiy 2020] and demonstrated its negative role in assessing and developing ways to overcome the growing threats to humanity, including threats related to digital transformation and the development of AI.

Prospects of AI in the context of the global crisis of the world civilization

It is this context of considering the AI development and its social significance, often pushed to the background (especially in publications of the type we discussed above), that is now of paramount and fateful importance for humanity. The ongoing global crisis of our consumer civilization leads to emergence of increasingly large-scale economic, political, and social conflicts, uncompromising struggle for resources, for spheres of influence, and ultimately for a redistribution of the entire global structure of socio-economic and political self-organization. And in this regard, AI becomes a weapon in the confrontation of various forces.

In this regard, as already noted above, special attention is drawn to the task of creating an AGI, the solution of which has become in recent years the subject of competition between the largest hightech companies, and also between the states leading in the field of AI, among which are our strategic opponents. This forces us to concentrate our efforts in this direction as much as possible, and to get ahead of our competitors. Recently, in Russia there was published the first book dedicated to AGI [Vedyakhin et al. 2021]. It analyzes in detail the main theoretical and methodological issues of the development of AGI and the scientific approaches necessary for this, identifies those specific functions that must be created and, most interestingly, the supposed large-scale changes that it is able to produce in the emerging new world order.

The authors identify two key capabilities of AGI, which are similar to natural intelligence. Unlike the narrow AI that solves one specific task, it must be integral, i.e., capable of solving many types of tasks. And it must acquire autonomy, i.e., the ability to act effectively and independently in various environments. All these increase the opera-

tional capabilities of AI systems, their use for military equipment and military operations, for solving problems of production, management, planning, organization of economic processes, optimization of a wide variety of spheres of public life and scientific research, which is extremely important for our country, under the current circumstances. And this is just as important for understanding and implementing those future historic changes in the development of the Earth's civilization, which are associated with the collapse of the principle and practice of the monopolar world.

At the same time, the development of AGI will raise new complex theoretical questions concerning its interactions with natural intelligence, the creation and prospects of hybrid intelligence, its possible competition with natural intelligence, possible risks and threats to man and society. Achieving a high degree of autonomy of AGI may lead to the appearance of such types and methods of its "amateur activity," which may pose a danger to man and society and will require new methods of ensuring security. Here we will face a new aspect of the same problem of *agency* of AI systems, which remains highly relevant.

Is AI a subject or an agent controlled by natural intelligence subjects?

Let us turn once again to the interpretations of theoretical questions about the agency of AI systems in relation to real human subjects. The approach of Iason Gabriel, the author of the above-mentioned article [Gabriel 2022], is to transfer to AI the same principles that are established for humans. The author writes that "AI increasingly shapes elements of the basic structure" of society, and, consequently, "the development and deployment of AI systems represent a new site for the operation of principles of distributive justice." According to the author, "AI interacts with the behavior of human decision-makers to shape the character of these practices, including how they distribute benefits and burdens across the population" [Gabriel 2022].

It turns out that for the author technologies seem to "come alive": "What is important for our purpose," he writes, "is that in modern societies, background justice is increasingly mediated algorithmically" [Gabriel 2022]. Here, Gabriel introduces the concept of background justice (apparently, mass-oriented justice) and assume that this justice is subject to an algorithm. Further, he states: "By making assessments

or predictions based upon an individual's past choices, and by providing decisions or recommendations that then shape that person's opportunity set, these systems exert a strong influence on the unfolding relationship between individual choices and collective outcomes" [Gabriel 2022]. As we can see, it is again assumed that the systems perform "intelligently" and create something. The author grants agency to AI technologies and demands that we extend the laws of social relations to AI, which "should support citizens' basic liberties, promote fair equality of opportunity, and provide the greatest benefit to those who are worst-off" [Gabriel 2022].

This anthropomorphic approach to artificial intelligence is fraught with grave consequences. Firstly, by transferring agency to AI, responsibility is implicitly removed from the human person who uses this technology, which hinders legal procedures. All technology is imperfect and cannot make a decision by itself. The error in the operation of the lie detector will never equal zero, and the same is true for using AI. It is the job of people to take into account the limitations of technology, and not just try to put them in some given framework. It is not by chance that in judicial practice, granted all the possibilities of criminology, humans do make the final decision. AI systems are not different from other technologies, so technologies should not remove responsibility from a person, and therefore should not have human characteristics (i.e., be fair, or humane, et al.).

The triumph of AI turned out to be so significant that many have decided that this technology can replace a human person, it only it needs placing in a certain framework. In fact, artificial intelligence has become a fetish of the 21st century, which some people extol as our future, while others fight against it. In fact, AI can recognize what the human eye cannot and identify a correspondence that a person cannot find. Yet itlags far behind the true cognitive capabilities of a human. Maybe in future it will be possible to create an AI integrated into the social environment, the so-called "general" AI, but now we are too far from this point and granting it agency is not only wrong, but also quite dangerous.

Ethical issues of AI development

The ethics of using AI are now widely discussed not only by public figures, politicians, but also by scholars. Much attention was drawn

to the public argument over the draft paper on the ethics of artificial intelligence by a Google employee, Timnit Gebru, one of the world's leading experts on the problems of bias in algorithms and data mining [Hao 2020]. The scientific and commercial positions fundamentally diverged, and as a result, the researcher left Google.

Discussions are also underway among Russian scientists. Along with constructive proposals, some recognized scholars offer very debatable and call for politicization of science and technology, endowing technology with the properties of "patriotism." Thus, on November 23, 2021, a meeting of the Presidium of the Russian Academy of Sciences was held, at which the academicians discussed, among other things, the ethical aspect of AI and the possibilities for AI to be trusted. One of the speakers stated, "From my point of view, artificial intelligence should not only be trusted, as was discussed today, but also should be patriotic, that is, it should primarily work for the interests of the country, and not against it" [Slavin 2021].

Today, many people protest against facial recognition technologies (in some US cities such recognition is prohibited, the European Parliament also proposed banning facial recognition technologies). However, it is not the recognition itself that is harmful, but its use for illegal purposes. People do not hide their illnesses from a doctor because they trusts this specialist and expect help. It is necessary to regulate the laws of use of technologies, including artificial intelligence, and to seriously punish those who harm a person or act illegally. Today, Russia appears poorly protected against fraudsters who use means of communication to deceive gullible citizens. All that the authorities can do today is to warn the population about new techniques of fraud. At the same time, the government is quite effective (also in the use of AI) in fighting its political opponents. The ethical problems of AI should be solved not by limiting the technologies, but primarily by limiting the actions of those people who misuse them.

Solving ethical issues of AI application is a very complex matter that cannot be accomplished only by adopting declarative codes of AI ethics. It is necessary to take into account that even our ethical principles, whatever ones we adopt – from the biblical commandments to the code of the builder of Communism – are only a few points in the infinite space of moral and ethical decisions in which we travel daily. Discussions of the ethics of AI are just beginning and professional com-

munities of philosophers and engineers should work closely together to develop answers to the new challenges of the time.

It should be emphasized that it is extremely important now to consider ethical problems in relation to the development of AI in a broader conceptual sense – from the viewpoint of the real functioning of ethical norms in our social life, the real state of morality in mass consciousness, in individual, collective, and institutional subjects. After all, there has always existed a gap between the knowledge of ethical norms and their implementation. Let us recall the Latin saying: *Video meliora, proboque, deteriora sequor* ("I see better things, and approve, but I follow worse"). Much too often, personal interest turned out to suppress moral precepts, and deception made a cozy nest for itself in the highest ethical instructions. There are no sufficient reasons to talk about moral progress in the development of mankind (detailed materials on this subject are presented in many philosophical studies [Dubrovsky 2007]). All these circumstances must be taken into account when we discuss the topic of "AI Ethics," both in relation to the creators of AI systems and to their users.

When trying to model the principles of ethics and implement them in AI operation, another challenge is that ethical norms cannot be organized as a clear hierarchical structure that allows an alternative choice. Here, the choice always depends on considering and evaluating the specific circumstances. Therefore, such modeling is only possible in specially selected cases.

At the same time, the problem of agency in AI design requires closer study. After all, in one way or another, the ability of an AI system to solve complex problems is associated with descriptions of some functions of natural intelligence. If it is not advisable to represent AI as a subject similar to a person, then how do we understand and define AI, when it is allowed to make decisions in certain situations and which defeats the world champions in chess and Go games? In our opinion, the most adequate approach may be the representation of AI as an agent endowed with a set of invariant simplified qualities that natural agents possess. These qualities include: deliberation, reflexivity, communicativeness, and simplified sociability. This representation of AI as an agent (or pseudo-subject) is consistent with the principle of distributed control in biology and psychology, called the principle of a dual subject [Lepskiy 1998]. This representation of AI, combined with systems of

principles and ontologies set in the concept of post-non-classical cybernetics of self-developing environments, allows using AI as a means of social innovation, while maintaining control over AI technologies as well as posing and solving the problem of integrating artificial and natural intelligence formations, yet preserving the basic qualities of natural intelligence [Lepskiy 2018; Lepskiy 2021].

Conclusion

AI is undergoing a phase of rapid growth. And yet, the impressive statistics of its successes should not mislead us: the current period is only the beginning of a total penetration of AI into our lives. That is why we should be very sensitive to possible cognitive distortions in the study of new phenomena. For example, anthropomorphism applied to AI can make us believe in the agency of the machine, which is false. The present paper, written by authors who specialize in philosophy and methodology, initiates a broader dialogue and a transition from creating codes of conduct for AI to creating the next generation of AI collaborating with humans and for them.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Дубровский 2007 – *Дубровский Д.И.* О нравственном прогрессе и нравственном регрессе (К проблематике развития морального сознания) // Философские науки. 2007. № 11. С. 81–102.

Дубровский 2021 – *Дубровский Д.И.* Задача создания Общего искусственного интеллекта и проблема сознания // Философские науки. 2021. Т. 64. № 1. С. 13–44.

Лепский 1998 – *Лепский В.Е.* Концепция субъектно-ориентированной компьютеризации управленческой деятельности. – М.: Институт психологии РАН, 1998.

Лепский 2020 – *Лепский В.Е.* Рефлексия пандемии COVID-19: субъектно-ориентированный подход // Экономические стратегии. 2020. № 8 (174). С. 66–71.

Лепский 2021 – *Лепский В.Е.* Искусственный интеллект в субъектных парадигмах управления // Философские науки. 2021. Т. 64. № 1. С. 88–101.

Прорывные инновации… 2022 – Прорывные инновации: человек 2.0: доклад к XXIII Ясинской (Апрельской) международной научной конференции по проблемам развития экономики и общества, Москва, 4–8 апреля 2022 г. / под ред. Л.М. Гохберга, А.Р. Ефимова, Ю.В. Мильшиной. – М.: НИУ ВШЭ, 2022.

Сильный искусственный интеллект… 2021 – Сильный искусственный интеллект: На подступах к сверхразуму / А. Ведяхин и др. − М.: Интеллектуальная литература, 2021.

Славин 2021 − *Славин Б.Б.* Может ли искусственный интеллект быть справедливым // БИТ. 2021. № 10 (113). С. 32−35.

Gabriel 2022 − *Gabriel I.* Toward a Theory of Justice for Artificial Intelligence // AI & Society. 2022. Vol. 151. No. 2. P. 218−231.

Hao 2020 – *Hao K.* We read the paper that forced Timnit Gebru out of Google. Here's what it says // MIT Technology Review. 4 December 2020. – URL: https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

Lepskiy 2018 – *Lepskiy V.* Evolution of Cybernetics: Philosophical and Methodological Analysis // Kybernetes. 2018. Vol. 47. No. 2. P. 249–261.

Rawls 1999 − *Rawls J.* A Theory of Justice. − Cambridge, MA: Harvard University Press, 1999.

The Right to Privacy… 2021 − The Right to Privacy in the Digital Age: Report of the United Nations High Commissioner for Human Rights. A/HRC/48/31 // Office of the High Commissioner for Human Rights. 13 September 2021. − URL: https://www.ohchr.org/en/documents/thematic-reports/ahrc4831-right-privacy-digital-age-report-united-nations-high

REFERENCES

Dubrovsky D.I. (2007) On Moral Progress and Moral Regress. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* No. 11, pp. 81−102 (in Russian).

Dubrovsky D.I. (2021) The Task of Creating a General Artificial Intelligence and the Problem of Consciousness. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 64, no. 1, pp. 13−44 (in Russian).

Gabriel I. (2022) Toward a Theory of Justice for Artificial Intelligence. *AI & Society.* Vol. 151, no. 2, pp. 218−231.

Gokhberg L.M., Efimov A.R., & Milshina Y.V. (Eds.) (2022) *Breakthrough Innovations: Man 2.0: Report to the 23rd Yasin (April) International Scientific Conference on Problems of Economic and Social Development, Moscow, April 4−8, 2022.* Moscow: National Research University Higher School of Economics (in Russian).

Hao K. (2020, December 4) We read the paper that forced Timnit Gebru out of Google. Here's what it says. *MIT Technology Review.* Retrieved from https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

Lepskiy V.E. (1998) *The Concept of Subject-Oriented Computerization of Managerial Activity.* Moscow: Institute of psychology Russian academy of sciences (in Russian).

Lepskiy V. (2018) Evolution of Cybernetics: Philosophical and Methodological Analysis. *Kybernetes.* Vol. 47, no. 2, pp. 249–261.

Lepskiy V.E. (2020) Reflection of the COVID-19 Pandemic: a Subject-Oriented Approach. *Ekonomicheskie strategii.* No. 8, pp. 66−71 (in Russian).

Lepskiy V.E. (2021) Artificial Intelligence in Subject-Oriented Control Paradigms. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 64, no. 1, pp. 88–101 (in Russian).

Office of the High Commissioner for Human Rights (OHCHR) (2021, September 13) *The Right to Privacy in the Digital Age: Report of the United Nations High Commissioner for Human Rights.* A/HRC/48/31. Retrieved from https://www.ohchr.org/en/documents/thematic-reports/ahrc4831-right-privacy-digital-age-report-united-nations-highRawls J. (1999) *A Theory of Justice.* Cambridge, MA: Harvard University Press.

Slavin B.B. (2021) Can Artificial Intelligence Be Fair? *BIT Journal.* No 10 (113), pp. 32−35 (in Russian).

Vedyakhin A.A. et al. (2021) *Strong Artificial Intelligence: On the Approaches to the Supermind.* Moscow: Intellektualnaya Literatura (in Russian).