УДК: 101.1:004.8
DOI: 10.15372/PS20260107
EDN: ZOSLGH

В.В. Плотников, М.С. Косников

ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ГРАНИЦЫ ОТВЕТСТВЕННОСТИ ЧЕЛОВЕКА И АЛГОРИТМА

Статья посвящена анализу распределения ответственности между человеком и искусственным интеллектом по мере увеличения автономности цифровых систем. Цель исследования – определить, может ли алгоритм выступать носителем моральной ответственности и каким образом изменяется субъектность человека при делегировании функций принятия решений искусственному интеллекту. В ходе работы использован междисциплинарный подход, основанный на анализе философских концепций ответственности (деонтология, утилитаризм, экзистенциализм), а также современных направлений философии техники и цифровой этики. Авторы показывают, что несмотря на функциональную автономность и способность алгоритмов вырабатывать решения без участия человека, искусственный интеллект не обладает сознанием, намерением и способностью осознавать последствия своих действий, а значит, не может выступать моральным агентом; ответственность неизбежно остается у человека – разработчика, пользователя или владельца системы. Особенность исследования заключается в разработке философской модели распределения ответственности между участниками взаимодействия с искусственным интеллектом, которая демонстрирует, что главным критерием является не сложность алгоритма, а сохранение человеческого контроля. Значение работы состоит в обосновании необходимости цифровой этики и сохранения человеческой субъектности в условиях роста автономности искусственного интеллекта.

Ключевые слова: искусственный интеллект; моральная ответственность; субъектность; автономность алгоритмов; цифровая этика; философия техники; взаимодействие человека и машины

V.V. Plotnikov, M.S. Kosnikov

PHILOSOPHY OF ARTIFICIAL INTELLIGENCE AND THE BOUNDARIES OF HUMAN AND ALGORITHM RESPONSIBILITY

The article analyzes the distribution of responsibility between humans and artificial intelligence as the autonomy of digital systems increases. The purpose of the study is to determine whether an algorithm can act as a carrier of moral responsibility and how human subjectivity changes when decision-making functions are delegated to artificial intelligence. The study uses an interdisciplinary approach based on the analysis of philosophical concepts of responsibility (deontology, utilitarianism, existentialism), as well as modern trends in the philosophy of technology and digital ethics. The authors show that, despite the functional autonomy and ability of algorithms to make decisions without human intervention, artificial intelligence lacks consciousness, intention, and the ability to realize the consequences of its actions, which means it cannot act as a moral agent. Responsibility inevitably remains with the human developer, user, or owner of the system. The study's peculiarity is the development of a philosophical model of the distribution of responsibility among participants in interactions with artificial intelligence which demonstrates that the primary criterion is not the complexity of the algorithm, but the preservation of human control. The importance of the work is in substantiating the need for digital ethics and the preservation of human subjectivity in the context of the growing autonomy of artificial intelligence.

Keywords: artificial intelligence; moral responsibility; subjectivity; autonomy of algorithms; digital ethics; philosophy of technology; human–machine interaction

Современная цивилизация живет в условиях беспрецедентного технологического ускорения. Информационные технологии перестали быть вспомогательным инструментом: они формируют новую среду человеческого существования. Искусственный интеллект (ИИ), который еще недавно воспринимался как набор вычислительных алгоритмов, сегодня способен анализировать данные, прогнозировать события, принимать решения и взаимодействовать с человеком на уровне сложных коммуникативных моделей. Развитие алгоритмов машинного обучения, появление больших массивов данных и совер-

шенствование вычислительной инфраструктуры привели к тому, что ИИ охватывает практически все сферы деятельности – от медицины и банковских сервисов до военной аналитики, правоохранительных структур и платформ электронного управления [11; 14].

Расширение функциональной автономии ИИ меняет логику принятия решений в социально значимых областях. Алгоритмы участвуют в оценке кредитных рисков, помогают определять приоритетность юридических дел, адаптируют информационные потоки в социальных сетях, отслеживают перемещение людей в городах, управляют транспортными системами. Подобные процессы усиливают зависимость общества от машинных моделей рассуждения. В результате возникает новая философская проблема: кто отвечает за последствия действий автономного программного агента? Если алгоритм принимает решение, повлиявшее на жизнь человека, ответственность несет программист, владелец системы, пользователь или сама система как потенциальный субъект нового типа?

Дискуссия вокруг статуса ИИ имеет давние корни. В научной литературе выделяют два направления. Первое основано на понимании ИИ как инструмента, не обладающего собственным намерением. Сторонники этой позиции утверждают, что любые действия алгоритма являются следствием человеческого выбора, так как человек создает систему, задает модель данных и формулирует критерии [3; 4; 10; 15]. Приверженцы второй точки зрения исходят из того, что современные алгоритмы способны вырабатывать решения без прямого участия человека. Возникает образ «интеллектуального агента», который действует не по заранее расписанным сценариям, а в зависимости от возникающей ситуации. При этом авторы, представляющие данное направление, нередко говорят о возможности появления «сильного ИИ», обладающего элементами самосознания и самостоятельного целеполагания [1; 2; 13]. Подобная перспектива вызывает тревогу, сопоставимую с изменением антропологической модели: человек перестает быть единственным носителем рациональности и единственным источником ответственности.

Таким образом, развитие технологий выводит проблему с уровня технической дискуссии в область онтологии и философии. Если алгоритм вступает в сферу принятия этически значимых решений, то неизбежно возникает вопрос о распределении ответственности за

последствия этих решений. Человек может начать воспринимать ИИ как объективного и беспристрастного арбитра и постепенно утрачивать личное чувство ответственности. В таком случае риск заключается не только в возможной ошибке алгоритма, но и в изменении самой структуры человеческого поведения, в смещении мотивации и ослаблении моральных оснований поступков.

Цель настоящего исследования – проанализировать основы распределения ответственности между человеком и алгоритмом при повышении автономности систем искусственного интеллекта.

Осмысление заявленной проблемы требует обращения к философскому наследию, в котором сформулированы представления об ответственности, субъектности, свободе и намерении. В данной работе авторы опираются на идеи этических школ и направлений, а также на современные исследования в области философии техники и цифровой этики, в которых рассматриваются вопросы машинного принятия решений, автономных систем и границ человеческого контроля.

* * *

Проблема распределения ответственности между человеком и искусственным интеллектом относится к числу фундаментальных вопросов современной философии техники. Ускорение цифровых процессов и появление автономных вычислительных систем требуют анализа границ моральной субъектности и возможностей делегирования ответственности нечеловеческому агенту. Для понимания этого феномена важно обратиться к основным направлениям философской мысли, которые рассматривают природу ответственности, свободы и намерения.

1. Деонтологическая традиция (ответственность как следование долгу). В деонтологической этике, ярко выраженной в философии И. Канта, ответственность определяется не последствиями действия, а внутренним мотивом субъекта [3; 13]. *Человек стремится действовать в соответствии с нравственным законом, а не ради достижения результата*. Это не только этическая позиция, но и онтологическое основание субъектности. Моральный субъект – это тот, кто обладает способностью к свободному выбору и осознанию долга.

В рамках данной концепции ИИ не может быть носителем моральной ответственности, поскольку действия алгоритма лишены намерения. Алгоритм действует в соответствии с заданной логикой или статистической моделью, но не из понимания долга. Уточняя кантовскую мысль применительно к технологиям, Х. Йонас предлагает этику ответственности, ориентированную на защиту будущего человечества [5]. Его позиция акцентирует необходимость предвидения последствий технологического взаимодействия человека и ИИ, но не допускает переноса моральной субъектности на машину.

2. Утилитаризм (последствия как критерий оценки решений). В утилитарной философии (Д. Бентам, Дж.С. Милль) моральность действия определяется его результатом [1]. *Правильным считается тот поступок, который приносит наибольшую пользу наибольшему числу людей*. Такой подход трактует ответственность как рациональный расчет последствий.

Перенос этого принципа в область ИИ позволяет рассматривать его как инструмент оптимизации выбора. С точки зрения утилитаризма, алгоритм способен проанализировать большое количество сценариев, устранить эффект когнитивных искажений и предложить решение, наиболее соответствующее критерию пользы. Однако возникает вопрос: кто несет ответственность, если оптимальное с точки зрения алгоритма решение вступает в конфликт с моральными нормами?

Алгоритм, принимающий решение в интересах «общего блага», не осознает этических последствий. Следовательно, ответственность остается за человеком, который определяет критерии оценки и задает цель.

3. Экзистенциализм (ответственность как проявление свободы человека). Экзистенциальная философия (М. Хайдеггер, Ж.-П. Сартр) рассматривает ответственность как проявление человеческой свободы. Свобода предполагает необходимость выбора и принятия последствий, а также способность действовать осознанно. Экзистенциальный подход актуализирует вопрос личной ответственности, особенно в ситуациях неопределенности. Х. Йонас развивает эту позицию, вводя понятие *эвристического страха*, который должен предшествовать моральному действию в условиях технологического риска. Осознание масштабов возможных последствий технологических ре-

шений должно усиливать ответственность человека, а не ослаблять ее посредством передачи полномочий алгоритму [3].

Экзистенциализм помогает увидеть, что делегирование ответственности ИИ может привести к утрате субъектности человеком. Если решения принимает алгоритм, человек избавляется от труда выбора и перестает быть моральным субъектом.

4. Философия техники (машины как участники социальных процессов). Современная философия техники, включая концепции Г. Зиммеля, Ж. Симондона и представителей акторно-сетевой теории (Б. Латур, М. Каллон), анализирует технологии как активные элементы социальной реальности. Согласно этой позиции, человеческие и нечеловеческие сущности образуют единую сеть, влияющую на распределение функций и ролей [1].

В рамках акторно-сетевой теории технология рассматривается как актор, который участвует в построении социальных процессов. ИИ способен трансформировать структуру принятия решений, изменяя не только способы действия, но и само понимание ответственности. Однако признание алгоритмов в качестве функциональных акторов не означает признания их моральной субъектности. Философия техники лишь фиксирует их влияние на социальные взаимодействия.

Дополнительный элемент в эту дискуссию вносит концепция антропоцентрического проектирования М. Кули. Согласно его подходу, технологии должны усиливать человеческий потенциал, а не замещать его. М. Кули настаивает на принципе приоритета человеческого участия в управлении цифровыми решениями, чтобы сохранить содержательность человеческой деятельности и ответственность субъекта [1].

Таким образом, классические философские подходы показывают, что ИИ не может быть носителем моральной ответственности, поскольку не обладает намерением, свободой или пониманием смысла. Однако современные концепции философии техники фиксируют растущую роль алгоритмов в социальных процессах, усложняющую распределение ответственности между человеком и системой.

Современные исследования философии ИИ отражают фундаментальное изменение в понимании природы ответственности и агентности. Если классические этические школы сосредоточены на субъ-

екте, обладающем сознанием и намерением, то актуальные исследования фиксируют появление новой разновидности агентности – алгоритмической. Алгоритмы начинают участвовать в принятии решений, которые влияют на социальные процессы, и это вызывает необходимость переосмысления роли человека в условиях цифровой среды [8].

Рост автономии ИИ изменяет представление о технологии как о пассивном инструменте. В системах с машинным обучением компьютер не просто выполняет заранее запрограммированный набор действий – он выбирает образцы поведения, формируя решение на основе анализа данных. Такой механизм создает впечатление самостоятельности и приводит к дискуссии о том, может ли алгоритм обладать признаками агентности.

С точки зрения философии, автономность алгоритма отличается от автономии человеческого субъекта. Алгоритм способен к выбору, но его выбор лишен намерения, поскольку основан на вычислительной процедуре. Он не задается вопросом о цели, не переживает сомнений и не рефлексирует над последствиями. Тем не менее во многих ситуациях именно алгоритм принимает решения, влияющие на судьбы людей: распределяет ресурсы, оценивает риски, отбирает кандидатов для трудоустройства, прогнозирует вероятность правонарушений.

В работах, посвященных проблемам ответственности, прослеживается мысль, что автономность алгоритма является следствием человеческого отказа от собственной субъектности [2; 3; 6]. Авторы ставят вопрос: если машина действует вместо человека, то кто остается субъектом ответственности?

Формирование цифровой этики связано с необходимостью переосмысления моральных отношений в среде, в которой взаимодействие происходит не только между людьми, но и между человеком и ИИ. В отличие от классической этики, цифровая этика рассматривает алгоритмы как участников этического взаимодействия, принимая во внимание их влияние на человеческие решения и социальные процессы [12].

В рамках цифровой этики предложены базовые принципы:

- справедливость (устранение дискриминации и предвзятости данных);

- прозрачность (возможность объяснить логику алгоритма);
- приоритет человека (ИИ должен усиливать человеческие возможности, а не заменять их).

Международные кодексы (например, Принципы Асиломара) отмечают необходимость проектирования ИИ с учетом гуманистических ценностей и запрета на разработку систем, способных причинить вред человеку [1; 2].

Одним из острых вопросов современной философии ИИ является тенденция к «снятию ответственности» с человека. Когда решения делегируются алгоритму, человек перестает воспринимать себя как источник действия. Возникает психологический и социальный эффект: алгоритм начинает восприниматься как объективный, рациональный и потому безошибочный [7].

Психологически человек склонен доверять алгоритму, воспринимая его как лишенный эмоций и ошибок. На практике это приводит к тому, что

- человек перестает воспринимать себя источником решения;
- снижается критическое мышление;
- формируется зависимость от алгоритмических рекомендаций.

Философски этот механизм можно трактовать как утрату субъектности. Человек перестает действовать самостоятельно и превращается в наблюдателя. Чем выше уровень автономности ИИ, тем выше риск перехода человека с роли активного морального субъекта на позицию пассивного исполнителя.

Анализ современных тенденций показал, что автономность ИИ и цифровая этика формируют новую структуру взаимодействия человека с технологиями. Искусственный интеллект влияет на принятие решений, создает иллюзию объективности и может смещать ответственность с человека на алгоритм. Но несмотря на функциональную самостоятельность ИИ, моральная ответственность остается в человеческом измерении, так как субъектность предполагает намерение и осознанный выбор – качества, которыми алгоритм не обладает.

Развитие ИИ изменяет представления о распределении действий между человеком и машиной. Повышение автономности алгоритмических систем приводит к тому, что ИИ участвует в процессах, кото-

рые ранее относились исключительно к сфере человеческой рациональности и волевого акта. Это вызывает необходимость анализа того, каким образом искусственный интеллект приобретает признаки агентности и каким образом формируется иллюзия субъекта в восприятии человека.

Современные алгоритмические системы не ограничены выполнением заранее заданного набора инструкций. Они формируют решения на основе анализа данных и могут изменять модель поведения в зависимости от ситуации. В литературе это определяется как функциональная автономность, т.е. система действует не как инструмент, а как участник процесса.

Сравним основные параметры, по которым искусственный интеллект проявляет агентные свойства в сфере принятия решений (табл. 1). Хотя ИИ способен формировать решения без вмешательства человека и демонстрирует предполагаемую «самодостаточность», он не переходит в состояние субъекта. Форма его деятельности основана на вычислении, а не на переживании смысла. Системы самообучения подтверждают этот вывод. Например, алгоритмы, способные обу-

Таблица 1

Признаки агентности искусственного интеллекта в сравнении с человеческим субъектом

Критерий	Искусственный интеллект	Человек
Основание принятия решений	Логика алгоритма, статистическая оптимизация	Осознанное намерение, волевое решение
Источник данных	Наборы данных, сенсоры, цифровая среда	Личный опыт, ощущения, рефлексия
Адаптация поведения	Перенастройка модели при поступлении новых данных	Переосмысление ситуации, обучение на опыте
Скорость реакции	Высокая, вне зависимости от масштаба данных	Ограничена физиологическими возможностями
Способность к рефлексии	Нет	Есть
Наличие ответственности	Нет	Есть

чаться на собственных ошибках (AlphaGo Zero), принимают решения, которые не были предусмотрены разработчиком заранее [9]. Однако отсутствие намерения и невозможность переживания последствий ограничивают возможность считать алгоритм носителем ответственности.

Несмотря на отсутствие сознания и способности к внутреннему выбору, ИИ воспринимается человеком как субъект взаимодействия, которое связано с особенностями человеческого восприятия:

1) приписывание замысла. Человек склонен объяснять поведение машины через намерение, хотя алгоритм действует на основе вычислений, а не целей;

2) эмоциональная реакция. Взаимодействие с цифровым помощником способно вызвать эмоции, аналогичные реакции на живого собеседника: радость, раздражение или привязанность;

3) культурная интерпретация. В искусстве и массовой культуре ИИ получает человеческий образ, голос и характер.

Так формируется иллюзия, что ИИ является участником коммуникации. Человек невольно наделяет систему признаками субъекта, из-за чего становится сложно различить, где решение принадлежит алгоритму, а где – человеку.

Таким образом, ИИ проявляет признаки агентности на уровне функции, но не на уровне субъекта. Машина принимает решения, но не несет за них моральной ответственности. Переживание субъектности формируется на стороне человека и связано с антропоморфизацией цифровых систем.

Вопрос распределения ответственности между человеком и ИИ напрямую связан с тем, кем фактически является инициатор решения. Поскольку ИИ может функционировать как инструмент, как контролируемый агент и как потенциальный автономный субъект (в гипотетическом будущем), формы распределения ответственности различаются. Анализ показывает, что степень автономности алгоритма определяет не только характер человеческого участия, но и глубину морального риска. В таблице 2 представлена сравнительная модель распределения ответственности.

Анализ представленных данных показывает, что главная переменная – не функциональная сложность алгоритма, а степень сохра-

Таблица 2

Модели распределения ответственности между человеком и искусственным интеллектом

Модель	Степень автономности ИИ	Источник принятия решения	Распределение ответственности	Философская интерпретация
ИИ как инструмент (слабая автономность)	ИИ выполняет предписания, алгоритм не инициирует цель	Человек принимает и обосновывает решение	Полная ответственность лежит на человеке (разработчике или операторе)	ИИ – средство, человек – субъект морали
Контролируемая автономность	ИИ предлагает варианты на основе анализа данных, действие ограничено рамками контроля	ИИ формирует предложение, человек утверждает решение	Ответственность разделена, но решающее слово остается за человеком	Человек сохраняет статус субъекта, но появляется риск переноса ответственности
Гипотетическая полная автономия (сильный ИИ)	ИИ сам формулирует цель и принимает решение	ИИ действует независимо от человека	Ответственность юридически и морально неопределенна	Субъектность ИИ ставится под сомнение, возникает этическая дилемма

нения *человеческого контроля*. Когда ИИ используется как инструмент, человек осознает свою ответственность. Когда ИИ действует автономно, возникает риск «вывода человека из системы».

В модели *контролируемой автономности* ответственность формально закреплена за человеком, но на практике она размывается. Если алгоритм предлагает решение, а человек лишь подтверждает его, человеческая роль становится формальной. В таком случае моральная оценка теряет свою силу, поскольку человек не способен полностью объяснить работу алгоритма. Особенно остро проблема проявляется при использовании самообучающихся систем. Алгоритм способен выработать вывод, не предусмотренный разработчиком.

Формально ответственность лежит на человеке, но фактически источник решения оказывается неясен.

Гипотетическая модель *полной автономии* («сильный ИИ») представляет собой философское допущение, в котором алгоритм приобретает потенциальную субъектность. Впервые возникает вопрос: может ли носитель ответственности быть нечеловеческой системой? Здесь возникает дилемма, связанная с принципом ответственности. Так, если появляется автономный агент, необходимо заранее обеспечить встроенные ограничения его поведения. Однако эти ограничения задает человек. Следовательно, ответственность остается внутри человеческого измерения.

Таким образом, философская проблема распределения ответственности выходит за рамки права или программирования. Она связана с фундаментальным вопросом: можем ли мы передать ответственность тому, кто не обладает сознанием?

Пока ИИ не имеет сознания и не является носителем намерения, он не может быть субъектом моральной ответственности. Даже если ИИ проявляет агентные признаки – действует автономно, обучается, принимает решения, ответственность остается у человека как единственного существа, осознающего последствия своего выбора. Это означает, что ответственность за будущее технологий – не только технический вопрос, но и акт человеческой зрелости и способности к моральному самоограничению.

Рост автономности ИИ указывает на необходимость определения границ ответственности между участниками, вовлеченными в создание и использование алгоритмических систем. Несмотря на расширение функциональных возможностей ИИ, вопрос субъектности остается принципиальным: система способна принимать решения, но не способна *осознавать* их последствия.

Анализ показывает, что ответственность возникает только там, где присутствуют сознание, намерение и способность к моральной рефлексии. Эти свойства присущи человеку, но отсутствуют у алгоритма. Поэтому в любом сценарии применения ИИ именно человек остается носителем моральной ответственности.

Для анализа структуры распределения ответственности целесообразно использовать модель, основанную на философских критериях: субъектности, наличии намерения, способности к пониманию

Таблица 3

Философская модель распределения ответственности между участниками взаимодействия с ИИ

Агент ответственности	Наличие субъектности	Намерение (интенциональность)	Способность понимать последствия
Разработчик	Да. Личность, обладающая свободой выбора и моральной рефлексией	Определяет цели работы системы, архитектуру и ограничения	Обязан учитывать долгосрочные риски применения технологии
Пользователь	Да. Полноценный носитель ответственности	Применяет систему в конкретной ситуации и принимает итоговое решение	Должен осознавать условия использования алгоритма
Алгоритм	Нет. ИИ является техническим средством	Намерение отсутствует, решения формируются вычислительным путем	Не понимает последствий, не испытывает эмоций, не способен к моральному выбору

последствий (табл. 3). Модель позволяет увидеть принципиальное отличие алгоритма от человека: система может принимать решение, но *не нести* ответственности за него. Причиной является отсутствие субъективного измерения – сознания, чувства вины, страха, способности к переживанию последствий и переосмыслению ошибок.

Даже в случае адаптивного обучения и высокой автономности поведение ИИ остается производной от человеческого кода, данных обучения и заданных целей. Если система принимает решение, которое приводит к негативному результату, ответственность лежит не на алгоритме, а на человеке как источнике первоначального замысла.

В философском понимании ответственность всегда предполагает *возможность действовать иначе*. Алгоритм не обладает этой возможностью, так как он не выбирает ценности, а оптимизирует параметры. Алгоритм может воспроизводить интеллектуальное действие, но не может совершать моральный поступок. Отсюда следует основ-

ной вывод: негативные последствия, возникающие в результате применения искусственного интеллекта, – проявление не «бесчеловечности машины», а *антропогенного бесчеловечного решения*, скрытого в алгоритме.

Снижение доли ответственности разработчиков и пользователей ведет к подмене человеческого решения машинной процедурой и нарушает этический баланс взаимодействия «человек – технология». Даже если алгоритм получает высокую степень автономности, он остается вычислительной системой, лишенной самосознания. Вследствие этого он неспособен к подлинному моральному выбору.

* * *

Современное внедрение ИИ в сферы принятия решений выявило фундаментальный философский вызов – необходимость определить границы ответственности между человеком и алгоритмом. Алгоритм может демонстрировать интеллектуальное поведение, проявлять признаки агентности и функционировать как самостоятельная вычислительная система, однако анализ подтверждает, что эти признаки не переводят ИИ в статус субъекта. Он остается инструментом, выполняющим задачи, поставленные человеком. Несмотря на внешнюю автономность, ИИ не обладает способностью к осмыслению своих действий, не переживает намерения и не может принимать решения в моральном смысле.

Главное ограничение ИИ заключается в отсутствии внутреннего сознания, эмоциональной рефлексии и способности оценивать последствия действий. Именно эти свойства формируют основу человеческой ответственности и морального выбора. Алгоритм может моделировать поведение человека и действовать продуктивнее его в отдельных задачах, но неспособен переживать вину, стыд, страх или сострадание – те феноменологические состояния, которые лежат в основании нравственного поступка. Следовательно, ответственность за последствия функционирования ИИ неизбежно остается в человеческом измерении. Она принадлежит тем, кто проектирует, внедряет и применяет технологию.

Увеличение автономности цифровых систем не отменяет человеческую ответственность, но меняет ее форму. Человек должен не только контролировать алгоритм, но и осознавать последствия вне-

дрения систем, способных влиять на социальные процессы. Существует риск смещения ответственности с человека на алгоритм, когда решения ИИ воспринимаются как объективные, что ведет к утрате человеческой субъектности и превращению человека в пассивного наблюдателя. Преодоление этого риска возможно только при осознании ценности человеческого участия и утверждении гуманистического подхода к технологиям. Развитие цифровой этики, нормативного регулирования и философского анализа автономности ИИ формирует основу для безопасного сосуществования технологий и человека. В конечном счете задача философии – обеспечить, чтобы развитие ИИ сопровождалось сохранением человеческой ответственности, гуманности и нравственного смысла.

Литература

1. *Абрамова О.А*. Общество и искусственный интеллект: путь к человекоцентрированному подходу // Информационное общество. 2020. № 5. С. 10–21.

2. *Айбуев А.-Х.А.-М., Цабаев Х.И*. Этические аспекты искусственного интеллекта: философский анализ взаимодействия человека и технологии // Экономика и управление: проблемы, решения. 2025. Т. 9, № 1 (154). С. 145–150. DOI: 10.36871/ek.up.p.r.2025.01.09.017.

3. *Бадмаева М.Х*. Этика искусственного интеллекта: принцип ответственности Ганса Йонаса // Вестник Бурятского государственного университета. 2022. № 1. С. 67–79. DOI: 10.18101/1994-0866-2022-1-67-79.

4. *Баженов С.С*. Искусственный интеллект как экзистенциальная угроза: от эсхатологического сюжета к нормативному регулированию // Информация – Коммуникация – Общество. 2023. Т. 1. С. 14–18.

5. *Баженов С.С*. Искусственный интеллект: от онтологических выводов к мировоззренческим установкам // Информация – Коммуникация – Общество. 2022. Т. 1. С. 14–18.

6. *Баженов С.С*. К проблеме моральной агентности искусственного интеллекта // Дискурсы этики. 2023. № 1 (17). С. 13–30.

7. *Величковский Б.М*. Психологические аспекты искусственного интеллекта // Новости искусственного интеллекта. 1995. № 2. С. 80–120.

8. *Горячев А.Ю., Смирнов С.В*. Учение о двойном последствии и искусственный интеллект // Социально-политические науки. 2024. Т. 14, № 2. С. 211–215. DOI: 10.33693/2223-0092-2024-14-2-211-215.

9. *Иерусалимский Ю.Ю., Лукашенко С.П*. Этические аспекты искусственного интеллекта // Медицинская этика. 2025. Т. 13, № 1. С. 15–19. DOI: 10.24075/medet.2025.032.

10. *Иоселиани А.Д*. «Искусственный интеллект» vs человеческий разум // Манускрипт. 2019. Т. 12, № 4. С. 102–107. DOI: 10.30853/manuscript.2019.4.21.

11. *Латфуллина Д.Р.* Человеческий разум и искусственный интеллект // Ученые записки Казанского филиала «Российского государственного университета правосудия». 2018. Т. 14. С. 512–516.

12. *Логунова Е.П., Бекетова Т.С.* Этические аспекты философии искусственного интеллекта // Наука и Образование. 2023. Т. 6, № 2.

13. *Попов Д.В.* Искусственный интеллект: между человечностью и бесчеловечным // Дискурс-Пи. 2022. Т. 19, № 3. С. 138–156. DOI: 10.17506/18179568_2022_19_3_138.

14. *Самойлова Л.П., Захаров И.С., Гатаулин Г.Р.* Ключевые проблемы философии XXI века: феномен сознания и этика искусственного интеллекта // Культура. Наука. Производство. 2024. № 13. С. 84–89. DOI: 10.52978/26187701_2024_13_84-89.

15. *Смаковский В.Н.* Философия искусственного интеллекта // Международный журнал информационных технологий и энергоэффективности. 2024. Т. 9, № 1 (39). С. 14–18.

References

1. *Abramova, O.A*. (2020). Society and artificial intelligence: path to the human-centered approach . Informatsionnoe obshchestvo [Information Society], 5, 10–21. (In Russ.).

2. *Aybuev, A.-H.A.-M. & H.I. Tsabaev*. (2025). Ethical aspects of artificial intelligence: a philosophical analysis of human–technology interaction. Ekonomika i upravlenie: problemy, resheniya [Economics and Management: Problems, Solutions], Vol. 9, No. 1 (154), 145–150. DOI: 10.36871/ek.up.p.r.2025.01.09.017. (In Russ.).

3. *Badmaeva, M.Kh*. (2022). Ethics of artificial intelligence: Hans Jonas' principle of responsibility. Vestnik Buryatskogo gosudarstvennogo universiteta [The Buryat State University Bulletin], 1, 67–79. DOI: 10.18101/1994-0866-2022-1-67-79. (In Russ.).

4. *Bazhenov, S.S.* (2023). Artificial intelligence as an existential threat: from an eschatological plot to normative regulation. Informatsiya – kommunikatsiya – obshchestvo [Information–Communication–Society], 1, 14–18. (In Russ.).

5. *Bazhenov, S.S.* (2022). Artificial intelligence: from ontological conclusions to worldview attitudes. Informatsiya – kommunikatsiya – obshchestvo [Information–Communication–Society], 1, 14–18. (In Russ.).

6. *Bazhenov, S.S.* (2023). Moral agency of artificial intelligence / S. S. Bazhenov. Diskursy etiki [Discourses of Ethics], 1 (17), 13–30. (In Russ.).

7. *Velichkovsky, B.M*. (1995). Psychological aspects of artificial intelligence. Novosti iskusstvennogo intellekta [Artificial Intelligence News], 2, 80–120. (In Russ.).

8. *Goryachev, A.Yu. & S.V. Smirnov*. (2024). The teaching of double effect and artificial intelligence . Sotsialno-politicheskie nauki [Socio-Political Sciences], Vol. 14, No. 2, 211–215. DOI: 10.33693/2223-0092-2024-14-2-211-215. (In Russ.).

9. *Ierusalimsky, Yu.Yu. & S.P. Lukashenko*. (2025). Ethical aspects of artificial intelligence. Meditsinskaya etika [Medical Ethics], Vol. 13, No. 1, 15–19. DOI: 10.24075/medet.2025.032. (In Russ.).

10. *Ioseliani, A.D*. (2019). “Artificial intelligence” vs human intelligence. Manuskript [Manuscript], Vol. 12, No. 4, 102–107. DOI: 10.30853/manuscript.2019.4.21. (In Russ.).

11. *Latfullina, D.R*. (2018). The human mind and artificial intelligence. Uchenye zapiski Kazanskogo filiala “Rossiyskogo gosudarstvennogo universiteta pravosudiya” [Scientific Notes of the Kazan Branch of the Russian State University of Justice], 14, 512–516. (In Russ.).

12. *Logunova, E.P. & T.S. Beketova*. (2023). Ethical aspects of the philosophy of artificial intelligence. Nauka i obrazovanie [Science and Education], Vol. 6, No. 2. (In Russ.).

13. *Popov, D.V*. (2022). Artificial intelligence: between humanity and the inhuman. Diskurs-Pi [Discourse-P], Vol. 19, No. 3, 138–156. DOI: 10.17506/18179568_2022_19_3_138. (In Russ.).

14. *Samoylova, L.P., I.S. Zakharov & G.R. Gataulin*. (2024). Key problems of 21st century philosophy: The phenomenon of consciousness and the ethics of artificial intelligence. Kultura. Nauka. Proizvodstvo [Culture. Science. Production], 13, 84–89. DOI: 10.52978/26187701_2024_13_84-89. (In Russ.).

15. *Smakovsky, V.N*. (2024). The philosophy of artificial intelligence. Mezhdunarodnyy zhurnal informatsionnykh tekhnologiy i energoeffektivnosti [International Journal of Information Technology and Energy Efficiency], Vol. 9, No. 1 (39), 14–18. (In Russ.).

Информация об авторах

Плотников Валерий Валерьевич – Кубанский государственный аграрный университет им. И.Т. Трубилина (350044, Краснодар, ул. Калинина, 13).
antidoxiya84@mail.ru

Косников Максим Сергеевич – Кубанский государственный аграрный университет им. И.Т. Трубилина (350044, Краснодар, ул. Калинина, 13).
sn_03@rambler.ru

Information about the authors

Plotnikov, Valery Valerievich – Kuban State Agrarian University named after I.T. Trubilin (13, Kalinina St., Krasnodar, 350044, Russia).
antidoxiya84@mail.ru

Kosnikov, Maxim Sergeevich – Kuban State Agrarian University named after I.T. Trubilin (13, Kalinina St., Krasnodar, 350044, Russia).
sn_03@rambler.ru

Дата поступления 25.10.2025.
Принята к публикации 02.02.2026.