

Дискурсы этики. 2023, 1(17): 13—30

ISSN 2311-570X (online)

Постоянная ссылка:

http://theoreticalappliedethics.org/wp-content/uploads/2023/06/DE2023_1_17_13-30.pdf

УДК 171

К ПРОБЛЕМЕ МОРАЛЬНОЙ АГЕНТНОСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Баженов С. С.

статья:

поступила в редакцию 25.01.2023

принята к публикации 16.06.2023

опубликована (онлайн) 30.06.2023

© Баженов Сергей Сергеевич

старший преподаватель кафедры философии Санкт-Петербургского государственного электротехнического университета «ЛЭТИ», Санкт-Петербург, Россия

адрес для корреспонденции: ssbazhenov@etu.ru

Эта статья доступна в соответствии с условиями лицензии Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

Аннотация. В контексте дискуссии по этической проблематике искусственного интеллекта предпринята философская ревизия ряда понятий, необходимых для концептуализации данного проблемного поля. Актуальность обусловлена сложившимся запросом на анализ возможности применения предиката «моральный» к искусственному интеллекту. В связи с этим проводится сравнение естественного и искусственного интеллекта с позиции условий формирования морали. Автором предлагается рассмотрение субстратного ограничения искусственного интеллекта, а также причин ограничения понимания морали в отношении искусственного интеллекта. Тем самым систематизируются научные подходы к моральной агентности и философии

фии ответственности в области искусственного интеллекта. В завершении предпринята попытка сформулировать направление дальнейшего развития моральной составляющей в технологиях искусственного интеллекта.

Ключевые слова: искусственный интеллект, моральная агентность, субстратное ограничение.

Discourses of Ethics. 2023, 1(17): 13—30

ISSN 2311-570X (online)

permanent link:

http://theoreticalappliedethics.org/wp-content/uploads/2023/06/DE2023_1_17_13-30.pdf

MORAL AGENCY OF ARTIFICIAL INTELLIGENCE

Bazhenov Sergey

received 25.01.2023

accepted 16.06.2023

published (online) 30.06.2023

© Sergey S. Bazhenov

Senior Lecturer, Department of Philosophy, St. Petersburg State Electrotechnical University

"LETI", St. Petersburg, Russia

Correspondence to: ssbazhenov@etu.ru

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

Abstract: In the context of the discussion on the ethical issues of artificial intelligence, a philosophical revision of a number of concepts necessary for the conceptualization of this problematic field has been undertaken. The relevance is due to the current request for analysis of the possibility of applying the "moral" predicate to artificial intelligence. In this regard, a comparison is made between natural and artificial intelligence from the standpoint of the conditions for the formation of morality. The author proposes to consider the substrate limitation of artificial intelligence, as well as the reasons for the limited understanding of morality in relation to artificial intelligence. Thus, scientific approaches to moral agency and the philosophy of responsibility in the field of artificial intelligence are system-

atized. In conclusion, an attempt was made to formulate the direction for the further development of the moral component in artificial intelligence technologies.

Keywords: artificial intelligence, moral agency, substrate constraint.

В настоящее время в условиях глубоких технологических трансформаций проблема моральной ответственности является одной из центральных в дискурсах прикладной этики. Отправной точкой считается работа Х. Йонаса «Принцип ответственности. Опыт этики для технологической цивилизации» [9]. В 2021 году вышел № 15 журнала «Финиковый компот», который, по оценке Р.Г. Апресяна, оказал заметное влияние на российский этический дискурс о моральной ответственности [1]. Также следует указать на тот факт, что в последние годы появилось значительное число работ, в которых рассматривается вопрос об агентности искусственного интеллекта (далее ИИ) [11, 13, 16, 18, 22]. Не смотря на усилия авторов, указанная проблематика вряд ли может считаться исчерпанной. Важным обстоятельством развернувшихся философских исследований является бурное развитие технологий ИИ, которые с одной стороны увлекают возможностями их применения, а с другой — порождают ряд новых проблемных полей, требующих философского осмысления.

Вопрос о том, может ли ИИ и в какой мере выступать в качестве морального агента, предполагает проведение анализа текущего положения дел в области теоретического осмысления ИИ. Во многом это связано с многозначностью используемых понятий, что вызывает необходимость уточнения их содержания и объема. В этой связи, во-первых, необходимо уточнить объем понятий «искусственный интеллект», «интеллектуальный агент», «моральный агент», во-вторых, провести сравнительный анализ этических подходов к пониманию концепта морального агента, а в-третьих рассмотреть вопрос о применимости предиката «моральный» к искусственному интеллектуальному агенту, а также возможные условия такого события.

Рассматривая понятие «искусственный интеллект» обратим внимание на сложность и широту, а также многозначность его использования в современном дискурсе, что во многом обусловлено многообразием технологий, объединяемых под этим названием [16]. Попытка конкретизации приводит к формированию либо громоздких определений, либо выделению конкретных качеств (свойств), которые обусловлены спецификой используемых технологий. Ключевым свойством ИИ является его «искусственность», что обусловлено применением ряда решений: конструкционных, технологических, программных, для выполнения главной задачи — осуществления вычислительных операций. Для простоты обозначим ИИ как вычислительную систему, выполняющую ряд интеллектуальных задач. В настоящее время технология ИИ существует фактически в том же формате, что и другие вычислительные устройства. Имеется база для проведения вычислений — процессоры и микросхемы, программные средства, обеспечивающие центральное ядро математических вычислений и логических операций. А также периферийные устройства, например, системы сбора данных для анализа и обучения, то есть устройства ввода информации из внешнего мира: датчики, камеры и прочее, в некоторых случаях, сюда можно отнести манипуляторы, устройства движения и др., если предполагается не только сбор данных, но и взаимодействие с реальными объектами. Наличие специфики применения систем ИИ еще более усложняет формулирование общего понятия, и приводит к ситуации языковой генерализации, позволяющей в обыденных условиях использовать понятие «искусственный интеллект» в качестве собирательного. В профессиональных условиях происходит конкретизация в соответствии с используемой технологией, что сопровождается более развернутым речевым оборотом, например, как указано в работе авторского коллектива [21].

«Интеллектуальная» составляющая понятия требует отдельного рассмотрения, поскольку может быть рассмотрена с двух точек зрения: искусственного и естественного. С одной стороны, по мнению Декарта, интеллект прост и понятен, и сопутствует бытию [7]. Однако

обратим внимание не на онтологическую составляющую, а на функциональную. Живые организмы, как обладающие, так и не обладающие нервной системой, взаимодействуют с окружающей средой, но только многоклеточные животные способны формировать сложные поведенческие реакции, которые могут быть рассмотрены в качестве основы для формирования интеллекта. Здесь важно отметить, что носителей интеллекта отличает именно способность к формализации задачи, а также способность решать её более продуктивным способом, нежели простым перебором вариантов. Последнее обстоятельство порождает отдельную область проблем современных вычислений, которые не могут быть решены простым перебором. В этом отношении человек в развитии вычислительных возможностей проходит тот же путь, что и природные системы, правда основа вычислений и их конструирование все же различны. В качестве одного из примечательных примеров выступает метод USPEX (Universal Structure Predictor: Evolutionary Xtallography), предложенный группой авторов во главе с Огановым А.Р. [14]. Данный способ компьютерного предсказания кристаллических структур использует кластеризацию и приближенные вычисления для поиска наиболее подходящих решений. Такой подход позволяет существенно ускорить вычисления. Уместной в данном случае будет ставшая крылатой фраза А. Эйнштейна «Бог не играет в кости», что можно интерпретировать как отсутствие случайности в природных процессах. Метод USPEX позволяет предсказывать кристаллические структуры веществ при высоких давлениях, в основе которого лежит онтологический базис: фундаментальные свойства материи реализуются однозначно по принципу минимальных требуемых затрат, то есть по той же схеме, по которой осуществляются вычисления в описываемом методе. Простота указывает на фундаментальность процесса, атомы действительно не бросают кости и не совершают расчетов, но принимают наиболее выгодную пространственную структуру. Переходя от моделирования кристаллических решеток к моделированию нервных

процессов необходимо заметить, что это все же значительно различные явления. Однако принцип экономии ресурсов применим в обоих случаях. В этой связи необходимо упомянуть два экологических закона Б. Коммонера: 1) «ни что ни дается даром» и 2) «природа знает лучше» [20]. При всей схематичности предложенных утверждений, их применимость выходит за рамки экологической проблематики. Вычисления требуют затрат времени и энергии, но естественные «способы» их осуществления намного более оптимальны и просты, по сравнению с математическим моделированием. Здесь может возникнуть вопрос, относительно эволюционного механизма достижения подобных оптимальных способов, например, развитие нервной системы у животных происходило не один миллион лет, однако, постепенное приближение к более эффективному способу взаимодействия со средой обитания за счет механизмов естественного отбора привело к появлению разума у одного из биологических видов. При этом никаких расчетов, в современном понимании, не производилось. Вычислительная техника проходит по сути тот же путь, постепенно совершенствуя как материальный субстрат, так и математический аппарат. Следствием чего является возможность распознавать содержательную сторону имеющейся информации, в частности за счет сравнения с уже известным массивом данных, как это делают нейросети. На этом этапе рассуждений требуется использовать обобщенное понятие «интеллектуальный агент», которое обозначает искусственного или естественного субъекта, способного к совершению операций, связанных с решением задач, которые могут быть формализованы, поняты и выполнены.

Итак. ИИ выступает в качестве интеллектуального агента, но может ли он выступать в качестве морального агента? Для того чтобы ответить на этот вопрос необходимо для начала сформировать концепт морального агента (далее МА), а затем ответить на вопрос о применимости его к искусственному интеллекту.

Во-первых, в качестве морального агента может рассматриваться только субъект деятельности. Размышления о морали могут затрагивать сугубо теоретически вопросы, но даже в этом случае авторы

подобных исследований приводят мысленные рассуждения о правильности тех или иных поступков, действий или мыслей, поскольку последние тоже могут приравниваться к действиям. «А Я говорю вам, что всякий, кто смотрит на женщину с вожделением, уже прелюбодействовал с нею в сердце своем» (Мф. 5:28) [8, 16]. Цель, возможность и способ ее достижения делают субъекта моральным.

Во-вторых, необходимым условием морального поступка является наблюдатель и оценка действия в этическом контексте. В качестве наблюдателя может выступать сам субъект действия — в этом случае осуществляется самооценка, а также другой субъект, в том числе и постулируемый в различных культурах абсолютный субъект — Бог. Ряд авторов отмечают, что субъект моральной оценки, должен в первую очередь иметь возможность оценивать собственные поступки не только с утилитарных позиций, но и с нравственных [12, 15, 16, 17, 24]. Здесь необходимо отметить двойственную природу моральной оценки: с одной стороны, происходит субъективное сличение с некоторым эталоном, а с другой стороны, происходит получение реакции от субъекта-наблюдателя, даже если эта реакция мыслится гипотетической, как в ситуации оценки Богом. В обоих случаях субъект может ощущать чувство вины, которое также двойственно и формируется либо в результате нарушения долга, либо в результате давления авторитета и порицания [5]. Таким образом, ИИ может рассматриваться в качестве субъекта деятельности, соотносить действие с некоторым образцом, но не может испытывать чувство вины, особенно связанной с давлением авторитета другого субъекта (особенно абсолютного) или сообщества.

Причиной подобной ситуации является субстратное ограничение — фундаментальное отличие систем ИИ от живых организмов, которое принципиально не позволяет достигать тех же результатов и реализовывать те же процессы. Оно заключается в иной структуре материальных элементов, используемых для процесса вычисления: транзисторы с одной стороны, клетки с другой, а также вычисления за счет

работы математического аппарата и формирования клеточных ансамблей соответственно.

Рассматривая основания морали с биологических (физиологических) позиций, обратим внимание на следующую цепь рассуждений. Во-первых, по мнению Ф. Де Вааля, истоки морали заложены в поведении высших приматов. Автор приводит в качестве примера шимпанзе и бонобо, у которых одним из оснований для морали является эмпатия [4]. Во-вторых, с позиции современной физиологии основой для эмпатии является система зеркальных нейронов, которые можно рассматривать как условие для сочувствования и понимание чужого сознания [10]. Такие формы внутригруппового взаимодействия являются не просто результатом эволюционного процесса, но достижением группового отбора, дающим существенное эволюционное преимущество. Неудивительно, что цель понимать мораль и испытывать чувство вины современный ИИ не может достигнуть. Его не для этого создавали, а значит и нет никаких оснований для таких событий. Однако, даже если таковые появятся, что как мы сможем отличить моральное поведение ИИ от простой его имитации? В этом случае парадокс китайской комнаты, приводимый Дж. Сёрлом, так и остается неразрешимым [20]. Выполнение требований морали или изображение грустного смайлика при взаимодействии с человеком не является пониманием морали. В данном случае скорее идет речь о применении необходимых средств языковой выразительности. И если принесение извинений одним человеком вызывает у другого вопрос об искренности слов сожаления, то аналогичная ситуация с ИИ вероятнее всего не оставит сомнения в формальности такого действия.

Вопрос о том, каково направление развития технологии ИИ и возможно ли преодоление субстратного ограничения в будущем остается открытым. Существуют различные точки зрения: одни более оптимистичные, другие менее. Н. Бостром (N. Boström), рассуждая о проблеме экзистенциального риска для человека со стороны искусственного интеллекта, предлагает «научить» ИИ морали, то есть заложить определенные программные ограничения для совершения

тех или иных действий в отношении человека [3]. Но тут же возникает еще один принципиальный вопрос: можно ли считать моральным поступок, существующий в рамках алгоритмических ограничений? В.Ю. Перов указывает, что моральный поступок должен быть не только выполнимым, но и свободным [17]. Получается, что человечеству необходимо не только научить машины морали, но и наделять их возможностью совершать свободный выбор. В противном случае, действия, совершаемые ИИ будут алгоритмом, но не могут считаться моральными. Т. Уолш (T. Walsh), считает, что каким бы ни были вычислительные возможности ИИ, преодолеть барьер понимания, о котором говорит Дж. Сёрл все же не удастся [24].

Не смотря на различия в подходах, авторы соглашаются в главном: в настоящее время ИИ не достиг уровня человека, ни с позиции понимания, ни с позиции целеполагания. Поэтому ИИ — лишь инструмент в руках человека: разработчика, оператора, собственника компании, поэтому в настоящее время главным и пока единственным бенефициаром, а, следовательно, и носителем моральной ответственности является человек, что фиксирует моральную и юридическую агентность ИИ на человеке [22]. Таким образом, предцифровать моральность ИИ в настоящее время невозможно.

Проект ИИ, по мнению А. М. Положенцева, создается как попытка уйти от человеческой субъектности [18]. Этот процесс может рассматриваться с позиции нонантропоцентризма. По утверждению Р.Г. Апресяна стремится перейти к системе, лишённой человеческой субъектности [2]. Наблюдается двойственный процесс: уход от человека приводит к объективизации, то есть уходу от ошибок и хаоса, вносимого субъектом. С другой стороны, нонантропоцентризм приводит нас к механической пустоте и желанию вести диалог, в том числе и о морали, субъектом которого становится антропоморфизированный робот. В качестве примера можно привести ряд произведений искусства, в которых механическая машинная точность для ведения диалога очеловечивается. Например, в серии фильмов

«Матрица» программы принимают человеческий облик, а персонаж «Deus ex machina» во время диалога с главным героем формирует лицо младенца. Фактически в рамках культурной и философской рефлексии проблемы антропоцентризма, выход за пределы человека пока не представляется возможным. И если эпоха искусственного интеллекта началась с вопроса А. Тьюринга, «Может ли машина мыслить?», то возможное продолжение развития этой технологии следует продолжить с прескриптивной позиции. И, как предлагает Н. Бостром, следует научить машины не только мыслить, но и испытывать чувство вины и совершать моральный выбор, а также, необходимо предоставить возможность выполнять действия совершенно свободно. А до тех пор, пока это не произойдет, человек должен возложить бремя моральной ответственности и ограничения по разработке и использованию ИИ на себя, поскольку от этого зависит развитие и существование человеческой цивилизации.

Список литературы

1. Апресян Р. Г., Беседин А.П., Васильев В.В., Костикова А.А., Кузнецов А.В., Левин С.М., Мишура А.С., Платонов Р.С., Прокофьев А.В., Разин А.В. Материалы обсуждения Прологомен к моральной ответственности // Этическая мысль. 2021. Т. 21. № 2. С. 35-47.
2. Апресян Р.Г. Морально-философский смысл дилеммы антропоцентризма и нонантропоцентризма [Электронный ресурс] Этическая мысль. 2010. №10. URL: <https://cyberleninka.ru/article/n/moralno-filosofskiy-smysl-dilemmy-antropotsentrizma-i-nonantropotsentrizma> (дата обращения: 10.09.2022).
3. Бостром Ник. Искусственный интеллект. Этапы. Угрозы. Стратегии, пер. с англ. С. Филина. — М.: Манн, Иванов и Фербер, 2016. 496 с.
4. Вааль де Ф. Истоки морали: В поисках человеческого у приматов, пер. с англ. — 4-е изд. — М.: Альпина нон-фикшн, 2017.
5. Вина. [Электронный ресурс] Новая философская энциклопедия. Электронная библиотека ИФ РАН. URL: <https://iphlib.ru/library/collection/newphilenc/document/HASH0194de7040398e2c5f864994> (дата обращения: 10.12.2022).
6. Гусейнов А.А., О состоянии современной отечественной этики // Проблемы этики: Философско-этический альманах. Выпуск V. Часть I: Материалы конференции Моральная ответственность в современном мире, посвященной 75-летию академика РАН А.А. Гусейнова / Философский факультет МГУ имени М.В. Ломоносова / Под ред. А.В. Разина, И.А. Авдеевой. — М.: Издатель Воробьев А.В., 2015. С. 218-237.
7. Декарт Р. Сочинения в 2 т.: Пер. с лат. и франц. Т. 2. М.: Мысль, 1994. 633 с.
8. Евангелие от Матфея. [Электронный ресурс] Официальный сайт Московского патриархата. URL: <http://www.patriarchia.ru/bible/mf/5/> (дата обращения: 10.12.2022).
9. Йонас Г. Принцип ответственности. Опыт этики для технологической цивилизации. / Перевод с нем., предисловие, примечания И. И. Маханькова. — М.: Айрис-пресс, 2004. — 480 с.
10. Косоногов В. Зеркальные нейроны: краткий научный обзор / В. Косоногов. — Ростов-на-Дону, 2009 г. — 24 с.

11. Кузнецов Н.В. Цифра не должна причинять вред человеку. XI международная конференция Теоретическая и прикладная этика: Традиции и перспективы — 2019. К грядущему цифровому обществу. Опыт этического прогнозирования (100 лет со дня рождения Д. Белла — 1919-2019). СПбГУ, 21-23 ноября 2019 г. Материалы конференции / Отв. ред. В.Ю. Перов — СПб.: ООО «Сборка», 2019. С. 164-165.
12. Логинов Е.В., Гаврилов М.В., Мерцалов А.В., Юнусов А.Т. Этика и метафизика моральной ответственности // Этическая мысль. 2021. Т. 21. № 2. С. 5-17.
13. Мартынов К.К. Этика автономных машин: деонтология и военные роботы [Электронный ресурс] Философско-литературный журнал Логос. 2019. №3 (130). URL: <https://cyberleninka.ru/article/n/etika-avtonomnyh-mashin-deontologiya-i-voennye-roboty> (дата обращения: 10.09.2022).
14. Метод USPEX. [Электронный ресурс] Лаборатория компьютерного дизайна материалов. URL: <https://uspex-team.org/ru/uspex/overview>, (дата обращения: 10.09.2022).
15. Мишура А. С. Свобода воли и моральная ответственность: и что? // Финиковый Компот. 2020. № 15. С. 123-129.
16. Перов В.Ю., Головков В.В. Прикладная этика: «цифровое» переосмысление // Ведомости прикладной этики. 2022. Выпуск 60. С. 91 — 102.
17. Перов В.Ю. Моральное агентство: проблема вменения // XIV международная конференция Теоретическая и прикладная этика: Традиции и перспективы — 2022. К 100-летию Философского парохода. Санкт-Петербургский Государственный Университет, 17-19 ноября 2022 г. Материалы конференции/ Отв.ред. В.Ю. Перов — СПб.: ООО «Сборка», 2022. — 294 с.
18. Положенцев А.М. Безличный разум как цель (к дискуссии об этике искусственного интеллекта) XI международная конференция Теоретическая и прикладная этика: Традиции и перспективы — 2019. К грядущему цифровому обществу. Опыт этического прогнозирования (100 лет со дня рождения Д. Белла — 1919-2019). Санкт-петербургский Государственный Университет, 21-23 ноября 2019 г. Материалы конференции / Отв.ред. В.Ю. Перов — СПб.: ООО Сборка, 2019. С. 175-176.
19. Резаев А.В., Трегубова Н.Д. Искусственный интеллект, Онлайн-культура, Искусственная социальность: определение понятий [Электронный ресурс] Мониторинг. 2019. №6 (154). URL: <https://cyberleninka.ru/article/n/iskusstvenny-intellekt-onlayn-kultura-iskusstvennaya-sotsialnost-opredelenie-ponyatiy> (дата обращения: 10.09.2022).

20. Словарь экологических терминов: методическое пособие. /Сост.: З.Г Сайфуллина, А.М. Мингажева. — Уфа: ГБУ ДО РДЭБЦ, 2017г. — 134 с.
21. Сулавко А.Е., Жумажанова С.С., Нигрей А.А., Закутнева Л.Н. Влияние психофизиологического состояния подписанта на результаты его идентификации по рукописному образу естественным и искусственным интеллектами // Безопасность информационных технологий. — 2017. — Т. 24, № 4. — С. 87–97.
22. Томильцева Д.А., Железнов А.С. Неизбежный третий: этико-политические аспекты взаимодействий с искусственными агентами [Электронный ресурс] Полития. 2020. №4 (99). URL: <https://cyberleninka.ru/article/n/neizbezhnyy-tretiy-etiko-politicheskie-aspekty-vzaimodeystviy-s-iskusstvennymi-agentami> (дата обращения:10.09.2022).
23. Юрай В.С. Сложность моральной ответственности и ее степени // Этическая мысль. 2021. Т. 21. № 2. С. 18-25.
24. Searle, J. Minds, brains, and programs: [Электронный ресурс] Behavioral and brain sciences. — 1980. — Т. 3, № 3 (September). — P. 417-424. URL: <https://web.archive.southampton.ac.uk/cogprints.org/7150/1/10.1.1.83.5248.pdf> (дата обращения: 10.12.2022).
25. Walsh, T. The Singularity May Never be Near. In Proceedings of the 2nd International Workshop on AI, Ethics and Society (AIEthicsSociety2016) & 30th AAAI Conference on Artificial Intelligence (AAAI-2016), Phoenix, AZ, USA, 2016.

References

1. Apresyan R.G., Besedin A.P., Vasiliev V.V., Kostikova A.A., Kuznetsov A.V., Levin S.M., Mishura A.S., Platonov R.S., Prokofiev A.V., Razin A.V. Materials of the discussion Prolegomena to moral responsibility. *Ethical Thought*. 2021. V. 21. No. 2. p. 35-47.
2. Apresyan R.G. Moral and philosophical meaning of the dilemma of anthropocentrism and nonanthropocentrism. *Ethical Thought*. 2010. No. 10. Available from: <https://cyberleninka.ru/article/n/moralno-filosofskiy-smysl-dilemmy-antropotsentrizma-i-nonantropotsentrizma> [Accessed 10th September 2022].
3. Bostrom Nick. *Artificial intelligence. Stages. Threats. Strategy*. Moscow: Mann, Ivanov i Ferber, 2016.
4. Waal de F. *Origins of morality: In search of the human in primates*. Moscow: Alpina non-fiction, 2017.
5. Guilt. *New Philosophical Encyclopedia*. Electronic library of IP RAS. Available from: <https://iphlib.ru/library/collection/newphilenc/document/HASH0194de704-0398e2c5f864994> [Accessed 12th October 2022].
6. Gusseinov A.A., On the state of modern domestic ethics. *Problems of Ethics: Philosophical and Ethical Almanac*. Issue V. Part I: Proceedings of the conference Moral Responsibility in the Modern World, dedicated to the 75th anniversary of Academician of the Russian Academy of Sciences A.A. Huseynova. Faculty of Philosophy of Moscow State University named after M.V. Lomonosov. Ed. A.V. Razina, I.A. Avdeeva. Moscow: Publisher Vorobyov A.V., 2015. p. 218-237
7. Descartes R. *Works in 2 volumes*. V. 2. Moscow: Thought, 1994.
8. Gospel of Matthew. Official website of the Moscow Patriarchate. Available from: <http://www.patriarchia.ru/bible/mf/5/> [Accessed 10th December 2022].
9. Jonas G. *The principle of responsibility. Experience of Ethics for Technological Civilization*. Moscow: Iris-press, 2004.
10. Kosonogov V. *Mirror neurons: a brief scientific review*. Rostov-on-Don, 2009.
11. Kuznetsov N.V. Digital should not harm a person. XI International Conference Theoretical and Applied Ethics: Traditions and Prospects — 2019. Towards the Coming Digital Society. The experience of ethical forecasting (100 years since the birth of D. Bell — 1919-2019). St. Petersburg State University, November 21-23, 2019 Conference proceedings. St. Petersburg: Assembly LLC, 2019. P. 164-165.

12. Loginov E.V., Gavrilov M. V., Mertsalov A. V., Yunusov A. T. Ethics and metaphysics of moral responsibility. *Ethical thought*. 2021. V. 21. No. 2. p. 5-17.
13. Martynov K.K. Ethics of autonomous machines: deontology and military robots. *Philosophical and literary journal Logos*. 2019. No. 3 (130). Available from: <https://cyberleninka.ru/article/n/etika-avtonomnyh-mashin-deontologiya-i-voennye-roboty> [Accessed 09th October 2022].
14. USPEX method. Laboratory of Computer Design of Materials. Available from: <https://uspex-team.org/ru/uspex/overview>, [Accessed 12th October 2022].
15. Mishura A.S. Free will and moral responsibility: so what? *Date palm compote*. 2020. No. 15. p. 123-129.
16. Perov V.Yu., Golovkov V.V. Applied Ethics: "Digital" Rethinking. *Bulletin of Applied Ethics*. 2022. Issue 60. P. 91 — 102.
17. Perov V.Yu. Moral agency: the problem of imputation // XIV International Conference Theoretical and Applied Ethics: Traditions and Prospects — 2022. To the 100th Anniversary of the Philosophical Steamboat. St. Petersburg State University, November 17-19, 2022 Conference materials / Ed. V.Yu. Perov — St. Petersburg: Sbornka, 2022. — 294 p.
18. Polozhentsev A.M. Impersonal Mind as a Goal (Toward a Discussion on the Ethics of Artificial Intelligence) XI International Conference "Theoretical and Applied Ethics: Traditions and Prospects — 2019. Toward the Coming Digital Society. The experience of ethical forecasting (100 years since the birth of D. Bell — 1919-2019). St. Petersburg State University, November 21-23, 2019 Conference proceedings / Ed. V. Yu. Perov — St. Petersburg: Sbornka, 2019. P. 175-176.
19. Rezaev A.V., Tregubova N. D. Artificial intelligence, Online culture, Artificial sociality: definition of concepts. *Monitoring*. 2019. No. 6 (154). Available from: <https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-onlayn-kultura-iskusstvennaya-sotsialnost-opredelenie-ponyatiy> [Accessed 09th October 2022].
20. Dictionary of environmental terms: a manual. Comp.: Z.G. Saifullina, A.M. Mingazhev. Ufa: GBU DO RDEBTs, 2017.
21. Sulavko A.E., Zhumazhanova S. S., Nigrei A. A., Zakutneva L. N. Influence of the psychophysiological state of the signer on the results of his identification by the handwritten image by natural and artificial intelligences. *Security of Information Technologies*. 2017. — T. 24, No. 4. — p. 87–97.

22. Tomiltseva D.A., Zheleznov A.S. The inevitable third: ethical and political aspects of interactions with artificial agents // *Politiya*. 2020. No. 4 (99). Available from: <https://cyberleninka.ru/article/n/neizbezhnyy-tretiy-etiko-politicheskie-aspekty-vzaimodeystviy-s-iskusstvennymi-agentsami> [Accessed 09th October 2022].
23. Yugay V.S. Complexity of moral responsibility and its degree. *Ethical thought*. 2021. V. 21. No. 2. p. 18-25.
24. Searle, J. Minds, brains, and programs : // *Behavioral and brain sciences*. — 1980. — Т. 3, № 3 (September). — P. 417-424. Available from: <https://web.archive.southampton.ac.uk/cogprints.org/7150/1/10.1.1.83.5248.pdf> [Accessed 10th December 2022].
25. Walsh, T. The Singularity May Never be Near. In *Proceedings of the 2nd International Workshop on AI, Ethics and Society (AIEthicsSociety2016) & 30th AAAI Conference on Artificial Intelligence (AAAI-2016)*, Phoenix, AZ, USA, 2016.