π-Economy. 2026, Том 19, № 5. С. 158–176.
π-Economy. 2026, Vol. 19, No. 5. Pp. 158–176.

Научная статья
УДК 005.53:004.8
DOI: https://doi.org/10.18721/JE.19509
EDN: https://elibrary/IOAYHP

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ВЗАИМОДЕЙСТВИЯ ЧЕЛОВЕКА И ИНТЕЛЛЕКТУАЛЬНОГО АГЕНТА В СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ

Н.С. Лукашевич ✉

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ lukashevich@spbstu.ru

Аннотация. Распространение генеративных моделей и автономных систем на основе искусственного интеллекта (ИИ-агентов) трансформирует архитектуру управленческих процессов, изменяя распределение функций, полномочий и ответственности в системах поддержки принятия управленческих решений. Такая трансформация актуализирует научную проблему обоснованного делегирования полномочий ИИ-агентам при сохранении субъектной роли и автономии человека в процессе принятия решений. Несмотря на значительное число исследований по алгоритмическому управлению и этике ИИ, в управленческой науке отсутствует методологический аппарат, позволяющий определять допустимый уровень автономности ИИ-агента с учетом характеристик управленческих задач и организационно-технической среды. Цель исследования состоит в разработке концептуальной модели и принципов человеко-агентного взаимодействия в системах поддержки принятия решений, позволяющих обосновывать делегирование ИИ-агенту полномочий и определять уровень его автономности. В исследовании использованы системный подход, сравнительный анализ существующих моделей автономности интеллектуальных систем, контент-анализ научной литературы, концептуальное моделирование и логико-структурный синтез. Методологическую основу составил междисциплинарный подход, объединяющий положения теории принятия решений, гуманистического и алгоритмического управления, организационной теории и прикладной этики ИИ. С позиций гуманистической теории организации обосновано разграничение автономности ИИ-агента и автономии человека. На этой основе разработана матрица «автономия–автономность», включающая четыре типа человеко-агентного взаимодействия, для каждого из которых определены управленческие механизмы сохранения автономии человека. Разработаны методологические принципы ответственного взаимодействия и делегирования полномочий ИИ-агенту, закрепляющие приоритет профессионального суждения человека в процессе принятия решений. Определены состав и взаимосвязи элементов концептуальной модели взаимодействия человека и ИИ-агента. Модель предусматривает механизмы контроля соблюдения условий делегирования и обратной связи. Разработана шкала автономности ИИ-агента, отражающая степень его участия в принятии управленческих решений. Формализована процедура определения допустимого уровня автономности на основе агрегирования экспертных оценок параметров управленческой задачи и технологической зрелости ИИ-агента. Научная новизна заключается в теоретическом обосновании концепции динамического и контекстно обусловленного делегирования ИИ-агенту полномочий по принятию управленческих решений. Автономность ИИ-агента рассматривается как управляемый параметр, зависящий от характеристик задачи и организационно-технической среды. Автономия человека выступает гуманистической ценностью и основанием для разграничения функциональной самостоятельности ИИ-агента и моральной субъектности человека, за которым сохраняется приоритет в осуществлении контроля и в принятии окончательного решения. Результаты исследования формируют методическую основу для обоснования делегирования ИИ-агенту полномочий применительно к различным типам управленческих задач и могут использоваться при ответственном внедрении ИИ в управленческие процессы, при

© Лукашевич Н.С., 2026. Издатель: Санкт-Петербургский политехнический университет Петра Великого

разработке внутренних регламентов и этических кодексов. Перспективы дальнейших исследований связаны с эмпирической апробацией результатов в различных организационных контекстах и с методической детализацией механизмов контроля и сохранения автономии человека в гибридных системах поддержки принятия решений.

Ключевые слова: человеко-агентное взаимодействие, интеллектуальный агент, ИИ-агент, управленческие решения, автономность ИИ, автономия человека, гуманистическое управление, делегирование

Для цитирования: Лукашевич Н.С. (2026) Концептуальная модель взаимодействия человека и интеллектуального агента в системах поддержки принятия управленческих решений. π-Economy, 19 (5), 158–176. DOI: https://doi.org/10.18721/JE.19509

Research article
DOI: https://doi.org/10.18721/JE.19509

A CONCEPTUAL MODEL OF HUMAN-INTELLIGENT AGENT INTERACTION IN MANAGERIAL DECISION SUPPORT SYSTEMS

N.S. Lukashevich ✉

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ lukashevich@spbstu.ru

Abstract. The proliferation of generative models and AI agents is transforming the architecture of managerial processes by changing the distribution of functions, authority, and responsibility within management decision support systems. This transformation raises the research problem of ensuring well-founded delegation of authority to AI agents while preserving human agency and autonomy in decision-making. Despite extensive research on algorithmic management and AI ethics, management science still lacks a methodological framework for determining the permissible level of AI-agent autonomy with due regard to the characteristics of managerial tasks and the organizational and technological environment. The study aims to develop a conceptual model of human–agent interaction in decision support systems that provide a basis for delegating authority to an AI agent and determining its level of autonomy. The study employs a systems approach, comparative analysis of existing models of intelligent-system autonomy, content analysis of the academic literature, conceptual modelling, and logical and structural synthesis. Its methodological foundation is an interdisciplinary approach integrating decision theory, humanistic and algorithmic management, organizational theory, and applied AI ethics. Drawing on the humanistic theory of organization, the study substantiates the distinction between AI-agent autonomy and human autonomy. On this basis, a human autonomy – AI-agent autonomy matrix is developed. It identifies four types of human-agent interaction and specifies management mechanisms for preserving human autonomy within each type. The study formulates methodological principles of responsible interaction and delegation of authority to an AI agent, establishing the priority of human professional judgement in decision-making. It also defines the components of the conceptual model of human–agent interaction and the relationships among them. The model incorporates mechanisms for monitoring compliance with the conditions of delegation and providing feedback. An AI-agent autonomy scale is developed to reflect the extent of the agent's involvement in managerial decision-making. A procedure for determining the permissible level of autonomy is formalized by aggregating expert assessments of the parameters of the managerial task and the technological maturity of the AI agent. The scientific novelty of the study lies in the theoretical substantiation of dynamic and context-dependent delegation of managerial decision-making authority to an AI agent. AI-agent autonomy is treated as a manageable parameter determined by task characteristics and the organizational and technological environment. Human autonomy is regarded as a humanistic value and as the basis for distinguishing the functional autonomy of an AI agent from human moral agency. Humans retain priority in exercising control and making the final decision. The findings provide a methodological basis for substantiating the delegation of

© Lukashevich N.S., 2026. Published by Peter the Great St. Petersburg Polytechnic University

authority to AI agents across different types of managerial tasks. They may support the responsible integration of AI into managerial processes and the development of internal regulations and codes of ethics. Further research should empirically validate the findings in different organizational contexts and elaborate the mechanisms for human control and the preservation of human autonomy in hybrid decision support systems.

Keywords: human-agent interaction, intelligent agent, AI agent, managerial decisions, AI autonomy, human autonomy, humanistic management, delegation

Citation: Lukashevich N.S. (2026) A conceptual model of human-intelligent agent interaction in managerial decision support systems. π-Economy, 19 (5), 158–176. DOI: https://doi.org/10.18721/JE.19509

Введение

Искусственный интеллект (ИИ) становится фактором развития многих отраслей[1], а его внедрение в управленческие процессы порождает фундаментальное противоречие. ИИ расширяет возможности обработки данных, выявления закономерностей и повышения оперативности принятия решений. Вместе с тем делегирование алгоритмам аналитических и управленческих функций ведет к диффузии ответственности, утрате профессионального суждения и ослаблению автономии человека [1–3]. Это противоречие усиливается переходом к организационным системам, в которых интеллектуальные агенты (ИИ-агенты) самостоятельно декомпозируют цели, планируют действия и взаимодействуют с другими системами, тогда как за человеком остается роль носителя целей, смыслов и ответственности. ИИ перестает быть инструментом поддержки и становится активным участником управленческого процесса, трансформируя саму природу управленческого решения.

Согласно отчету[2], 60% руководителей регулярно применяют ИИ для поддержки принятия решений, а к 2027 г. половина решений может быть автоматизирована ИИ-агентами. Экономика вступает в фазу «агентного скачка»: генеративный ИИ выполняет функции, ранее считавшиеся прерогативой человека, – формирует гипотезы, обобщает знания и действует при ограниченном участии оператора. В технологическом прогнозе[3] подчеркивается ускоренный рост алгоритмизации и развитие ИИ-агентов. При этом среди ключевых условий масштабирования ИИ выделяются доверие пользователя и безопасное человеко-машинное (агентное) взаимодействие. Российский рынок ИИ-агентов пока не сформирован как самостоятельный сегмент, но его развитие остается стратегическим приоритетом при сокращении венчурного финансирования[4]. В условиях формирующегося технологического рынка и недостаточной организационно-управленческой проработки вопросов применения ИИ-агентов организациям нужны критерии допустимого делегирования полномочий.

Функциональные возможности ИИ-агентов развиваются быстрее, чем организационные практики, правовое регулирование и этические нормы адаптируются к их применению. Классическая модель «человек в контуре» (human-in-the-loop) перестает гарантировать его реальное участие в принятии решений и порождает «алгоритмическое алиби», при котором человек формально сохраняет право решения, но передает алгоритму содержательную оценку ситуации. Эти

[1] Технологии Доверия (2023) *Искусственный интеллект в HR: Кейсы российского рынка*. [online] Available at: https://data.tedo.ru/publications/ai-in-hr.pdf [Accessed 30.05.2026]. (in Russian); IEEE Computer Society (2026) *Technology Predictions 2026*. [online] Available at: https://ict.moscow/static/pdf/files/4533eaf6-be20-5126-bdc6-bb29950e666d.pdf [Accessed 30.05.2026]; World Economic Forum (2025) *Future of Jobs: AI and Talent 2030*. [online] Available at: https://reports.weforum.org/docs/WEF_Four_Futures_for_Jobs_in_the_New_Economy_AI_and_Talent_in_2030_2025.pdf [Accessed 30.05.2026].

[2] Deloitte (2026) *Global Human Capital Trends 2026*. [online] Available at: https://www.deloitte.com/us/en/insights/topics/talent/human-capital-trends.html [Accessed 30.05.2026].

[3] World Economic Forum (2025) *Future of Jobs: AI and Talent 2030*. [online] Available at: https://reports.weforum.org/docs/WEF_Four_Futures_for_Jobs_in_the_New_Economy_AI_and_Talent_in_2030_2025.pdf [Accessed 30.05.2026].

[4] МТС (2025) *ИИ-агенты: Обзор мирового и российского рынков*. [online] Available at: https://d-economy.ru/wp-content/uploads/2026/02/mts-research-insights-ai-agents-2025.pdf [Accessed 30.05.2026] (in Russian)

риски особенно значимы в контексте парадигмы «Индустрия 5.0» [4, 5], утверждающей человекоцентричную, устойчивую и резильентную модель развития, в которой технологии должны расширять возможности человека, а не замещать его суждение.

Существующая литература обосновывает необходимость человекоцентричного ИИ [6–12], а нормативно-правовое регулирование[5] устанавливает риск-ориентированную классификацию ИИ-систем, принцип обязательного контроля со стороны человека и процессные требования к управлению ИИ на всем жизненном цикле. Однако недостаточно концептуализирована проблема определения допустимого уровня автономности ИИ-агента и критериев делегирования ему полномочий в зависимости от параметров организационной среды.

Таким образом, исследование направлено на решение научной проблемы организации человеко-агентного взаимодействия в системах поддержки принятия управленческих решений, содержательную основу которого составляет обоснованное распределение полномочий между человеком и ИИ-агентом при сохранении автономии, субъектности и ответственности человека.

Объектом исследования выступают процессы взаимодействия человека и ИИ-агента в человекоцентричной гибридной системе поддержки принятия управленческих решений. Предмет исследования составляют принципы, механизмы и процедуры делегирования ИИ-агенту полномочий, определения уровня его автономности и контроля его участия в принятии управленческих решений.

В исследовании ИИ-агент определяется в нескольких аспектах:

1) как программная «сущность», предназначенная для достижения целей в определенной среде [13];

2) как агент, обладающий автономностью – способностью выполнять задачи без участия человека, обучаться, изменять поведение и взаимодействовать со средой;

3) как участник человеко-агентных систем [14].

Распространение ИИ-агентов рассматривается как следующий этап развития технологий генеративного ИИ[6]. Организационно-техническая среда – это совокупность технологических, процессных и нормативно-культурных условий организации, формирующих контекст человеко-агентного принятия управленческих решений. В настоящем исследовании человеко-агентное взаимодействие рассматривается как совместное участие человека и ИИ-агента в принятии управленческих решений. Делегирование – это передача ограниченных полномочий на выполнение определенных действий и принятие решений с учетом нормативных требований, функциональных ограничений и уровня риска.

Исследовательская позиция состоит в том, что человекоцентричное управление в условиях развития ИИ требует перехода от общей парадигмы «человек в контуре» к моделям содержательного контроля. Модель задает границы делегирования, формы контроля и способы защиты автономии, достоинства и субъектности человека при внедрении ИИ.

Исследования человеко-агентного взаимодействия являются предметом теории алгоритмического и гуманистического управления, техноэтики, психологии личности и проектирования алгоритмических систем. Вместе с тем эти направления не образуют целостной методологической основы для определения допустимой автономности ИИ и распределения полномочий по принятию решений.

---

[5] EUR-Lex (2024) *Regulation* (*EU*) *2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations* (*EC*) *No 300/2008*, (*EU*) *No 167/2013*, (*EU*) *No 168/2013*, (*EU*) *2018/858*, (*EU*) *2018/1139 and* (*EU*) *2019/2144 and Directives 2014/90/EU*, (*EU*) *2016/797 and* (*EU*) *2020/1828* (*Artificial Intelligence Act*) (*Text with EEA relevance*). [online] Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng [Accessed: 30.05.2026]; ISO/IEC (2023) ISO/IEC 42001: *Information technology – Artificial intelligence – Management system*. [online] Available at: https://cdn.standards.iteh.ai/samples/81230/4c1911ebc9a641f-cb6ee21aa09c28ad3/ISO-IEC-42001-2023.pdf [Accessed 30.05.2026].

[6] МТС (2025) *ИИ-агенты: Обзор мирового и российского рынков*. [online] Available at: https://d-economy.ru/wp-content/uploads/2026/02/mts-research-insights-ai-agents-2025.pdf [Accessed 30.05.2026]. (in Russian)

Анализ концепции «Общество 5.0» [14] выявляет противоречие между декларируемым принципом человекоцентричности и применением ИИ-агентов в управленческих процессах. Противоречие отражено в тезисе М. Гладдена [14] о распределенной агентности, согласно которому, управленческий результат совместно формируют человек и ИИ-агенты в определенной цифровой организационной среде.

Исследования, основанные на теории самодетерминации и модели многоуровневого технологического опыта [15], отражают двойственное влияние ИИ-агентов на автономию человека. Технологии могут поддерживать самостоятельность выбора, однако манипулятивное воздействие и адаптация предпочтений ограничивают автономию человека. Поэтому формальное включение человека «в контур» принятия решений не гарантирует контроля, который предполагает понимание логики решения, право отклонить его, возможность оспаривания и полноценное участие в процедуре.

Литература по алгоритмическому управлению уточняет, как именно цифровые системы воздействуют на работника. Алгоритмы становятся «новым полем управленческого контроля» [16], формируя архитектуру выбора, фиксируя поведение и дисциплинируя через автоматические стимулы. Развивающая этот подход концепция дуальности алгоритмического управления [17] показывает, что алгоритмы способны ограничивать или поддерживать автономию работников в зависимости от способа их проектирования и применения, а сами работники своими действиями изменяют систему. Это положение легло в основу теории алгоактивизма [18], рассматривающей формы организованного коллективного сопротивления алгоритмическим практикам управления. Для настоящего исследования алгоактивизм значим как индикатор нарушения баланса между автономностью ИИ и автономией человека, что делает механизмы защиты последней необходимым элементом устойчивой человекоцентричной системы.

Результаты исследований по проблемам распределения ролей между человеком и ИИ представлены, в частности, шкалой агентности человека [13], разграничивающей задачи, полностью выполняемые ИИ, задачи равноправного партнерства и задачи с ведущей ролью человека в принятии решений. Эмпирически подтверждается неоднородность предпочтений работников. Например, согласие на автоматизацию рутинных операций обусловлено стремлением сохранить участие в значимых креативных задачах. Шкала агентности человека остается диагностическим инструментом и не определяет механизмы изменения уровня автономности и поведения ИИ-системы в ситуациях риска, ошибки или этического конфликта.

Исследования в области оценки экономического эффекта ИИ[7] [19] показывают, что влияние ИИ-инструментов на производительность зависит от типологии задач, готовности и компетенций работников, уровня доверия, нормативно-правового регулирования и способности организаций трансформировать процессы. Оценка эффекта алгоритмизации требует разграничения ее экономической целесообразности и соответствия гуманистическим принципам. В логике человекоцентричных организаций [20, 21] внедрение ИИ-агентов должно оцениваться по их эффективности, влиянию на развитие человекоцентричных практик и на трансформацию управленческой парадигмы в целом [9].

Социальное измерение проблемы отражают данные индекса цифрофобий[8], свидетельствующие о том, что тревоги граждан обусловлены угрозами утраты контроля, приватности, справедливого отношения и автономии мышления, возникающими при применении цифровых технологий. В этом контексте цифровые фобии можно рассматривать как социальный индикатор недостаточной управляемости и человекоцентричности ИИ-систем. Это согласуется с тезисом

[7] World Economic Forum (2025) *Future of Jobs: AI and Talent 2030*. [online] Available at: https://reports.weforum.org/docs/WEF_Four_Futures_for_Jobs_in_the_New_Economy_AI_and_Talent_in_2030_2025.pdf [Accessed 30.05.2026]; МТС (2025) *ИИ-агенты: Обзор мирового и российского рынков*. [online] Available at: https://d-economy.ru/wp-content/uploads/2026/02/mts-research-insights-ai-agents-2025.pdf [Accessed 30.05.2026] (in Russian)

[8] ВЦИОМ (2025) *Индекс цифрофобии: Всероссийский репрезентативный опрос «ВЦИОМ-Спутник»*. [online] Available at: https://wciom.ru/fileadmin/user_upload/indeks_cifrofobii_061025.pdf [Accessed 30.05.2026]. (in Russian)

о социальном конструировании технологического будущего[9]. Задача гуманистического управления заключается в определении перспектив организационного развития, предусматривающих согласование экономической эффективности с развитием человека, демократизацией управления и сохранением его субъектности.

Внимания заслуживает положение о необходимости согласования автоматизации и аугментации в сфере управления человеческими ресурсами, имеющей высокую этическую значимость [22–24]. Кадровые решения влияют на достоинство, карьеру и профессиональное развитие человека, поэтому автономность ИИ-агента в этой сфере требует нормативных ограничений.

Отдельное направление исследований связано с научной дискуссией о возможности признания ИИ-агентов самостоятельными моральными субъектами. Исследования по ответственному ИИ[10] [11, 13, 15, 23] определяют несоответствие между декларируемыми этическими принципами и практикой их внедрения в управленческие процессы. Повышение автономности ИИ-агента способствует диффузии ответственности и формирует «алгоритмическое алиби», позволяющее обосновывать рекомендациями ИИ-агента ограниченное участие человека в этически значимых решениях. Одновременно возрастает риск дегуманизации управления, поскольку ориентация ИИ-агентов на экономическую эффективность может приводить к неполноценному анализу эмоционального, социального и неформального контекстов и не гарантировать соответствие эффективного решения гуманистическим ограничениям [10]. Систематическое делегирование ИИ-агенту анализа и проверки решений влияет на практику профессионального суждения и компетенции критической оценки, что приводит только к формальному экспертному участию, создающему видимость контроля при снижении качества профессионального суждения. Дополнительные риски обусловлены неопределенностью, связанной с качеством исходных данных и непрозрачностью процедур их отбора [22]. Ориентация ИИ-агента на экстраполяцию прошлых трендов и закономерностей может исключать новые сценарии развития, поэтому человекоцентричный подход к внедрению ИИ требует механизмов критического анализа и обоснованного отклонения алгоритмических рекомендаций.

Обзор ИИ-решений на основе больших языковых моделей [25] подтверждает, что внедрение полностью автономных ИИ-агентов ограничивается «галлюцинациями», недостаточной надежностью и неустойчивостью результатов, рисками безопасности и несанкционированными действиями, а также ограниченной прозрачностью решений и диффузией ответственности. Исследователи [25] предлагают снижать указанные риски посредством распределения функций между человеком и ИИ. Агентоцентричность такого подхода определяет человека как средство повышения надежности алгоритмической системы, недостаточно учитывая его роль как субъекта морального суждения. В работе [26] предлагается классификация, разделяющая автономность (способность ИИ самостоятельно выполнять задачи) и моральную компетентность (способность оценивать нравственные последствия и принимать этически обоснованные решения).

Теоретический обзор выявил два ограничения существующих подходов. Методическое ограничение заключается в недостаточной разработанности критериев и процедур обоснования допустимого уровня автономности ИИ-агента. Нормативно-этическое ограничение состоит в отсутствии критериев отнесения управленческих задач к неделегируемым.

Целью исследования являются разработка и теоретическое обоснование концептуальной модели взаимодействия человека и ИИ-агента, определяющей правила делегирования полномочий и выбора допустимого уровня автономности при принятии управленческих решений.

---

[9] Фёдоров В. (2025) *Моделирование будущего: зачем и каким образом?* [online] Available at: https://wciom.ru/fileadmin/user_upload/modelirovanie_budushchego_zachem_i_kakim_obrazom.pdf [Accessed 30.05.2026]. (in Russian)

[10] EUR-Lex (2024) *Regulation* (*EU*) *2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations* (*EC*) *No 300/2008*, (*EU*) *No 167/2013*, (*EU*) *No 168/2013*, (*EU*) *2018/858*, (*EU*) *2018/1139 and* (*EU*) *2019/2144 and Directives 2014/90/EU*, (*EU*) *2016/797 and* (*EU*) *2020/1828* (*Artificial Intelligence Act*) (*Text with EEA relevance*). [online] Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng [Accessed: 30.05.2026].

Проектирование гибридных систем поддержки управленческих решений целесообразно ориентировать не на достижение максимальной автономности ИИ-агента, а на обеспечение автономии человека и его профессиональной компетенции. Определение допустимого уровня автономности ИИ-агента должно осуществляться с учетом типа управленческого решения, степени его этической значимости, уровня рисков, организационно-технических возможностей реализации соответствующего уровня автономности.

Методы и материалы

Исследование основано на теоретическом анализе публикаций о человекоцентричности, алгоритмическом управлении и человеко-агентном взаимодействии в контексте распределения ответственности, объяснимости алгоритмов, доверия и автономии человека.

На первом этапе сформирована гуманистическая перспектива исследования и разработаны методологические принципы проектирования гибридных систем поддержки принятия управленческих решений. Гуманистическое управление рассматривает человека как субъекта, обладающего достоинством, автономией и способностью осмысленно участвовать в управлении. Далее проведен анализ понятий и категорий, включая теоретическое разграничение автономности ИИ-агента и автономии человека. На втором этапе проведен анализ моделей человеко-агентного взаимодействия и делегирования ИИ-агентам полномочий по принятию решений с точки зрения типа управленческой задачи, уровня риска, технологической зрелости ИИ-агента. На третьем этапе на основе системного подхода предложены элементы и организационно-управленческие механизмы концептуальной модели.

Методы и последовательность этапов позволяют перейти от теоретической предпосылки «человек в контуре» к процедуре, учитывающей параметры организационно-технической среды и характеристики управленческих задач, а также обеспечивающей распределение полномочий в системах поддержки принятия управленческих решений.

Для достижения поставленной цели в исследовании решались следующие задачи:

1) провести анализ теоретических подходов к пониманию автономности ИИ-агентов и автономии человека, раскрыв принципиальное различие между этими категориями, а также выявить ограничения существующих подходов к человеко-агентному взаимодействию при распределении полномочий и определении допустимого уровня автономности;

2) разработать методологические принципы проектирования гибридных систем поддержки принятия решений и обосновать необходимость ограничения автономности ИИ-агента с позиций теории гуманистического управления;

3) определить состав и взаимосвязи элементов концептуальной модели, включая механизмы контроля участия ИИ-агента в принятии решений;

4) разработать критерии и процедуру экспертного определения допустимого уровня автономности ИИ-агента.

Результаты и обсуждение

Обзор литературы, посвященной автономии человека в человеко-агентных системах [1–3, 15, 27–31], показывает, что этот феномен трактуется неоднозначно и объединяет несколько подходов:

1) автономия как защита от манипуляций [3, 28];

2) автономия как способность сохранять контроль над алгоритмической системой [29];

3) автономия как возможность оспаривать алгоритмическое решение;

4) автономия как антропологическая ценность, нередуцируемая к проблеме ценностного выравнивания [1];

5) автономия как организационное свойство, подлежащее диагностике [27].

Автономия человека	**Автономность ИИ-агента**	
	НИЗКАЯ	**ВЫСОКАЯ**
ПОДДЕРЖИВАЕТСЯ	ИНСТРУМЕНТАЛИЗАЦИЯ *Аугментация* (human-in-the-loop)	ДЕЛЕГИРОВАНИЕ *Управляемое делегирование* (human-on-the-loop)
	Квадрант II	Квадрант III
ПОДРЫВАЕТСЯ	ПОДЧИНЕНИЕ *Алгоритмическая медиация* (human-shaped-by-the-loop)	ЗАМЕЩЕНИЕ *Автоматизация* *Диффузия ответственности* (human-out-of-the-loop)
	Квадрант I	Квадрант IV

Источник: составлено автором.

Рис. 1. Матрица «автономия–автономность»

Fig. 1. Normative autonomy – technical autonomy matrix

Автономия человека не ограничивается формальным наличием выбора и предполагает способность понимать условия организационной среды, формулировать цели, оценивать последствия и действовать согласно собственным ценностям и интересам.

Автономность ИИ-агента как техническая характеристика системы и автономия человека как антропологическая характеристика субъекта принятия решений относятся к теоретически различным категориям и требуют раздельного анализа. В матрице (рис. 1) показаны четыре типа человеко-агентного взаимодействия. Модель «человек в контуре» предполагает, что решение принимается только человеком. Модель «человек над контуром» (human-on-the-loop) ограничивает автономность ИИ-агента, предоставляя человеку полномочия по отмене и изменению решений. Модель «человек вне контура» (human-out-of-the-loop) соответствует полной автономности ИИ-агента в пределах заданных человеком ограничений. Модель «человек, формируемый контуром» (human-shaped-by-the-loop), предлагаемая в настоящем исследовании, сохраняет за человеком формальное право принятия решения, однако его восприятие, предпочтения и набор альтернатив формируются под влиянием алгоритмической системы.

В модели «человек, формируемый контуром» доминирует алгоритмическая медиация [3, 16, 28], при которой ИИ-агент влияет на внимание, предпочтения, интерпретацию ситуации и набор доступных альтернатив пользователя. В модели «человек вне контура» – механизм автоматизации, при котором ИИ-агент принимает решение самостоятельно, а человек исключается из процесса содержательного контроля.

Теоретическое обоснование этого различия опирается на положения многополярного управления [30] и гуманистической теории организации. В интерпретации Г.Б. Клейнера центральной категорией научного наследия академика Д.С. Львова [31] выступает развитие, понимаемое как движение от человека к человечеству и обратно. В этой логике ИИ-агенты могут усиливать аналитические, организационные и управленческие возможности предприятия, но не должны подменять человека как носителя целей и смыслов, субъекта развития и общественной ответственности. Идеи академика Д.С. Львова интерпретируют взаимодействие человека с ИИ-агентом не как распределение функций и ролей, а как нравственно-экономическую задачу подчинения экономики и цифровых технологий целям развития человека и человечества. Г.Б. Клейнер подчеркивает, что ИИ может оказывать поддержку разным типам лидеров, но не способен их заменить, поскольку именно духовный лидер определяет смысл деятельности коллектива [31].

Гуманистическая теория организации утверждает, что ответственность за смыслообразование и единство коллектива остается исключительной компетенцией человека, поскольку алгоритмическая оптимизация не заменяет ценностных механизмов интеграции [9]. Следовательно, системный менеджмент и гуманистическая теория организации теоретически обосновывают нормативно-этические ограничения автономности ИИ-агента и приоритет профессионального суждения человека. С этой позиции ограничение автономности ИИ-агента рассматривается как этически обоснованное требование, предотвращающее сведение человека к объекту алгоритмического управления.

Допустимый уровень автономности ИИ-агента должен определяться с учетом его влияния на автономию человека. В контексте технодетерминизма несостоятельным является предположение, что низкий уровень автономности ИИ-агента обеспечивает сохранение автономии человека в цифровой среде. Рекомендации ИИ-агента способны постепенно изменять стиль мышления и суждения лица, принимающего решение, усиливая его ориентацию на количественные показатели, снижая внимание к этическому контексту и сужая круг рассматриваемых альтернатив.

В отличие от существующих подходов, формулирующих только общие требования к ответственному ИИ, в исследовании разработаны методологические принципы ответственного делегирования полномочий ИИ-агенту в человекоцентричных гибридных системах поддержки принятия управленческих решений.

Теоретической предпосылкой *принципа синергии ИИ-агента и человека* выступает двойственная природа алгоритмизации [17]. Ценность ИИ определяется способностью усиливать профессиональное суждение человека и качеством совместного управленческого решения. Распространенная упрощенная логика связывает рост алгоритмизации с сокращением участия и контроля со стороны человека, сводя проектирование к выбору между человеком и ИИ-агентом. Повышение автономности ИИ-агента не может выступать самостоятельной целью проектирования гибридной системы. Принцип предполагает, что допустимый уровень автономности ИИ-агента определяется двумя условиями:

1) способностью создавать совместный эффект, при котором алгоритмы усиливают аналитические и когнитивные возможности человека, а человек обеспечивает ценностную интерпретацию;

2) распределением функций, при котором человек и ИИ-агент совместно принимают обоснованные, своевременные и этически приемлемые решения.

При этом человек должен понимать, на каких данных, критериях и логике основано решение ИИ, и иметь возможность изменить, отклонить или оспорить его[11]. В этом контексте ИИ-агент не может рассматриваться как самостоятельный моральный субъект [26].

Способность ИИ-агента самостоятельно анализировать данные и инициировать действия не означает наличия правовых и этических оснований такого делегирования. Риск-ориентированный подход[12] предполагает, что высокая технологическая готовность (зрелость) ИИ-агента сохраняет необходимость его ограничения и контроля. Положение о трансцендентности [32] в теории гуманистического управления, закрепляя за человеком способность определять смысл и общественную значимость деятельности, исключает безусловное делегирование ИИ полномочий, затрагивающих достоинство, права, возможности, цели и смысл деятельности. *Принцип ограниченной автономности* предусматривает этические, правовые, стратегические ограничения, нарушение которых служит основанием для обязательного участия человека в процессе принятия решения.

[11] EUR-Lex (2024) *Regulation* (*EU*) *2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations* (*EC*) *No 300/2008*, (*EU*) *No 167/2013*, (*EU*) *No 168/2013*, (*EU*) *2018/858*, (*EU*) *2018/1139 and* (*EU*) *2019/2144 and Directives 2014/90/EU*, (*EU*) *2016/797 and* (*EU*) *2020/1828* (*Artificial Intelligence Act*) (*Text with EEA relevance*). [online] Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng [Accessed: 30.05.2026].

[12] Там же.

Ответственный контроль требует компетенций и навыков, позволяющих распознавать этические аспекты решений и оценивать их последствия [33]. Нарушения этических требований при использовании ИИ должны рассматриваться как финансовые, правовые и репутационные риски организации. Если участие человека в процедуре принятия решения может основываться на формальном утверждении рекомендаций ИИ-агента, то система мотивации должна поощрять действия сотрудников, направленные на обнаружение ошибок и отказ от решений с недопустимым уровнем риска. *Принцип развития этической компетентности и ответственности* предполагает целенаправленное формирование и совершенствование у лиц, принимающих решение, компетенций, необходимых для взаимодействия с ИИ-агентами, выявления этических рисков, критического анализа результатов и оспаривания рекомендаций алгоритмических систем.

Теория алгоритмического управления свидетельствует о распределенном характере процедуры принятия решений, на которую влияют действия пользователя и разработчика человеко-агентных систем, архитектура алгоритма, качество данных, интерфейс, регламенты и цели внедрения [16, 22]. *Принцип распределенной ответственности* предполагает ее разграничение между участниками проектирования, внедрения и эксплуатации человеко-агентных систем в соответствии с их полномочиями и степенью влияния на управленческое решение.

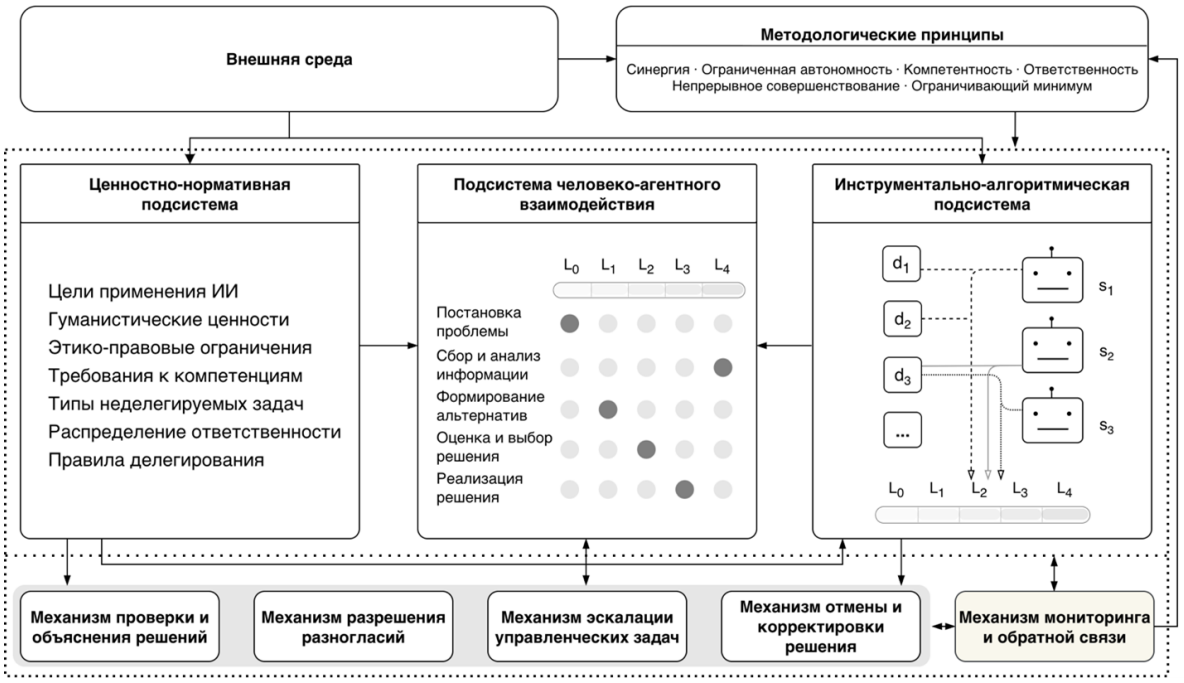
В авторской концепции автономность ИИ-агента рассматривается как организационно-управленческий параметр, определяющий допустимый уровень делегирования полномочий и подлежащий пересмотру при изменении условий организационно-технической среды и решаемых задач. *Принцип непрерывного совершенствования* предусматривает регулярный пересмотр уровня автономности с учетом результатов мониторинга, выявленных инцидентов и случаев оспаривания, а также изменений нормативных требований, характеристик ИИ-системы, компетенций пользователей, состава и параметров управленческих задач.

Высокий риск, необратимость последствий или высокая этическая значимость решения не могут компенсироваться точностью и производительностью ИИ-агента. Например, ИИ может обладать высокой точностью и устойчивостью, но применяться к необратимому или этически чувствительному решению. Использование средней интегральной оценки в таком случае допускает взаимную компенсацию критериев, при которой высокая технологическая зрелость скрывает критический уровень риска. *Принцип ограничивающего минимума* устанавливает, что уровень автономности ИИ-агента определяется по наиболее ограничивающему (минимальному) параметру организационно-технической среды. Подход согласуется с международной нормативно-правовой практикой[13], которая закрепляет риск-ориентированную градацию и обязательность надзора для систем высокого риска независимо от их технического качества. В основе практики лежит предположение о том, что допустимый уровень автономности ИИ-агента определяется приемлемым уровнем риска, а не имеющимися техническими возможностями.

В существующих подходах уровень автономности ИИ-агента определяется через двухвариантную логику, предусматривающую выполнение задачи либо человеком, либо алгоритмической системой. Традиционные модели распределения ролей, полномочий и ответственности, в частности модель RACI, также имеют статический характер и не учитывают участие ИИ-агентов в принятии решений, изменение объема делегированных полномочий и необходимость согласования рекомендаций человека и ИИ-агента.

В развитие исследований, посвященных проблеме диагностики автономии в социотехнических системах [27], рационализации человеко-машинного взаимодействия и анализа экзистенциальных рисков [1, 28], в настоящей работе автономия человека рассматривается в

[13] EUR-Lex (2024) *Regulation* (*EU*) *2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations* (*EC*) *No 300/2008*, (*EU*) *No 167/2013*, (*EU*) *No 168/2013*, (*EU*) *2018/858*, (*EU*) *2018/1139 and* (*EU*) *2019/2144 and Directives 2014/90/EU*, (*EU*) *2016/797 and* (*EU*) *2020/1828* (*Artificial Intelligence Act*) (*Text with EEA relevance*). [online] Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng [Accessed: 30.05.2026]; ISO/IEC (2023) ISO/IEC 42001: Information technology – Artificial intelligence – Management system. [online] Available at: https://cdn.standards.iteh.ai/samples/81230/4c1911ebc9a641f-cb6ee21aa09c28ad3/ISO-IEC-42001-2023.pdf [Accessed 30.05.2026].



Источник: составлено автором.

Рис. 2. Ключевые элементы концептуальной модели взаимодействия человека и ИИ-агента

Fig. 2. Key elements of the conceptual model of human – AI agent interaction

организационно-управленческом контексте и предлагается модель, формализующая распределение полномочий между человеком и ИИ-агентом. Методологические принципы образуют ядро концептуальной модели, которая обеспечивает переход от парадигмы «человек в контуре» к концепции управляемой автономности ИИ-агента и реализуется в трех функционально взаимосвязанных подсистемах (рис. 2).

Ценностно-нормативная подсистема основана на принципах ограниченной автономности, развития этической компетентности и ответственности. Принимая во внимание положения гуманистического управления [32], подсистема задает этические, правовые, стратегические и количественные ограничения допустимых действий (решений) ИИ-агента. Инструментально-алгоритмическая подсистема реализует принципы синергии и ограничивающего минимума, а также определяет допустимый уровень автономности на основе многокритериальной оценки управленческих задач и технологической зрелости (готовности) ИИ-агента. Подсистема человеко-агентного взаимодействия реализует принципы распределенной ответственности и непрерывного совершенствования, устанавливая процедуру мониторинга результатов и пересмотра уровня автономности, объединяя механизмы обоснованного делегирования, возврата полномочий человеку, отмены и оспаривания решений в ситуациях высокой неопределенности, этической значимости или повышенного риска.

Технологическая зрелость (готовность) ИИ-системы может оцениваться по четырем ключевым критериям:

1) критерий точности характеризует долю корректных результатов, формируемых системой;

2) критерий объяснимости характеризует прозрачность логики принятия решений для пользователя;

3) критерий устойчивости определяет способность системы корректно функционировать при изменении качества и типа входных данных;

4) критерий охвата характеризует долю ситуаций определенного класса задач, которые система способна обрабатывать автономно.

Веса критериев не являются универсальными и определяются контекстом применения: для решений с высокими этическими требованиями приоритетным становится критерий объяснимости, для систем критической инфраструктуры – критерий устойчивости, для массовых операционных задач – критерий охвата.

В концептуальной модели автономность ИИ-агента предлагается оценивать с помощью пятиуровневой шкалы:

1) уровень L_0 (информирование) – ИИ-агент осуществляет сбор и анализ данных, решение полностью принимается человеком;

2) уровень L_1 (рекомендация) – ИИ-агент формирует варианты решения, выбор осуществляется человеком;

3) уровень L_2 (утверждение) – ИИ-агент предлагает решение, подлежащее обязательному подтверждению или отклонению со стороны человека;

4) уровень L_3 (условная автономность) – ИИ-агент принимает решения автономно, а полномочия передаются человеку при возникновении аномалий или снижении достоверности оценок системы;

5) уровень L_4 (полная автономность) – ИИ-агент действует автономно в пределах заданных человеком ограничений.

Разработана система критериев «7R» (табл. 1) управленческой задачи для определения уровня автономности ИИ-агента. Критерии объединены в три категории, отражающие условия делегирования. Допустимость, оцениваемая по риску, обратимости и этической значимости, устанавливает нормативные ограничения делегирования и определяет возможность передачи ИИ-агенту полномочий по решению задачи. Реализуемость (формализуемость, оперативность) характеризует когнитивно-технологические условия, при которых ИИ-агент способен своевременно выполнить задачу, если объем информации и требуемая скорость обработки превышают когнитивные возможности человека. Целесообразность (стабильность среды, повторяемость) отражает организационно-экономическую обоснованность делегирования. В соответствии с принципом ограничивающего минимума итоговый уровень автономности определяется наименьшей оценкой среди учитываемых параметров.

Рассматривается управленческая задача d, решаемая с использованием ИИ-агента s. Эксперты оценивают (выбирают) допустимый уровень автономности ИИ-агента по двум группам критериев:

$$l_{ie}^{R}(d) \in \mathcal{L},\ \ l_{je}^{M}(s,d) \in \mathcal{L},\ \ i = 1 \ldots I,\ \ \ j = 1 \ldots J,\ \ \ e = 1 \ldots E,$$

где $l_{je}^{M}(s,d)$ – допустимый уровень автономности ИИ-агента s по критерию технологической зрелости j, определяемый экспертом e; $l_{ie}^{R}(d)$ – допустимый уровень автономности для управленческой задачи d по критерию i, определяемый экспертом e; J – количество критериев зрелости; I – количество критериев управленческой задачи; E – количество экспертов; d – тип управленческой задачи; L – множество уровней автономности L_0–L_4.

Агрегированная экспертная оценка по каждой группе критериев определяется медианой:

$$\tilde{L}_j^M(s,d) = Med_{e=1,\ldots,E}\left\{l_{je}^{M}(s,d)\right\},\ \ \tilde{L}_i^R(d) = Med_{e=1,\ldots,E}\left\{l_{ie}^{R}(d)\right\},$$
$$l_{i(1)}^{R} \le \cdots \le l_{i(E)}^{R},\ \ l_{j(1)}^{M} \le \cdots \le l_{j(E)}^{M},$$

где $\tilde{L}_j^M(s,d)$ – агрегированная оценка уровня автономности по критерию технологической зрелости j ИИ-агента s при решении управленческой задачи d; $\tilde{L}_i^R(d)$ – агрегированная оценка уровня автономности по критерию i управленческой задачи d.

При нечетном количестве экспертов медианной категорией является центральная упорядоченная оценка. При четном количестве экспертов используется нижняя медиана, то есть меньшая из двух центральных оценок. Допустимый уровень автономности ИИ-агента определяется по принципу ограничивающего минимума:

$$L(s,d) = \min\left\{L_R(d), L_M(s,d)\right\}, \quad L_R(d) = \min_{i=1,\ldots,I} \tilde{L}_i^R(d),$$
$$L_M(s,d) = \min_{j=1,\ldots,J} \tilde{L}_j^M(s,d),$$

где $L(s, d)$ – итоговый допустимый уровень автономности ИИ-агента s при решении управленческой задачи d; $L_R(d)$ – допустимый уровень автономности, определяемый критериями управленческой задачи d; $L_M(s, d)$ – допустимый уровень автономности для ИИ-агента s при решении управленческой задачи d.

Согласованность экспертной процедуры может оцениваться отдельно для каждой группы критериев с использованием альфы Криппендорфа. Для каждого критерия рассчитываются доля экспертов, чьи оценки совпадают с медианой, и межквартильный размах.

Механизм отмены решения определяет порядок аннулирования принятого ИИ-агентом решения и предполагает решение следующих основных задач:

1) определение оснований для отмены (например, отклонение от установленных норм и правил, этические соображения, неполнота или недостоверность данных, изменение контекста, мотивированное несогласие ответственного лица с рекомендацией алгоритма);

2) обеспечение простого интерфейса инициирования отмены и обязательное документирование ее причин;

3) организация обратной связи для совершенствования ИИ-агента.

Механизм возврата полномочий человеку (эскалации) обеспечивает сохранение содержательного контроля человека в ситуациях, выходящих за пределы компетенции ИИ-агента, и предполагает решение следующих основных задач:

1) определение сигналов передачи решения человеку (например, снижение уверенности, обнаружение аномалий и этически значимых аспектов решения);

2) классификация уровня сложности решения;

3) определение адресата эскалации (профильный эксперт, руководитель, коллегиальный орган или высшее руководство).

Механизм разрешения разногласий определяет порядок выявления и урегулирования расхождений между оценками человека и ИИ-агента и предполагает решение следующих задач:

1) определение признаков конфликта, например расхождение между рекомендациями ИИ-агента и человека, несогласие с критериями, подозрение на ошибку;

2) установление порядка принятия окончательного решения: например, в рутинных задачах может учитываться алгоритмическая рекомендация, в высокорисковых и этически значимых задачах окончательное решение принимает человек либо коллегиальный орган.

Заключение

Развитие ИИ-агентов влияет на распределение функций, полномочий и ответственности в процессе принятия управленческих решений. Проведенное исследование позволило обосновать необходимость пересмотра традиционной управленческой парадигмы и сформировать концептуальную основу проектирования гибридных систем поддержки принятия решений, в которых за

Таблица 1. Параметры управленческой задачи (система критериев «7R») и градация уровней автономности ИИ-агента
Table 1. Management task parameters (7R criteria system) and AI agent autonomy level gradation

Критерий	Характеристика	Автономное принятие решений (L_4)	Совместное принятие решений (L_2, L_3)	Инструментальная поддержка (L_0–L_1)
Риск (Risk)	Вероятность и масштаб негативных последствий	Низкий (минимальный ущерб, не влияющий на функционирование системы)	Умеренный (локальный ущерб, компенсируемый в рамках операционных ресурсов)	Критический (угроза жизни, значительный финансовый или репутационный ущерб)
Стабильность среды (Regularity)	Степень предсказуемости внешней среды	Стабильная (поведение системы предсказуемо)	Изменчивая (поведение системы частично прогнозируемо)	Хаотичная (поведение системы нельзя спрогнозировать)
Обратимость (Reversibility)	Время и ресурсы, необходимые для отмены ошибочного решения	Высокая (отмена без потери ресурсов)	Средняя (отмена возможна с умеренными затратами времени и ресурсов)	Низкая (решение невозможно отменить)
Формализуемость (Routinisation)	Возможность описания задачи в виде правил и алгоритмов	Формальная (жесткая логика алгоритма)	Гибридная (сочетание алгоритмических процедур и экспертной оценки)	Неформализуемая (требуются интуиция и экспертная оценка)
Оперативность (Responsiveness)	Ограничения по времени на принятие решения	Мгновенная (требуемая скорость превышает когнитивные возможности человека)	Оперативная (решение требуется в сжатые сроки, но допускает верификацию)	Отложенная (отсутствуют жесткие ограничения по времени)
Этичность (Rectitude)	Воздействие решения на этические ценности	Нейтральная (решение не затрагивает этические ценности)	Ограниченная (затрагивает нормативные и репутационные аспекты)	Критическая (решение затрагивает достоинство, права и свободы человека)
Повторяемость (Recurrence)	Частота возникновения задачи	Высокая (задача регулярная, повторяется постоянно)	Средняя (эпизодически повторяется с вариативным контекстом)	Низкая (возникает редко, в уникальных ситуациях)

Источник: составлено автором.

человеком сохраняются ведущая роль в принятии решений и возможность содержательного контроля. Концептуальная модель определяет основания распределения полномочий и предлагает способ снижения рисков алгоритмического управления, описанных в работах [16, 18].

Теоретическая значимость исследования состоит в развитии концептуальных и методологических основ проектирования гибридных систем поддержки принятия управленческих решений. Полученные результаты формируют основу для дальнейшего изучения человекоцентричного алгоритмического управления. Результаты исследования создают предпосылки для внедрения ответственного ИИ в систему управления и разработки регламентов, этических кодексов и политик применения ИИ. Возможность их практического применения требует экспериментальной апробации.

Получены следующие результаты исследования:

1. Обоснован теоретический подход, основанный на разграничении автономности ИИ-агента и автономии человека. Автономность ИИ-агента теоретически определяется как технико-организационное свойство, отражающее степень его самостоятельности в действиях. Автономия человека выступает гуманистической ценностью и основанием для разграничения функциональной самостоятельности ИИ-агента и моральной субъектности человека, за которым закрепляется приоритет в контроле и принятии окончательных решений. Разработана матрица «автономия–автономность», включающая четыре типа человеко-агентного взаимодействия, для которых определены управленческие механизмы сохранения автономии человека.

2. Разработанные методологические принципы конкретизируют модель «человека в контуре» и обеспечивают трансформацию общих установок ответственного ИИ в систему требований к делегированию, распределению ответственности и содержательному контролю. В системе принципов закрепляется концептуальное различие между самостоятельностью ИИ-агента и моральной субъектностью человека, при котором исключительные полномочия по контролю, отмене решений и разрешению разногласий сохраняются за человеком.

3. Разработана концептуальная модель взаимодействия человека и ИИ-агента, в которой делегирование полномочий рассматривается как динамический процесс, адаптируемый к изменениям решаемых задач и организационной среды с учетом установленных ограничений. Модель, в отличие от логики максимальной автономности, позволяет обосновывать состав и ограничения полномочий ИИ-агента, определять степень участия человека в процессе принятия решения, выбирать механизмы контроля и унифицировать правила делегирования для однотипных управленческих задач.

4. Разработаны критерии и процедура экспертного определения допустимого уровня автономности ИИ-агента на основе оценки параметров управленческой задачи и технологической зрелости ИИ-агента. Авторский вклад состоит в формализации экспертной процедуры, основанной на агрегировании оценок и принципе ограничивающего минимума, который исключает компенсацию ограничений высокими значениями других параметров. Разработана шкала автономности ИИ-агента, определяющая степень (уровень) его участия в принятии управленческих решений. Автономность при этом рассматривается как динамический организационно-управленческий параметр, подлежащий корректировке при изменении управленческих задач, рисков, условий организационной среды и технических возможностей ИИ-агента.

Направления дальнейших исследований

Перспективным направлением дальнейших исследований является эмпирическая апробация результатов в различных организационных контекстах, различающихся по типам принимаемых решений и допустимому уровню автономности ИИ-агента. К таким контекстам относятся управление человеческими ресурсами (подбор, оценка и развитие персонала), управление производственными процессами, стратегическое планирование (постановка целей и распределение ресурсов). Сравнительный анализ результатов апробации позволит уточнить ограничения

применимости разработок, оценить практическую ценность и дифференцировать механизмы обеспечения автономии человека применительно к различным типам управленческих решений и условиям организационной среды. Дальнейшего исследования и конкретизации также требуют механизмы, обеспечивающие автономию человека в гибридных системах поддержки принятия решений.

СПИСОК ИСТОЧНИКОВ

1. Krook J. (2025) When autonomy breaks: the hidden existential risk of AI. *AI & Society*, 40, 6011–6024. DOI: 10.1007/s00146-025-02397-5

2. Kaber D. (2026) From Automation to Autonomy Through AI: Enabling and Retaining Human Controllability. In: *Handbook of Human-Centered Artificial Intelligence* (ed. W. Xu), Singapore: Springer, 1173–1216. DOI: 10.1007/978-981-95-3658-0_5

3. Prunkl C. (2024) Human Autonomy at Risk? An Analysis of the Challenges from AI. *Minds and Machines*, 34 (3), art. no. 26. DOI: 10.1007/s11023-024-09665-1

4. Бабкин А.В., Федоров А.А., Либерман И.В., Клачек П.М. (2021) Индустрия 5.0: понятие, формирование и развитие. *Экономика промышленности*, 14 (4), 375–395. DOI: 10.17073/2072-1633-2021-4-375-395

5. Бабкин А.В., Шкарупета Е.В., Плотников В.А. (2021) Интеллектуальная киберсоциальная экосистема Индустрии 5.0: понятие, сущность, модель. *Экономическое возрождение России*, 4 (70), 39–62. DOI: 10.37930/1990-9780-2021-4-70-39-62

6. Bujold A., Roberge-Maltais I., Parent-Rocheleau X., Boasen J., Sénécal S., Léger P.-M. (2024) Responsible artificial intelligence in human resources management: a review of the empirical literature. *AI and Ethics*, 4, 1185–1200. DOI: 10.1007/s43681-023-00325-1

7. Cui T., Tan B., Shi Y. (2024) Fostering humanistic algorithmic management: A process of enacting human-algorithm complementarity. *The Journal of Strategic Information Systems*, 33 (2), art. no. 101838. DOI: 10.1016/j.jsis.2024.101838

8. Rodgers W., Murray J.M., Stefanidis A., Degbey W.Y., Tarba S.Y. (2023) An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. *Human Resource Management Review*, 33 (1), art. no. 100925. DOI: 10.1016/j.hrmr.2022.100925

9. Rosanas J.M. (2011) A humanistic approach to organizations and organizational decision-making. In: *The Future of Leadership Development* (ed. J. Canals), London: Palgrave Macmillan, 143–176. DOI: 10.1057/9780230295087_8

10. Lee M.K., Nigam I., Zhang A., Afriyie J., Qin Z., Gao S. (2021) Participatory algorithmic management: Elicitation methods for worker well-being models. In: *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (*AIES'21*), 715–726. DOI: 10.1145/3461702.3462628

11. Shneiderman B. (2020) Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *arXiv:2002.04087*. DOI: 10.48550/arXiv.2002.04087

12. Facchini A., Mangili F. (2024) Human-Centered AI (Also) for Humanistic Management. In: *Humanism in Marketing: Responsible Leadership and the Human-to-Human Approach* (eds. P. Kotler, W. Pfoertsch, F. Ancarani, I. Ureta), Cham: Palgrave Macmillan, 225–255. DOI: 10.1007/978-3-031-67155-5_11

13. Shao Y., Zope H., Jiang Y., Pei J., Nguyen D., Brynjolfsson E., Yang D. (2025) Future of Work with AI Agents: Auditing Automation and Augmentation Potential across the U.S. Workforce. *arXiv:2506.06576*. DOI: 10.48550/arXiv.2506.06576

14. Gladden M.E. (2019) Who Will Be the Members of Society 5.0? Towards an Anthropology of Technologically Posthumanized Future Societies. *Social Sciences*, 8 (5), art. no. 148. DOI: 10.3390/socsci8050148

15. Calvo R.A., Peters D., Vold K., Ryan R.M. (2020) Supporting human autonomy in AI systems: A framework for ethical enquiry. In: *Ethics of Digital Well-Being: A Multidisciplinary Approach* (eds. C. Burr, L. Floridi), Cham: Springer, 31–54. DOI: 10.1007/978-3-030-50585-1_2

16. Kellogg K.C., Valentine M.A., Christin A. (2020) Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14 (1), 366–410. DOI: 10.5465/annals.2018.0174

17. Meijerink J., Bondarouk T. (2023) The Duality of Algorithmic Management: Toward a Research Agenda on HRM Algorithms, Autonomy and Value Creation. *Human Resource Management Review*, 33 (1), art. no. 100876. DOI: 10.1016/j.hrmr.2021.100876

18. Cobonpue J.S., Wurm B., Hess T. (2024) Fighting back the algorithm: A systematic literature review on algoactivism. *ICIS 2024 Proceedings*, art. no. 16. [online] Available at: https://aisel.aisnet.org/icis2024/lit_review/lit_review/16 [Accessed: 30.05.2026].

19. Aghion P., Bunel S. (2024) *AI and Growth: Where Do We Stand?* [online] Available at: https://www.frbsf.org/wp-content/uploads/AI-and-Growth-Aghion-Bunel.pdf [Accessed 30.05.2026].

20. Лукашевич Н.С. (2025) Человекоцентричная организация: от теоретических моделей к категориально-системной методологии. *Лидерство и менеджмент*, 12 (8), 1699–1714. DOI: 10.18334/lim.12.8.123655

21. Townsend M., Romme A.G.L. (2024) The Emerging Concept of the Human-Centered Organization: A Review and Synthesis of the Literature. *Humanistic Management Journal*, 9 (1), 53–74. DOI: 10.1007/s41463-024-00168-w

22. Köchling A., Wehner M.C. (2020) Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13, 795–848. DOI: 10.1007/s40685-020-00134-w

23. Fenwick A., Molnar G., Frangos P. (2024) Revisiting the role of HR in the age of AI: Bringing humans and machines closer together in the workplace. *Frontiers in Artificial Intelligence*, 6, art. no. 1272823. DOI: 10.3389/frai.2023.1272823

24. Minbaeva D. (2021) Disrupted HR? *Human Resource Management Review*, 31 (4), art. no. 100820. DOI: 10.1016/j.hrmr.2020.100820

25. Zou H.P., Huang W.-C., Wu Y., Chen Y., Miao C., Nguyen H., Zhou Y., Zhang W., Fang L., He L., Li Y., Cao Y., Li D., Jiang R., Yu P.S. (2025) A Survey on Large Language Model Based Human-Agent Systems. *arXiv preprint*. [online] Available at: https://arxiv.org/html/2505.00753v3 [Accessed 30.05.2026].

26. Dyrkolbotn S.K., Pedersen T., Slavkovik M. (2017) Classifying the Autonomy and Morality of Artificial Agents. *CARe-MAS@ PRIMA*, 2017, 67–83.

27. Subías-Beltrán P., Unceta I., de Lecuona I. et al. (2026) Asking the right questions: a governance approach to uphold human autonomy in artificial intelligence. *AI & Society*, 41, 1149–1173. DOI: 10.1007/s00146-025-02611-4

28. Catena E. (2025) AI and human autonomy: A literature review. *AI and Ethics*, 6, art. no. 126. DOI: 10.1007/s43681-025-00958-4

29. Mitchell M., Ghosh A., Luccioni A.S., Pistilli G. (2025) Fully Autonomous AI Agents Should Not Be Developed. *arXiv:2502.02649*. DOI: 10.48550/arXiv.2502.02649

30. Клейнер Г.Б. (2024) Многополярное управление организацией. *Российский журнал менеджмента*, 22 (2), 163–178. DOI: 10.21638/spbu18.2024.201

31. Клейнер Г.Б. (2024) Системный менеджмент: информационная реконструкция для новой экономики. *Стратегическое планирование и развитие предприятий*, 8–18. DOI: 10.34706/978-5-8211-0821-0-8-18

32. Melé D. (2016) Understanding Humanistic Management. *Humanistic Management Journal*, 1 (1), 33–55. DOI: 10.1007/s41463-016-0011-5

33. Işık Ö., Goswami A. (2026) The Three Obstacles Slowing Responsible AI. *MIT Sloan Management Review*, 67 (2), 70–74. DOI: 10.63383/FBhK7635

REFERENCES

1. Krook J. (2025) When autonomy breaks: the hidden existential risk of AI. *AI & Society*, 40, 6011–6024. DOI: 10.1007/s00146-025-02397-5

2. Kaber D. (2026) From Automation to Autonomy Through AI: Enabling and Retaining Human Controllability. In: *Handbook of Human-Centered Artificial Intelligence* (ed. W. Xu), Singapore: Springer, 1173–1216. DOI: 10.1007/978-981-95-3658-0_5

3. Prunkl C. (2024) Human Autonomy at Risk? An Analysis of the Challenges from AI. *Minds and Machines*, 34 (3), art. no. 26. DOI: 10.1007/s11023-024-09665-1

4. Babkin A.V., Fedorov A.A., Liberman I.V., Klachek P.M. (2021) Industry 5.0: concept, formation and development. *Russian Journal of Industrial Economics*, 14 (4), 375–395. DOI: 10.17073/2072-1633-2021-4-375-395

5. Babkin A.V., Shkarupeta E.V., Plotnikov V.A. (2021) Intelligent cyber-social ecosystem of Industry 5.0: definition, essence, model. *Economic Revival of Russia*, 4 (70), 39–62. DOI: 10.37930/1990-9780-2021-4-70-39-62

6. Bujold A., Roberge-Maltais I., Parent-Rocheleau X., Boasen J., Sénécal S., Léger P.-M. (2024) Responsible artificial intelligence in human resources management: a review of the empirical literature. *AI and Ethics*, 4, 1185–1200. DOI: 10.1007/s43681-023-00325-1

7. Cui T., Tan B., Shi Y. (2024) Fostering humanistic algorithmic management: A process of enacting human-algorithm complementarity. *The Journal of Strategic Information Systems*, 33 (2), art. no. 101838. DOI: 10.1016/j.jsis.2024.101838

8. Rodgers W., Murray J.M., Stefanidis A., Degbey W.Y., Tarba S.Y. (2023) An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. *Human Resource Management Review*, 33 (1), art. no. 100925. DOI: 10.1016/j.hrmr.2022.100925

9. Rosanas J.M. (2011) A humanistic approach to organizations and organizational decision-making. In: *The Future of Leadership Development* (ed. J. Canals), London: Palgrave Macmillan, 143–176. DOI: 10.1057/9780230295087_8

10. Lee M.K., Nigam I., Zhang A., Afriyie J., Qin Z., Gao S. (2021) Participatory algorithmic management: Elicitation methods for worker well-being models. In: *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (*AIES'21*), 715–726. DOI: 10.1145/3461702.3462628

11. Shneiderman B. (2020) Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *arXiv:2002.04087*. DOI: 10.48550/arXiv.2002.04087

12. Facchini A., Mangili F. (2024) Human-Centered AI (Also) for Humanistic Management. In: *Humanism in Marketing: Responsible Leadership and the Human-to-Human Approach* (eds. P. Kotler, W. Pfoertsch, F. Ancarani, I. Ureta), Cham: Palgrave Macmillan, 225–255. DOI: 10.1007/978-3-031-67155-5_11

13. Shao Y., Zope H., Jiang Y., Pei J., Nguyen D., Brynjolfsson E., Yang D. (2025) Future of Work with AI Agents: Auditing Automation and Augmentation Potential across the U.S. Workforce. *arXiv:2506.06576*. DOI: 10.48550/arXiv.2506.06576

14. Gladden M.E. (2019) Who Will Be the Members of Society 5.0? Towards an Anthropology of Technologically Posthumanized Future Societies. *Social Sciences*, 8 (5), art. no. 148. DOI: 10.3390/socsci8050148

15. Calvo R.A., Peters D., Vold K., Ryan R.M. (2020) Supporting human autonomy in AI systems: A framework for ethical enquiry. In: *Ethics of Digital Well-Being: A Multidisciplinary Approach* (eds. C. Burr, L. Floridi), Cham: Springer, 31–54. DOI: 10.1007/978-3-030-50585-1_2

16. Kellogg K.C., Valentine M.A., Christin A. (2020) Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14 (1), 366–410. DOI: 10.5465/annals.2018.0174

17. Meijerink J., Bondarouk T. (2023) The Duality of Algorithmic Management: Toward a Research Agenda on HRM Algorithms, Autonomy and Value Creation. *Human Resource Management Review*, 33 (1), art. no. 100876. DOI: 10.1016/j.hrmr.2021.100876

18. Cobonpue J.S., Wurm B., Hess T. (2024) Fighting back the algorithm: A systematic literature review on algoactivism. *ICIS 2024 Proceedings*, art. no. 16. [online] Available at: https://aisel.aisnet.org/icis2024/lit_review/lit_review/16 [Accessed: 30.05.2026].

19. Aghion P., Bunel S. (2024) *AI and Growth: Where Do We Stand?* [online] Available at: https://www.frbsf.org/wp-content/uploads/AI-and-Growth-Aghion-Bunel.pdf [Accessed 30.05.2026].

20. Lukashevich N.S. (2025) Human-centered organization: from theoretical models to categorical-system methodology. *Leadership and Management*, 12 (8), 1699–1714. DOI: 10.18334/lim.12.8.123655

21. Townsend M., Romme A.G.L. (2024) The Emerging Concept of the Human-Centered Organization: A Review and Synthesis of the Literature. *Humanistic Management Journal*, 9 (1), 53–74. DOI: 10.1007/s41463-024-00168-w

22. Köchling A., Wehner M.C. (2020) Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13, 795–848. DOI: 10.1007/s40685-020-00134-w

23. Fenwick A., Molnar G., Frangos P. (2024) Revisiting the role of HR in the age of AI: Bringing humans and machines closer together in the workplace. *Frontiers in Artificial Intelligence*, 6, art. no. 1272823. DOI: 10.3389/frai.2023.1272823

24. Minbaeva D. (2021) Disrupted HR? *Human Resource Management Review*, 31 (4), art. no. 100820. DOI: 10.1016/j.hrmr.2020.100820

25. Zou H.P., Huang W.-C., Wu Y., Chen Y., Miao C., Nguyen H., Zhou Y., Zhang W., Fang L., He L., Li Y., Cao Y., Li D., Jiang R., Yu P.S. (2025) A Survey on Large Language Model Based Human-Agent Systems. *arXiv:2505.00753v3*. [online] Available at: https://arxiv.org/html/2505.00753v3 [Accessed 30.05.2026].

26. Dyrkolbotn S.K., Pedersen T., Slavkovik M. (2017) Classifying the Autonomy and Morality of Artificial Agents. *CARe-MAS@ PRIMA*, 2017, 67–83.

27. Subías-Beltrán P., Unceta I., de Lecuona I., Pujol O. (2026) Asking the right questions: a governance approach to uphold human autonomy in artificial intelligence. *AI & Society*, 41, 1149–1173. DOI: 10.1007/s00146-025-02611-4

28. Catena E. (2025) AI and human autonomy: A literature review. *AI and Ethics*, 6, art. no. 126. DOI: 10.1007/s43681-025-00958-4

29. Mitchell M., Ghosh A., Luccioni A.S., Pistilli G. (2025) Fully Autonomous AI Agents Should Not Be Developed. *arXiv:2502.02649*. DOI: 10.48550/arXiv.2502.02649

30. Kleiner G.B. (2024) Multipolar management of organization. *Russian Management Journal*, 22 (2), 163–178. DOI: 10.21638/spbu18.2024.201

31. Kleiner G.B. (2024) Sistemnyi menedzhment: informacionnaia rekonstrukciia dlia novoi ekonomiki [Systems Management: Information Reconstruction for the New Economy]. *Strategicheskoe planirovanie i razvitie predpriiatii* [*Strategic Planning and Enterprise Development*], 8–18. DOI: 10.34706/978-5-8211-0821-0-8-18

32. Melé D. (2016) Understanding Humanistic Management. *Humanistic Management Journal*, 1 (1), 33–55. DOI: 10.1007/s41463-016-0011-5

33. Işık Ö., Goswami A. (2026) The Three Obstacles Slowing Responsible AI. *MIT Sloan Management Review*, 67 (2), 70–74. DOI: 10.63383/FBhK7635

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

ЛУКАШЕВИЧ Никита Сергеевич
E-mail: lukashevich@spbstu.ru
Nikita S. LUKASHEVICH
E-mail: lukashevich@spbstu.ru
ORCID: https://orcid.org/0000-0001-5719-4844

Поступила: 04.06.2026; Одобрена: 12.10.2026; Принята: 12.10.2026.
Submitted: 04.06.2026; Approved: 12.10.2026; Accepted: 12.10.2026.