

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И РЕКОНСТРУКЦИЯ ЧЕЛОВЕЧЕСКОЙ СОВЕСТИ



<https://doi.org/10.24412/2181-1784-2025-24-187-196>

Холмуминов Джафар Мухаммадиевич,
доктор философских наук (DSc),
профессор Института ISFT
(г. Ташкент, Республика Узбекистан).
jafarkholmomi@gmail.com

Аннотация: Настоящая статья посвящена философскому анализу возможностей и ограничений реконструкции человеческой совести с помощью систем искусственного интеллекта (ИИ). Автор, используя онтологический, гносеологический и этический подходы, показывает, что совесть как проявление духовного познания, нравственного самосознания и свободной воли обладает такими характеристиками, которые делают её недоступной для полной реконструкции посредством алгоритмов ИИ. В исследовании рассматривается эволюция понимания совести в философской традиции от Сократа до Хайдеггера и Йонаса, подчеркивается, что живой духовный опыт, индивидуальное самосознание и моральная ответственность не могут быть воссозданы искусственными моделями. Путём анализа возможностей искусственного интеллекта и связанных с ним рисков статья акцентирует необходимость формирования ориентированной на будущее этики и принципа предостерегающей ответственности в эпоху технологического развития. Исследование освещает глубокие философские вопросы взаимоотношения человека и технологий и предлагает философский ответ на задачу сохранения ценностей свободы, ответственности и достоинства человека в цифровом пространстве.

Ключевые слова: искусственный интеллект (ИИ), совесть и самосознание, моральная ответственность, духовный опыт, этика ориентированная на будущее, философия бытия и морали, ханс йонас и принцип предостерегающей ответственности, ограничения моделирования нравственности, искусственные технологии и человеческие ценности, онтология человека и технологии.

ARTIFICIAL INTELLIGENCE AND RECONSTRUCTION OF HUMAN CONSCIENCE

Kholmuminov Jafar Muhammadiyevich

Doctor of Philosophical sciences (DSc),

Professor of the “INTERNATIONAL SCHOOL OF FINANCE
TECHNOLOGY AND SCIENCE” INSTITUTE

(Tashkent, Uzbekistan)

jafarkholmomi@gmail.com

Abstract: *This article is dedicated to the philosophical analysis of the possibilities and limitations of reconstructing human conscience through artificial intelligence (AI) systems. The author, employing ontological, epistemological, and ethical approaches, demonstrates that conscience, as an embodiment of spiritual cognition, moral self-awareness, and free will, possesses characteristics that place it beyond full reconstruction through artificial algorithms. The study examines the evolution of the notion of conscience within the philosophical tradition – from Socrates to Heidegger and Jonas – and emphasizes that living spiritual experience, individual self-awareness, and moral responsibility cannot be fully replicated by artificial models. Through an analysis of the capabilities and potential risks of AI, the article stresses the necessity of forming a future-oriented ethics and the principle of precautionary responsibility in the era of technological development. This research highlights profound philosophical issues regarding the relationship between human beings and technology and offers a philosophical response to the challenge of preserving the human values of freedom, responsibility, and dignity within the digital sphere.*

Keywords: *Artificial Intelligence (AI), Conscience and Self-Awareness, Moral Responsibility, Spiritual Experience, Future-Oriented Ethics, Philosophy of Being and Morality, Hans Jonas and Precautionary Responsibility, Limitations of Moral Modeling, Artificial Technology and Human Values, Ontology of the Human Being and Technology.*

В современном сознании, находящемся под влиянием беспрецедентного развития цифровых технологий и автоматизации, вопрос возможности воссоздания человеческой совести с помощью систем искусственного интеллекта (Artificial Intelligence, AI) становится одной из важнейших тем этической философии и онтологии человека.

Совость, которая в философской традиции понимается как высшее проявление нравственного самосознания и внутренней силы, оценивающей человеческие поступки, с времён Сократа, Платона и Аристотеля до Иммануила Канта и Эдмунда Гуссерля была центральной темой философских размышлений в области онтологии и этики. С философской точки зрения, совесть – это не просто вычислительный механизм или результат позднего этапа эволюции познания, а воплощение духовного проявления и способности к саморегуляции ценностей.

С появлением систем обучения нейронных сетей и самообучающихся алгоритмов сегодня в научно-философской среде обсуждается теоретическая возможность создания искусственных агентов, которые могли бы давать этическую оценку своим действиям.

Однако ключевой вопрос, стоящий в центре этих дискуссий, таков: возможно ли воссоздание совести с помощью искусственного вычислительного инструмента или же совесть, как сугубо человеческий, экзистенциальный и духовный феномен, остаётся вне пределов искусственного моделирования?

Таким образом, цель данного исследования – философский анализ возможности и ограничений воссоздания совести с помощью искусственного интеллекта. Это исследование проводится с комплексного онтологического, гносеологического и этического подходов, затрагивая не только вопрос технической возможности, но и вопросы человеческого познания, самосознания и смысла бытия.

В условиях современного мира, где способности искусственного интеллекта могут изменять социальные и этические структуры, а также даже наши представления о том, что значит быть человеком, возникает необходимость философского осмысления. Поэтому предлагаемое исследование направлено на попытку выработать этические и онтологические границы и критерии взаимодействия человека с искусственной технологией.

Совесть издавна рассматривалась как одно из важнейших проявлений человеческого самосознания.

В античной философии, особенно в учениях Сократа и Платона, совесть описывалась как «внутренний голос» или «говорящий истину», который направляет человека к добру и истине [1].

В философии Иммануила Канта совесть выступает как внутренний суд, побуждающий человека к оценке своих поступков с нравственной точки зрения [2].

По Канту, совесть – это не просто внутреннее чувство, а нормативное и универсальное предписание, обязывающее человека следовать моральному закону внутри себя.

С точки зрения феноменологии Эдмунда Гуссерля, совесть – это опыт связи с собственным бытием, в котором самосознание и этичность сливаются [3].

Гуссерль акцентирует внимание на способе непосредственного переживания совести в познании себя и мира, характеризуя её как динамику нравственного сознания.

Таким образом, в философской традиции совесть обладает следующими чертами:

Характер интериоризации (замещение внешних источников морали внутренним источником);

Результат самосознания и самокритики;

Универсальная ориентация на ценности и внутреннюю ответственность.

С онтологической точки зрения совесть рассматривается как сугубо человеческий духовный феномен, который выходит за пределы интеллектуальной сферы и достигает уровня подлинного опыта бытия. Совесть является внутренним вопросом «кем быть» и «что делать», что соответствует фундаментальной проблеме практической философии бытия.

В этом контексте Мартин Хайдеггер трактует совесть не только как этическое явление, но и как онтологическое: она выступает как «голос бытия», пробуждающий человека от забвения подлинного существования [4]. С этой точки зрения, совесть не носит исключительно рационального характера и не может быть выражена лишь через вычисления и программирование.

Напротив, она представляет собой живой духовный опыт связи человека с собственной сутью и сутью бытия, который лежит за пределами возможностей вычислительных систем и искусственных алгоритмов.

Совесть обладает и гносеологическим характером, поскольку она не только чувствует, но и знает – и на основе этого знания действует. Это знание имеет неполный, но живой и подлинный характер. Согласно этическим теориям Джона Ролза и Саймона Блэкберна, совесть представляет собой своего рода практический разум, в котором практическая оценка неразрывно связана с нравственным знанием и опытом [5][6].

Эта особенность показывает, что основное различие между искусственной вычислительной системой и человеческой совестью проявляется в форме и источнике знания:

Система искусственного интеллекта получает знания на основе данных и алгоритмов;

Совость же постигает знание через духовный опыт и экзистенциальное отношение.

Таким образом, вопрос о воссоздании совести с помощью AI – это не просто технологический вопрос, а вопрос возможности воссоздания духовного опыта и связи с бытием. Одним из главных ограничений воссоздания совести средствами искусственного интеллекта является отсутствие субъективности и духовного опыта. С онтологической точки зрения совесть рождается из живого духовного опыта, формирующегося в непосредственном общении с «Я» – сущностью и в постижении существования других [7].

Хайдеггер в «Бытие и время» указывает, что совесть – это «голос бытия», а не логическая функция [8]. Это означает, что совесть – это воплощение связи человека с сутью существования, которое не может быть воспроизведено вычислительными алгоритмами или искусственными моделями воли.

Система AI, даже при самых сложных моделях обучения, не способна создать духовный опыт присутствия и духовной связи.

Совость как источник знания и нравственной оценки отличается от знаний, получаемых искусственным интеллектом.

Знания ИИ строятся на основе данных и алгоритмов, тогда как этическое знание, лежащее в основе совести, формируется из духовных, нравственных и моральных переживаний, которые имеют личностный и непрерывный характер [9].

Фридрих Ницше подчеркивает, что всякое знание связано с ценностями и внутренним процессом преобразования [10]. Это невозможно свести к статистическим данным или логическим уравнениям. ИИ, опирающийся на вычислительную логику, лишен такого живого и опытного знания, а потому не способен к подлинному воссозданию совести.

Алгоритмы ИИ зависят от обучения на исторических данных и накопленном опыте. Однако эти данные могут быть предвзятыми, и в результате этические решения искусственного интеллекта могут воспроизводить неэтичные или дискриминационные структуры [11].

Кроме того, нравственная ответственность в человеческих этических решениях устанавливается через самосознание и понимание последствий своих поступков.

Система ИИ, напротив, в подлинном смысле не может нести моральную ответственность, поскольку отсутствие самосознания и самоуважения не позволяет ей стать настоящей совестью.

Одна из угроз «воссоздания» совести с помощью AI – это опасность превращения этики в бюрократическую систему. Если этические решения будут полностью переданы алгоритмам, этика превратится из живого духовного феномена в механический порядок, в котором теряются свобода, выбор и личная ответственность – суть совести [12].

Ханс Йонас в книге «Императив ответственности» подчеркивает, что в технологическую эпоху человек должен сохранять моральную ответственность, а не перекладывать её на автоматическую систему [13].

С ростом систем, способных к этической оценке и принятию решений, появляются новые философские вопросы этики. С этической и философской точки зрения человечества, ощущается необходимость разработки искусственных этических принципов (Artificial Moral Principles, AMP) как основы формирования новой технологической ответственности. Необходимо помнить, что такие принципы не могут опираться только на критерии эффективности – они должны исходить из общих источников морали и духовных ценностей. Без опоры на практическую философию человека и духовную ответственность любая «искусственная этика» рискует превратиться в систему бездушных решений [14].

В философии практики (Аристотель, Кант, Йонас) этическая деятельность определяется как акт свободной и сознательной воли. Этика – это результат осознанной и свободной воли, с помощью которой человек осознает свою ответственность перед другими и перед бытием.

Воссоздание такого акта в рамках системы искусственного интеллекта сталкивается с глубокими онтологическими препятствиями:

Автоматизация этики сводит решения к инженерной алгоритмической схеме;

Отсутствие духовного опыта приводит к утрате чувства подлинной ответственности.

Поэтому философия искусственной этики должна признать ограничения автоматизации этики и сохранить решающую роль самосознания и воли как основы подлинной человеческой ответственности [15].

Учитывая расширение возможностей систем искусственного интеллекта, Ханс Йонас в своей теории «Императив ответственности» [16] предлагает

новый принцип осторожной ответственности в отношении будущего человека и природы как руководящий для наших действий.

Этот императив требует:

Предвидения и анализа практических и этических последствий применения новых технологий;

Определения этических ограничений использования AI в жизненно важных решениях;

Сохранения человечества как высшей цели и смысла каждого технологического развития.

Этика, устремленная в будущее, учит нас действовать не только из сиюминутной выгоды, но и с точки зрения глобальной и исторической ответственности.

Чтобы избежать угрозы необратимого разрушения этики и человеческого бытия через искусственные технологии, предлагаются философские модели управления развитием этических систем AI:

Требование гуманности: каждое решение и этическая оценка системы AI должны основываться на правах человека и его достоинстве.

Мультидисциплинарное сотрудничество: философия этики, биоэтика, право и научные эксперты должны совместно участвовать в разработке этических стандартов для AI.

Этический мониторинг: необходимость постоянного мониторинга и аудита этических алгоритмов для предотвращения предвзятости и недопустимых последствий.

Разработка глобального этического кодекса: создание международной этической базы для AI, обеспечивающей согласование универсальных ценностей [17].

Философия «искусственной совести» в эпоху стремительного роста возможностей AI призывает нас к осторожности и этическому балансу. Без должной этической мудрости и ответственного самосознания развитие искусственной «совести» может поставить под угрозу основные ценности человечества – свободу, ответственность и достоинство.

Поэтому каждое действие, связанное с развитием искусственного интеллекта, должно подвергаться философской и этической оценке, а зрелая человеческая ответственность должна оставаться в центре наших технологических решений.

Таким образом, данное исследование показывает, что вопрос «воссоздания» человеческой совести с помощью систем искусственного

интеллекта (AI) превращается в онтологическую, гносеологическую и этическую проблему.

Совость, как воплощение духовного познания и осознанного бытия человека, выходит за пределы уровня информации и вычислений и связана с такими явлениями, которые могут возникнуть только в контексте живого духовного опыта.

На основе философского анализа ограничения «воссоздания» совести с помощью AI выделяются на трех уровнях:

На онтологическом уровне – отсутствие субъективности и духовного опыта;

На гносеологическом уровне – глубокое различие между знанием и этическим познанием;

На этическом уровне – отсутствие самосознания и моральной ответственности в системе AI.

Эти ограничения подчеркивают, что вычислительная система может создавать модели практической этики, но не способна воспроизвести тайну духовной совести и свободы воли.

Опыт совести как нравственного познания и ответственной воли принципиально отличается от любой алгоритмической автоматизации.

Поэтому политика взаимодействия с искусственным интеллектом должна основываться на принципе осторожной ответственности (по Хансу Йонасу).

Это означает, что в процессе развития этических технологий человек должен твёрдо хранить моральные стандарты и общечеловеческие духовные ценности и использовать искусственный интеллект не как замену человеческой ответственности, а как вспомогательный инструмент, находящийся под этическим контролем.

Современная философская мудрость, способная к онтологическому, этическому и познавательному анализу, играет центральную роль в обеспечении надлежащего баланса между технологическим прогрессом и защитой человеческих ценностей.

Будущее искусственного интеллекта должно оставаться человеческим будущим – будущим, в котором этика, свобода и достоинство человека будут движущей силой любого технологического развития.

СПИСОК ЛИТЕРАТУРЫ (REFERENCES)

- [1] Платон. Апология. Перевод Бенджамина Джоветта. Оксфорд: Oxford University Press, 1891, С. 58. (Plato. Apology. Translated by Benjamin Jowett. Oxford University Press, 1891, P. 58.)
- [2] Иммануил Кант. Критика практического разума. Кембридж: Cambridge University Press, 1997, С. 175. (Kant, Immanuel. Critique of Practical Reason. Cambridge University Press, 1997, P. 175.)
- [3] Эдмунд Гуссерль. Идеи к чистой феноменологии и феноменологической философии. Springer, 1982, С. 234. (Husserl, Edmund. Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. Springer, 1982, P. 234.)
- [4] Мартин Хайдеггер. Бытие и время. Harper & Row, 1962, С. 312. (Heidegger, Martin. Being and Time. Harper & Row, 1962, P. 312.)
- [5] Джон Ролз. Теория справедливости. Harvard University Press, 1971, С. 165. (Rawls, John. A Theory of Justice. Harvard University Press, 1971, P. 165.)
- [6] Саймон Блэкберн. Этика: очень краткое введение. Oxford University Press, 2001, С. 79. (Blackburn, Simon. Ethics: A Very Short Introduction. Oxford University Press, 2001, P. 79.)
- [7] Эдмонд Авад и др. Эксперимент «Моральная машина». Nature, т. 563, 2018, С. 64. (Awad, Edmond et al. "The Moral Machine experiment." Nature, vol. 563, 2018, P. 64.)
- [8] Хьюберт Дрейфус. Почему компьютеры до сих пор не могут думать. Springer, 1992, С. 243. (Dreyfus, Hubert. What Computers Still Can't Do: A Critique of Artificial Reason. MIT Press, 1992, P. 243.)
- [9] Фридрих Ницше. Генеалогия морали. Oxford University Press, 1996, С. 120. (Nietzsche, Friedrich. The Genealogy of Morals. Oxford University Press, 1996, P. 120.)
- [10] Аласдэр Макинтайр. После добродетели. University of Notre Dame Press, 1981, С. 88. (MacIntyre, Alasdair. After Virtue. University of Notre Dame Press, 1981, P. 88.)
- [11] Поль Рикёр. Себя как другого. University of Chicago Press, 1992, С. 165. (Ricoeur, Paul. Oneself as Another. University of Chicago Press, 1992, P. 165.)
- [12] Сафия Умоджа Нобл. Алгоритмы угнетения. NYU Press, 2018, С. 140. (Noble, Safiya Umoja. Algorithms of Oppression. NYU Press, 2018, P. 140.)
- [13] Кэти О'Нил. Оружие математического разрушения. Crown Publishing, 2016, С. 108. (O'Neil, Cathy. Weapons of Math Destruction. Crown Publishing, 2016, P. 108.)

[14] Шошана Зубофф. Эпоха капитализма наблюдения. Public Affairs, 2019, С. 298. (Zuboff, Shoshana. The Age of Surveillance Capitalism. Public Affairs, 2019, P. 298.)

[15] Ханс Йонас. Императив ответственности. University of Chicago Press, 1984, С. 82. (Jonas, Hans. The Imperative of Responsibility: In Search of an Ethics for the Technological Age. University of Chicago Press, 1984, P. 82.)

[16] Джеймс Мур. Природа, значение и трудности этики машин. IEEE Intelligent Systems, 2006, С. 21. (Moor, James H. «The Nature, Importance, and Difficulty of Machine Ethics». IEEE Intelligent Systems, 2006, P. 21.)

[17] Шеннон Валлор. Технологии и добродетели. Oxford University Press, 2016, С. 92. (Vallor, Shannon. Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. Oxford University Press, 2016, P. 92.)

[18] Пола Боддингтон. К этическому кодексу для искусственного интеллекта. Springer, 2017, С. 160. (Boddington, Paula. Towards a Code of Ethics for Artificial Intelligence. Springer, 2017, P. 160.)