Научная статья
УДК 17:004.8
DOI: 10.18101/1994-0866-2022-1-67-79

ЭТИКА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ПРИНЦИП ОТВЕТСТВЕННОСТИ ГАНСА ЙОНАСА

© Бадмаева Маина Харлановна
аспирант,
Бурятский государственный университет имени Доржи Банзарова
Россия, 670000, г. Улан-Удэ, ул. Смолина, 24а
badmaevamaina@gmail.com

Аннотация. Искусственный интеллект (ИИ) сегодня стал неотъемлемой частью жизни современного общества. Перспективы его применения требуют глубокого осмысления, поскольку чреваты неблагоприятными последствиями для человека и человечества. В статье рассмотрены наиболее актуальные аспекты применения искусственного интеллекта, описаны возможные сложности и преимущества, обусловленные внедрением технологий искусственного интеллекта. Понимание причин, смысла и последствий такого внедрения, по мнению автора статьи, требует разработки новых, адекватных предмету изучения теоретических методов и положений. Автор обосновывает мысль о том, что учет и соблюдение этических принципов в процессе разработки технологий искусственного интеллекта и ее применения могут содействовать наиболее гармоничному развитию человечества. В этой связи необходим пересмотр традиционной этики и поиск таких ее оснований, которые в перспективе обеспечат плодотворное взаимодействие человека и систем ИИ. Попытка создания подобной этической системы взглядов была осуществлена немецким философом-экзистенциалистом Гансом Йонасом. В статье автор исследует этику ответственности Ганса Йонаса в качестве новой регулирующей составляющей процесса применения технологий искусственного интеллекта. Этика ответственности способна уменьшить экзистенциальные риски, сохраняя возможность дальнейшего внедрения ИИ в жизнь человека.

Ключевые слова: технократизм, искусственный интеллект, узкий искусственный интеллект, этика, этика искусственного интеллекта, этика ответственности, Ганс Йонас, экзистенциализм, страх за будущее, аксиологическая онтология.

Для цитирования
Бадмаева М. Х. Этика искусственного интеллекта: принцип ответственности Ганса Йонаса // Вестник Бурятского государственного университета. Философия. 2022. Вып. 1. С. 67–79.

Сегодня мы еще не обладаем достаточно точным и общепринятым определением искусственного интеллекта. В 1956 г. Джон Маккарти, американский информатик, предпринял попытку дать определение искусственному интеллекту. По его словам, искусственный интеллект представляет науку, работающую над машинами и компьютерными программами, напоминающими своими действиями работу человеческого интеллекта, способными реагировать на окружающую действительность как человек, т. е. оценивать результаты «чувственного» познания, рассуждать, обучаться, адаптироваться, принимать решения и т. д. В рамках данной статьи под искусственным интеллектом будет подразумеваться узкий ИИ

(слабый ИИ, прикладной ИИ). Узкий ИИ — это система, назначением которой является решение каких-либо конкретных интеллектуальных задач (например, распознавание речи). Сегодня большинство систем ИИ являются узкими, поскольку они еще не способны решать любые интеллектуальные задачи, как это делал бы гипотетический универсальный ИИ.

В настоящее время искусственный интеллект начинает все больше проникать в человеческую жизнь. Технологии искусственного интеллекта довольно быстро расширяют свой спектр применения: диагностика и создание лекарств в медицине, беспилотные автомобили, домашний помощник, педагогика, криминалистика, машинный перевод, финансовые модели и т. д. Постепенно в связи с более широким применением искусственного интеллекта его участие в социальных, политических, экономических процессах и взаимодействиях становится все более масштабным и привычным. Оценивая значимость технологий искусственного интеллекта и перспективы их развития, мы приходим к выводу о необходимости учета и переосмысления основополагающих этических принципов, следование которым, на наш взгляд, обязательно в процессе разработки и применения систем ИИ.

Обзор последних достижений в области применения систем ИИ демонстрирует все более широкое их внедрение во все сферы нашей жизнедеятельности. Такие компании, как Google, Apple, Facebook, Microsoft, Uber, Яндекс, активно применяют системы ИИ и превращаются благодаря им во все более влиятельных игроков современной глобальной экономики.

В области сельского хозяйства компания Autonomous Tractor Cooperation представила беспилотный трактор Spirit, предназначенный для агрономических работ [1]. Система ИИ позволяет машине ездить по маршруту, ранее пройденному с водителем. Компания Cognitive Technologies продемонстрировала беспилотный трактор с системой компьютерного зрения, который может работать даже по ночам [2].

В области государственной службы и охраны правопорядка была протестирована программа Series Finder, выявляющая на основании анализа шаблона преступлений потенциальных преступников и предсказывающая тем самым возможные преступления в будущем [3]. Для раскрытия уже совершенных краж и предсказания новых программа использовала данные отдела анализа преступности Кембриджского полицейского управления (КПУ), позволившие ей выявить своего рода modus operandi преступника, т. е. тип поведения, набор привычек, которым следует правонарушитель. На основании полученных девяти поведенческих моделей, начав буквально с пары преступлений, Series Finder смогла восстановить большинство совершенных преступлений, зарегистрированных КПУ. Программа также выявила еще девять преступлений, укладывающихся в рамки созданных ею моделей, о которых КПУ ранее ничего не было известно.

В сфере государственной службы широкое применение получили системы распознавания изображений. Как только точность распознавания лиц преодолела рубеж в 94%, в обществе возникли серьезные опасения в связи с опасностью нарушения конфиденциальности, вторжения в частную жизнь и возможной дискриминации прав граждан [4] как неизбежных последствий использования данных технологий.

В распознавании голоса, речи и текста, в машинном переводе ИИ тоже имеет существенные успехи. Microsoft продемонстрировала, что ИИ способен транскрибировать речь лучше, чем профессиональные стенографисты. Сайт Google переводит тексты на 37 языков, а речь — на 32. WaveNet от Google и Deepspeech от Baidu являют собой примеры глубоких нейронных сетей, способных автоматически синтезировать голос [5]. Умные устройства в домах могут значительно повлиять на изменение наших повседневных привычек. Например, на сайте Indiegogo идет сбор средств на «первого в мире социального робота для дома» Jibo, который способен общаться с людьми, распознавать лица своих владельцев и запоминать их предпочтения[1]. В сфере образования большая часть технологий ИИ пока задействована в качестве систем проверки посещаемости занятий и выполнения заданий, оценивания и анализа экзаменационных ответов, составления персональных планов обучения. Так, система с ИИ AutoTutor занимается обучением языкам программирования, физике и развитию критического мышления. Такие онлайн-платформы, как Udasity, EdX, оценивают написание тестов и эссе. В банковской отрасли кредитование частично поручено системам ИИ, которые оценивают платежеспособность клиентов. Также системы ИИ помогают в обнаружении мошеннических транзакций, анализе уровня налогов и доходов, управлении финансами.

В сфере транспортной системы активно тестируют свои беспилотные автомобили такие компании, как Google, General Motors, Tesla, BMW, Ford. Систему автопилота считают призванной обеспечить безопасность пассажиров в стандартных ситуациях, в случае возникновения неожиданных, критических обстоятельств управление передается в руки человека. Умное освещение дорожного покрытия с системами ИИ может параллельно анализировать состояние дорог, предотвращать возникновение «пробок» на наиболее сложных и востребованных участках дорог. В области промышленности роботы Sawyer и Baxter от компании Rethink Robotics выполняют различные монотонные операции, начиная от упаковки коробок до выравнивания изделий на конвейерной ленте, оптимизируя тем самым процесс производства [6].

Медицина является одной из основных отраслей, в которой на ИИ возлагаются большие надежды. Благодаря системам ИИ активно развивается отрасль медицины, позволяющая заблаговременно предотвратить возникновение вспышек новых заболеваний. Также с помощью систем ИИ идет разработка новых лекарственных препаратов, происходит оптимизация труда медработников, проводится диагностика самых различных заболеваний.

Однако наряду с очевидными преимуществами применения ИИ возникает все больше проблем, решение которых уже не может быть ограничено только обсуждением технологических и инженерных вопросов. Все большую актуальность в этой связи приобретают вопросы этики. Так, в Германии на правительственном уровне 30 сентября 2016 г. усилиями Федерального министерства транспорта и

[1] URL: https://www.jibo.com (дата обращения: 18.01.2022). Текст: электронный.

цифровой инфраструктуры Германии[1] и IEEE[2] была создана специальная комиссия по этике автоматизированного вождения. В комиссию вошли специалисты из области философии, права и социальных наук, оценки технологий, защиты прав потребителей и автомобильной промышленности, а также разработки программного обеспечения.

Сегодня можно выделить целый ряд проблем применения ИИ, имеющих этическую составляющую: автономность систем ИИ, эмоциональное взаимодействие с этими системами, обучение без понимания, нестабильность обучения, проблема спецификации, предвзятость в отборе данных.

Системы ИИ проявляют автономность, когда самостоятельно, без участия человека принимают решение о своем дальнейшем поведении. Однако остается не ясным, на каком основании и по каким причинам это происходит, где пролегает граница автономии машины, каковы последствия этой автономии, если она пересекается с автономией человека?

С автономностью систем ИИ тесно связана проблема прослеживаемости, интерпретируемости и лояльности (лояльность в мире автономных систем означает свойство, когда система делает то, что она декларирует, сообщает о своих действиях, ничего не утаивая в фоновых процессах). Действия системы для человека должны быть максимально прозрачными, легко интерпретируемы и однозначно и безусловно лояльными по отношению к человеку. В настоящее время на рынке есть немало алгоритмов (например, интерпретирующих полученные в рамках медицинских исследований снимки), и эти алгоритмы не предполагают необходимость объяснять свою работу и ее результаты. Так, в 2015 г. проводилось исследование, нацеленное на выявление тех пациентов из числа госпитализированных с диагнозом «пневмония», у которых в перспективе следует ожидать серьезные осложнения. Алгоритм ошибочно предсказал, что астматики лучше переносят пневмонию. Данный прогноз потенциально мог привести к необоснованным решениям об отказе в госпитализации со стороны медицинских учреждений и нанести значительный вред здоровью людей, страдающих астмой [7].

Еще один из вопросов, отражающих рассматриваемый аспект проблемы применения систем с ИИ, может быть сформулирован следующим образом: «На кого накладывать гражданскую ответственность, если случится дорожно-транспортное происшествие с участием автономного беспилотного транспорта?». Размышления над ответом неизбежно отсылают нас к этической категории справедливости, обсуждению проблемы равенства машин и людей. Будет ли обоснована выдача разрешения на использование автономного транспорта, по сути предоставляющая машинам равные с людьми права и прерогативы? Кроме того, автономные транспортные средства в процессе перемещения будут собирать большое количество данных об объектах, находящихся как снаружи, так и внутри салона. Следова-

[1] BMVI. Ethics Commission's complete report on automated and connected driving. URL: https://www.bmvi.de/goto?id=354980 URL: https://www.jibo.com (дата обращения: 18.01.2022). Текст: электронный.

[2] IEEE Global Initiative. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. URL: http://standards.ieee.org/develop/indconn/ec/autonomous-systems.html. (дата обращения: 18.01.2022). Текст: электронный.

тельно, могут возникнуть и проблемы нарушения конфиденциальности персональных данных пассажиров. Еще один вопрос: насколько будет проявлена автономность системы при современной версии «проблемы вагонетки»? Каким образом беспилотный транспорт будет выбирать наименьшее зло в случае возникновения чрезвычайной ситуации, когда независимо от решения все же будут жертвы, и ни в одном из сценариев не окажется положительного или правильного решения? Очевидно, что при разработке алгоритма управления автономным транспортом решение этой проблемы будет невероятно сложным.

Сообщество автомобильных инженеров (SAE) разработало пятиуровневую классификацию автоматизации автомобилей, в которой пятый уровень соответствует «полной автоматизации» автомобиля, то есть машина не только делает все сама и едет куда надо в любое время и в любых условиях, но и исключает участие человека в управлении [8, с. 54]. Одной из первых стран, которая выказывает желание и готовность внедрить автономный транспорт является Сингапур. В январе 2019 г. Сингапур утвердил национальные стандарты по автономному транспорту Technical Reference 68 (TR 68) и рамочную модель добровольного внедрения систем на базе ИИ (Model AI Governance Framework) [9]. Эмоциональное взаимодействие человека с системами ИИ очень наглядно показано в фильме «Из машины». Главный герой эмоционально вовлекается в отношения с гуманоидным роботом Ава и влюбляется в нее, несмотря на понимание того, что перед ним не живой человек, а машина. В диалоге с ним машина раскрывает свои, совершенно отличные от человеческих ценности.

Более сильное эмоциональное вовлечение демонстрирует пример с танатосенситивными чат-ботами. «Танатосенситивность — это новый гуманистически обоснованный подход к исследованиям и планированию информационных технологий, который распознает и активно использует факты смертности, умирания и смерти человека при создании интерактивных систем» [10, с. 2466]. Танатосенситивными чат-ботами являются программы, создающиеся на основе цифровых данных погибшего человека и имитирующие его личность [11, с. 1205]. Эти чат-боты содержат практически всю цифровую информацию о погибшем (данные его переписок, аудио- и видеоматериалы, фотографии), на основании которых и производится имитация этой личности. Так, например, в 2016 г. была создана программа «Лука» под руководством Е. Куйды и Ф. Дудчака, которая сканировала поведение умершего пользователя цифрового аккаунта и создавала на основе этих данных цифровую копию человека. Как показал опыт, данное приложение, основанное на системе ИИ, очень точно имитировало личность погибшего.

Попытка создания и внедрения этого приложения пока не удалась в полной мере из-за отсутствия финансирования. При этом стремление запустить подобные проекты не исчезло. Восприятие и переживание смерти составляют сущностную характеристику человека, придавая уникальность и осознанность его существованию. Однако перенос личности человека в цифровое пространство и возникновение танатосенситивной цифровой среды, на наш взгляд, как бы отчуждает человека, личность от смерти и, создавая принципиально новое, другое виртуальное содержание для переживания человеком собственной смерти и смерти близких, может привести к кризису его идентичности.

Обучение без понимания связано с проблемой задавания метрики системам ИИ. К примеру, трудно представить, каким образом возможно объяснить машине, что такое человек в трактовке, в понимании самого человека. Отсюда исходит трудность в общении с системами ИИ и непереводимость многих фундаментальных категорий на язык машины и обратно, с языка машины на человеческий язык. Так, в сфере медицины, где есть необходимость интерпретации медицинских снимков как для самого врача, так и для пациента, данные проблемы проявляются наиболее очевидным образом. В радиологических снимках, если система ИИ распознает некую область, как представляющую угрозу для жизнедеятельности человека, возникает проблема, как именно диагностировать эту «некую область». Распознавание не равнозначно пониманию: есть такая вещь, как нулевой контекст, о чем говорит Фэй-Фэй Ли в лекции TED по компьютерному зрению (2015) [12]. Например, интерпретация машиной изображения памятника всаднику: «человек едет верхом на коне по улице». Другой пример связан с распознаванием эмоций, когда ИИ пытается определить грусть. Чаще всего, когда человек грустит, его лицо не меняется довольно длительное время. Сетка замечает нулевую динамику в экспрессии и делает предположение, что это либо «нейтральное состояние», либо «грусть», и уже потом уточняет остальные признаки и делает вывод о наблюдаемой эмоции [13]. Это означает, что машинное распознавание образов пребывает сейчас на самых ранних стадиях обучения. Одного лишь распознавания образов недостаточно для интерпретаций процессов [14, с. 115].

Проблема нестабильности обучения связана с тем, что пока системы ИИ часто допускают ошибки в более точном распознавании образов. Например, если система ИИ в беспилотном транспорте без труда распознает объект на дороге в обычных условиях, в измененных условиях (добавление лишнего шума или иной информации, устранение которой не заложено в алгоритмах) она может уже не распознать его, что потенциально создает ситуацию, представляющую угрозу для безопасности людей.

Развитие технологий заставляет политиков и общество в целом задуматься о происходящих изменениях, обусловленных использованием систем ИИ. Речь идет о том, является ли все более широкое внедрение систем ИИ этически ответственным или даже необходимым? Положительный ответ на вопрос неизбежно ставит нас перед необходимостью отвечать на следующий, еще более сложный вопрос: насколько мы готовы стать технически зависимыми от сложных систем ИИ? Какие меры предосторожности необходимы для обеспечения управляемости, прозрачности и лояльности действий таких систем? Какие технические принципы развития необходимы для того, чтобы не размыть контуры гуманного общества, в котором личность, ее свобода развития, физическая и умственная неприкосновенность, ее притязания на общественное уважение будут находиться в центре правовой системы?

Помочь с ответами на эти вопросы может технологическая этика, обсуждающая этические основы жизни человека и человеческой цивилизации в современную эпоху внедрения ИИ и новых технологических разработок. Речь идет об этике ответственности Ханса Йонаса, которая, по нашему мнению, предлагает решение, заслуживающее внимания. Х. Йонас еще в 1979 г. опубликовал книгу «Принцип

ответственности. Опыт этики для технологической цивилизации», которая и сегодня остается интересной и актуальной. По словам Йонаса, технологическое развитие представляет все большую угрозу для будущего человечества. Автор утверждает, что людям необходим пересмотр положений и принципов традиционной этики и создание новой этики, способной осмыслить чрезвычайно возросшую технологическую мощь человечества, а также предвидеть возможные последствия ее применения. Йонас критикует традиционную этику за то, что она не адаптирована под новые условия в связи с изменившимся характером человеческих действий. «Ни одна предшествующая этика не научит нас нормам добра и зла, которые бы вместили совершенно новые модальности власти и ее возможные творения. Целина коллективного праксиса, на которую мы вступаем вместе с высокими технологиями, является для этической теории еще ничейной землей» [15, с. 189].

В этой ситуации на первый план должна выйти этика ответственности, в которой одноименная категория выступает в качестве центральной системообразующей категории и фокусируется главным образом на последствиях поступков, действий человека, а не на его «благих намерениях». М. Вебер, немецкий социолог, выделил два вида этики, а именно этику ответственности (Verantwortungsethik) и этику добрых намерений (Gesinnungsethik) [16]. Согласно этике добрых намерений моральная ценность поступка определяется намерением преступника. Действия определяются как хорошие, если они совершаются исходя из добрых побуждений. Хорошо известным примером этого является этика И. Канта. Согласно его этическим принципам, фактический эффект действий, совершаемых с добрыми намерениями, не имеет значения для морального суждения. Нравственная ценность выводится не из результатов действия, а из цели. Однако в реальном мире может быть множество значимых целей, достижение которых возможно только посредством нарушения моральных норм, поэтому этика добрых намерений становится принципиально не реализуема.

Согласно этике ответственности центр тяжести переносится на результат, на последствия наших действий. При этом делается упор на разумную природу человека, присущую ему способность к рациональному познанию. Йонас подчеркивает, что человек должен руководствоваться рациональными соображениями, чтобы сравнивать, оценивать возможные результаты, последствия каждого своего поступка и выбирать наиболее подходящую альтернативу для осуществления своих намерений. Только в этом случае человек, совершенные им деяния могут быть оправданы как перед самим собой, так и перед другими.

По мысли Йонаса, этика ответственности получает новое понимание в контексте новых технологий [15, с. 234]. Люди начинают осознавать, что применение последних необратимым образом изменяет практически все сферы жизнедеятельности человека, начиная с самой природы человека как на физиологическом уровне, так и на ментальном. Причем эти изменения далеко не всегда означают благо для человека и его будущего. Скорее, напротив, угрожают разрушением его привычному образу жизни, месту человека в мире, угрожают самому существованию человечества. Осознание «апокалиптической ситуации», считает Йонас, должно породить эвристический страх, люди должны представить долгосрочные последствия технологического развития цивилизации [17]. После того, как мы во-

образим то, что может произойти в будущем, у нас возникнут чувства, соответствующие тому, что мы вообразили. Именно эти чувства позволят в корне изменить наш образ жизни (например, породят заботу об окружающей среде). Так возникает осознание ответственности за предотвращение будущих бедствий, или то, что Йонас называет принципом ответственности: «Поступай так, чтобы последствия твоих действий были сообразны целям сохранения истинной человеческой жизни на земле» [18].

Философ предлагает и другие варианты вариантов формулирования принципа ответственности:

- «Поступай так, чтобы последствия твоего действия были совместимы с непрерывностью подлинной человеческой жизни на Земле»;

- «Включай в свой нынешний выбор будущую совокупность людей в качестве предмета твоего воления».

- «Поступай так, чтобы последствия твоих действий не были разрушительны для будущей возможности жизни как таковой».

- «Не подвергай опасности условия неограниченного дальнейшего существования человечества на Земле» [15, с. 36].

При этом важно, считает Йонас, чтобы предсказания негативных последствий имели для нас приоритет перед позитивными предсказаниями или надеждой на то, что технологии могут сделать людей более человечными. Ответственность не является нашей обязанностью, она основана на вызовах реальности, находящейся под угрозой.

Согласно традиционной этике, природа человека и природа вещей по сути своей неизменны и неуязвимы. С античного периода традиционная этика считала, что человек не может нарушить равновесие природы, и отсюда проистекает антропоцентрический характер этически значимой области, как считает Йонас. «Неприкосновенность природы… ее существенная неизменность как космического порядка фактически являлась фоном всех предприятий смертного человека, включая его вмешательства в сам этот порядок. Его жизнь разворачивалась между постоянным и переменным: постоянное было природой, переменным — его собственные произведения» [15, с. 20]. Посредством бурного развития техники человек освоил огромные природные ресурсы вплоть до их истощения, расширил область искусственной среды обитания, меняя не только окружающую среду, но и свою собственную природу. Теперь, согласно Йонасу, человек должен ограничивать свое чрезмерно потребительское отношение к природе, заботиться о ней, нести ответственность за ее будущее, за будущее своих потомков. Под угрозой находится не только окружающая среда, но фундаментальные характеристики самого человека, его физической природы. Искусственное поддержание жизни, успехи генной инженерии, клонирование, пересадка искусственных органов, прогнозирование синтеза белков, киберимпланты в мозге, позволяющие распознавать речь со скоростью мысли, бионические протезы — все эти новшества заставляют нас иначе взглянуть на «природу» человека и задаться вопросом, где стираются границы между physis и techne в человеческом бытии. Даже смертность как сущностная особенность человека уже не представляется чем-то неотъемлемым и неотвратимым благодаря последним открытиям в области биологии и медицины. Такое неограниченное господство человека над природой и своей сущностью требует

пересмотра традиционной этики и включения в сферу этического новых объектов моральной ответственности. Исходя из современных реалий этически значимая область не должна, как прежде, быть ограничена сферой общения между людьми. В новой этике ответственность человека за сохранность природы и существование человечества должны занять главное место, стать краеугольным камнем всех теоретических построений. Сам же человек мыслится как центральный пункт бытия, поскольку является единственным существом, которое способно нести ответственность за будущее всего мира, способно устанавливать надежный контроль над всеми (в том числе своими собственными) потенциальными возможностями его изменения.

Следующий тезис, который критикует Йонас, касается пространственной и временной ограниченности универсума традиционной этики. Традиционная этика не принимала во внимание отдаленные последствия человеческой деятельности и этические требования были ограничены обозримыми пространственными и временными рамками. «Нравственный универсум состоит из современников и его временной горизонт ограничен предполагаемыми отрезками их жизней. Схожим образом обстоит дело и с пространственным горизонтом, в котором действующий и другой встречаются как соседи, друзья или враги…» [15, с. 23]. Моральность поступков исходит из таких параметров, как их непосредственные результаты, соблюдение морального закона, соответствие идее блага, общественному благополучию и т. п.

При этом высшее благо, как правило, имеет трансцендентную природу, что проявляется в неизменности и вечности природы. Например, в философии Платона мир идей стоит в вертикальной направленности по отношению к конечному и преходящему миру вещей. Только мир идей обладает подлинной ценностью, соответственно, этическая составляющая существует в отношении не к миру вещей, а направлена далеко за его пределы, к миру истинного бытия. Согласно Йонасу, такая установка продолжает свое развитие и в философии Канта («категорический императив»), и в философии Гегеля («абсолютный дух»). Йонас утверждает необходимость пересмотра этой установки и замены ее этикой ответственности, которая заключается, с одной стороны, в доказательстве ценности и самодостаточности настоящего, а с другой — подразумевает то, что необходимо учитывать временное измерение, оценивать и прогнозировать отдаленные последствия для будущего всякого действия, совершенного в настоящем, причем любыми имеющимися способами, начиная с футурологии и заканчивая научной фантастикой. Временной аспект этики ответственности требует, чтобы действия человека ни при каких условиях и никогда не угрожали дальнейшему существованию мира и человека в нем.

По словам Йонаса, сама осознаваемая человеком опасность должна служить ориентиром построения этики ответственности: «В свете ее зарниц будущее в его планетарном масштабе и его человеческой глубине становится всесторонне открыто новым этическим принципам, из которых дедуцируются новые обязанности новой власти» [15, с. 7]. Йонас настаивает на том, чтобы человечество сместило акцент не на наши желания и надежды, а на то, что вызывает у нас страх и озабоченность. Он пишет об эвристической функции страха, которая оценивает последствия наших действий и предостерегает нас от недостаточно взвешенных решений.

Если в предшествующей этической традиции страх интерпретировался как негативный аффект, требующий преодоления и борьбы, то Йонас, следуя идеям философов-экзистенциалистов, придает страху иной смысл. Страх способствует устранению будущей угрозы человечеству, его самосохранению. У Хайдеггера Angst в виде экзистенциального страха Вот-бытия перед Ничто помогает человеку осознать свое бытие в качестве Бытия-в-мире, осознать собственную свободу и конечность. Посредством страха раскрывается предельная, безотносительная и вместе с тем неопределенная возможность вот-бытия, конституирующая его целостность и полноту. Этой предельной возможностью, по Хайдеггеру, является смерть [19, с. 164]. «Я сам есть эта постоянная, предельная возможность меня самого, а именно — возможность более не быть» [20, с. 201]. Если у Хайдеггера вот-бытие переживается единичным и уникальным индивидом, то у Йонаса вот-бытие представлено совокупной субъект-коллективной деятельностью. А. Н. Ермолаенко пишет, что Йонас вводит понятие Furcht как страх перед чем-то конкретным, страх, который имеет отношение к миру временному, конечному. Например, страх перед ядерной катастрофой, страх за жизнь Другого» [21, с. 144].

Стоит заметить, что этика ответственности стала неким логическим продолжением «философии организма» Г. Йонаса [22, с. 127]. По мнению философа, организм — это онтологический центр или «узел бытия». Органическое через разные формы свободы уже в своих низших проявлениях подготавливает дух, остающийся в своих высших формах частью органического. Эта идея утверждается Г. Йонасом в прологе книги «Организм и свобода». В эпилоге произведения он заключает, что философия духа неизбежно включает в себя этику, которая также становится частью философии природы благодаря неразрывному единству духа и организма, организма и природы [23, с. 330]. Таким образом, йонасовская идея жизни подразумевает развитие цели, имманентной органическому, в свободный и способный к ответственности человеческий дух, а организм толкует как уникальное явление в поэтапном ряду жизни. В процессе восхождении к человеку в концепции Йонаса неизбежно присутствует телеологический компонент, нашедший наиболее полное выражение в книге «Принцип ответственности», в которой автор пишет: «Создавая жизнь, природа делает явной, по крайней мере, одну определенную цель, а именно саму жизнь» [15, с. 145].

Фундаментальный и категорический принцип теоретических построений Йонаса может быть выражен так: «человечество должно существовать». Он пишет, что только данный принцип можно считать поистине категорическим и безусловным, потому что все иные отпадают, если самому человеку угрожает опасность исчезновения. Телеология Йонаса здесь перетекает в аксиологическую онтологию, где бытие имеет аксиологическое преимущество перед не-бытием из-за опасности превращения его в не-бытие. Этическая способность к ответственности в системе Йонаса становится онтологической, способность человека осознавать ответственность за будущее мира и человечества — центральным условием бытия.

Аксиологическая онтология, полагает Г. Йонас, должна спасти такие феномены, как жизнь, органическое, субъективность, свобода и т. п., от редукционизма, дав им именно ценностную интерпретацию.

Таким образом, открывающиеся современные аспекты взаимодействия ИИ с человеком приводят к возникновению сложных этических вопросов, ответы на которые необходимы для предотвращения значительных трудностей и проблем в будущем существовании человеческой цивилизации. Технологическое изменение мира приобретает все более масштабный, кумулятивный и необратимый характер. Сложившаяся беспрецедентная ситуация с применением систем ИИ практически во всех жизненных процессах человека требует пересмотра традиционных этических категорий, осознания нового масштаба ответственности человека перед лицом будущих поколений. В силу своей разумной природы он способен предвидеть, уменьшить, устранить негативные эффекты, вызванные применением систем ИИ, обеспечив тем самым безопасное будущее всего человечества.

Этика ответственности Йонаса, включающая в себя понятие страха в качестве важнейшего принципа осуществления заботы о будущем, акцентирует свое внимание именно на этой важнейшей для человека способности. Страх за будущее, страх перед возможными изменениями природы и человека сегодня может быть осмыслен в качестве главного ценностно-образующего принципа, необходимого элемента ответственности и источника долженствования в деятельности и поступках человека, в качестве регулятивного принципа новой этики, порожденной эпохой технократизма и ИИ.

Литература

1. Gronau I. Autonomous tractor cooperation lays out plan to slowly introduce driverless tractor // Precision. 2016. February 17. URL: https://www.precisionfarmingdealer.com/articles/2013-autonomous-tractor-cooperation-lays-out-plan-to-slowly-introduce-driverlesstractor (date of the application: 12.01.2022).

2. Зыков В. Беспилотный трактор прошел тесты на полях в России // Известия IZ. 2016. 10 июня. URL: https://iz.ru/news/617516 (date of the application: 14.01.2022).

3. Rudin C., Sloan M. Predictive policing: using machine learning to detect patterns of crime // Wired. URL: https://www.wired.com/insights/2013/08/predictive-policing-using-machine-learning-to-detect-patterns-of-crime/ (date of the application: 12.01.2022).

4. Snow J. Google's New AI Smile Detector Shows How Embracing Race and Gender Can Reduce Bias. MIT Tecnology Review, 2017. P. 189.

5. Steinberg R. Areas Where Artificial Neural Networks Outperforms Humans. Venture Beat, 2017. P. 17.

6. Baxter и Sawyer. URL: https://robogeek.ru/novosti-kompanii/sozdateli-robotov-baxter-i-sawyer-pereshli-na-rabotu-v-universal-robots# (date of the application: 18.01.2022).

7. Caruana R. et al. Intelligible Models for Health Care: Predicting Pneumonia Risk and Hospital 30-day Readmission. ACM. 2015. P. 23.

8. Shladover S. The Truth About 'Self-Driving' Cars. Scientific American, 2016. P. 53–57.

9. Irene Tham. World Economic Forum: Singapore updates AI governance model with real-world cases // Strait Times. 2020. 22 January. P. 34.

10. Massimi M., Charise A. Dying, death, and mortality: Towards thanatosensitivity in HCI // CHI'09 Extended Abstracts on Human Factors in Computing Systems. ACM. 2009. P. 2459–2468.

11. Шаповалов И. С. Разрушение идентичности и отчуждение от смерти в танатосенситивной цифровой среде // Манускрипт. 2021. Т. 14, вып. 6. С. 1202–1208. Текст: непосредственный.

12. Фэй Фэй Ли. О компьютерном зрении (лекция). URL: https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures?language=ru (дата обращения: 21.12.2021). Текст: электронный.

13. Малышев А. Компьютерное зрение видит эмоции, пульс, дыхание и ложь — но как построить на этом стартап. Разговор с Neurodata Lab. 2019. 27 авг. URL: https://habr.com/ru/company/habr_career/blog/465153/ (дата обращения: 20.12.2021). Текст: электронный.

14. Тополь Э. Искусственный интеллект в медицине. Как умные технологии меняют подход к лечению / перевод с английского. Москва: Альпина Паблишер, 2022. 696 с. Текст: непосредственный.

15. Йонас Г. Принцип ответственности. Москва: Айрис-Пресс, 2004. 480 с. Текст: непосредственный.

16. Jonas H. Technology and responsibility: Reflections on the new tasks of ethics // Ethics and Emerging Technologies. Boston, 2014. P. 37–47. URL: https://doi.org/10.1057/9781137349088_3 (date of the application: 12.01.2022).

17. Jonas H. Technology as a subject for ethics. Social Research. 49(4). P. 891–898. URL: https://www.jstor.org/stable/40971222 (date of the application: 18.01.2022).

18. Gordon J. S., Burckhart H. Global ethics and moral responsibility. Hans Jonas and his critics. Routledge, 2020. 238 p.

19. Гаджикурбанова П. А. Страх и ответственность: этика технологической цивилизации Ганса Йонаса. Этическая мысль/Ethical Thought. 2019. № 4. P. 161–178. URL: https://et.iphras.ru/article/view/2506 (дата обращения: 12.01.2022). Текст: электронный.

20. Хайдеггер М. Пролегомены к истории понятия времени. Томск, 1998. 383 с. Текст: непосредственный.

21. Ермолаенко А. Н. Этика ответственности и социальное бытие человека. Киев, 1994. 198 с. Текст: непосредственный.

22. Пугачева Н. П. Два «Принципа» Ганса Йонаса: апология жизни // Философия и общество. 2020. № 2. С. 123–130. Текст: непосредственный.

23. Jonas H. Das Prinzip Leben. Anzaetzezueiner philosophischen Biologie. Frankfurt a. Mein: Insel, 2011. 401 p.

Статья поступила в редакцию 01.02.2022; одобрена после рецензирования 18.02.2022; принята к публикации 21.02.2022.

ETHICS OF ARTIFICIAL INTELLIGENCE: HANS JONAS' PRINCIPLE OF RESPONSIBILITY

Maina Kh. Badmaeva
Research Assistant,
Dorzhi Banzarov Buryat State University
24a Smolina St., Ulan-Ude 670000, Russia
badmaevamaina@gmail.com

Abstract. Artificial intelligence (AI) has now become an integral part of the life of modern society. The prospects for its application require deep reflection, since they are fraught with adverse consequences for mankind. The article discusses the most relevant aspects of the use of artificial intelligence, describes the possible difficulties and benefits of the introduction of these technologies. In our opinion, understanding the reasons, meaning and consequences of such an introduction requires the development of new theoretical methods and

provisions adequate to the subject of study. We believe that taking into account and observing ethical principles in the process of developing artificial intelligence technologies and its application will contribute to the most harmonious development of mankind. In this regard, it is necessary to revise traditional ethics and search for the grounds, which will ensure fruitful interaction between a person and AI systems in the future. German existentialist philosopher Hans Jonas made an attempt to create such an ethical system of views. In the article, we study Hans Jonas' ethics of responsibility as a new regulatory component in applying artificial intelligence technologies. Ethics of responsibility can reduce existential risks, while maintaining the possibility of further introduction of AI into human life.
Keywords: technocracy, artificial intelligence, narrow artificial intelligence, ethics, ethics of artificial intelligence, ethics of responsibility, Hans Jonas, existentialism, fear of the future, axiological ontology.

For citation
Badmaeva M. Kh. Ethics of Artificial Intelligence: Hans Jonas' Principle of Responsibility. *Bulletin of Buryat State University. Philosophy.* 2022; 1: 67–79 (In Russ.).

The article was submitted 01.02.2022; approved after reviewing 18.02.2022; accepted for publication 21.02.2022.