Galactica Media: Journal of Media Studies. 2026. No 4 | ISSN: 2658-7734
AI, Technology & Culture | https://doi.org/10.46539/gmd.v8i4.722

Human Agency under Conditions of Algorithmic Mediation by Artificial Intelligence

Daria V. Chirva*

National Research University Higher School of Economics (HSE University); ITMO University (ITMO). Saint Petersburg, Russia. Email: dashachirva[at]yandex.ru ORCID https://orcid.org/0000-0001-7804-8803

* — corresponding author

Received: 20 December 2025 | Revised: 5 February 2026 | Accepted: 17 February 2026

Abstract

The widespread integration of artificial intelligence systems into everyday and professional practices challenges classical philosophical conceptions of agency, autonomy, and responsibility. Under conditions of algorithmic mediation, human action is increasingly shaped within hybrid sociotechnical systems in which cognitive, epistemic, and normative functions are distributed between humans and AI. This situation calls for a reexamination of which properties of agency remain distinctively human and how they manifest within this new configuration of human–AI interaction. The aim of this article is to provide a philosophical analysis of the specific properties of human agency in hybrid human–artificial intelligence systems and to identify how intentionality and control are transformed under algorithmic mediation. The article demonstrates that human agency in AI-mediated sociotechnical systems is characterized by weakened epistemic autonomy, contextually distributed control, and a hybrid form of intentionality that incorporates algorithmically generated elements. The article identifies several distinctive features, including vulnerability to algorithmic framing and cognitive codependence on AI systems, that distinguish hybrid agency from classical instrumental models of human action. The article concludes that, despite these transformations, humans retain a unique role in normative reflection and meaning-making. This role enables them to critically evaluate and reconfigure the architectures that structure interaction with AI. The article is intended for philosophers, media scholars, digital technology researchers, specialists in AI ethics, and others engaged in analyzing sociotechnical systems and questions of responsibility in algorithmically mediated social realities.

Keywords

Human Agency; Algorithmic Mediation; Artificial Intelligence; Sociotechnical Systems; Hybrid Agency; Distributed Control; Intentionality; Epistemic Autonomy; Normative Reflection

This work is licensed under a Creative Commons “Attribution” 4.0 International License
© 2026 Daria V. Chirva

Galactica Media: Journal of Media Studies. 2026. № 4 | ISSN: 2658-7734
ИИ, технологии и культура | https://doi.org/10.46539/gmd.v8i4.722

Человеческая агентность в условиях алгоритмического посредничества искусственного интеллекта

Чирва Дарья Викторовна*

Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ); Национальный исследовательский университет ИТМО (ИТМО). Санкт-Петербург, Россия.
Email: dashachirva[at]yandex.ru ORCID https://orcid.org/0000-0001-7804-8803

* — корреспондирующий автор

Рукопись получена: 20 декабря 2025 | Пересмотрена: 5 февраля 2026 | Принята: 17 февраля 2026

Аннотация

Широкое внедрение систем искусственного интеллекта в повседневные и профессиональные практики приводит к трансформации форм человеческого действия и ставит под вопрос классические философские представления об агентности, автономии и ответственности. В условиях алгоритмического посредничества человеческое действие всё чаще формируется в рамках гибридных социотехнических систем, в которых когнитивные, эпистемические и нормативные функции распределены между человеком и ИИ. Это требует переосмысления того, какие свойства агентности остаются специфичными для человека и каким образом они проявляются в новой конфигурации взаимодействия. Целью статьи является философский анализ специфических свойств человеческой агентности в гибридных системах «человек — искусственный интеллект», а также выявление изменений в человеческой интенциональности, контроле и познании в условиях алгоритмического посредничества. В работе показано, что человеческая агентность в социотехнических системах, насыщенных ИИ, характеризуется ослабленной эпистемической автономией, контекстуально распределенным контролем и гибридной формой интенциональности, включающей алгоритмически сгенерированные элементы. Делается вывод о том, что человек в гибридных системах сохраняет уникальную роль носителя нормативной и смысловой рефлексии, позволяющей оценивать и трансформировать саму архитектуру взаимодействия с ИИ. Статья адресована философам, исследователям медиа и цифровых технологий, специалистам в области этики ИИ, а также всем, кто исследует социотехнические системы и проблемы ответственности в условиях алгоритмизации социальной реальности.

Ключевые слова

человеческая агентность; алгоритмическое посредничество; искусственный интеллект; социотехнические системы; гибридная агентность; распределенный контроль; интенциональность; эпистемическая автономия; нормативная рефлексия

Это произведение доступно по лицензии Creative Commons “Attribution” («Атрибуция») 4.0
© 2026 Чирва Дарья Викторовна

Введение

Широкое распространение систем искусственного интеллекта (далее — ИИ), включенных в повседневные и профессиональные практики, радикально изменяет конфигурацию взаимодействия человека и технологии. Современные социотехнические системы (от алгоритмических платформ до автономных сервисов принятия решений) обладают собственными когнитивными и операционными возможностями, которые влияют на человеческое познание, его выбор, нормы поведения и формы социального порядка. В этих условиях человек как единственный центр агентности уходит в прошлое: направление его действий совместно определяется человеком и алгоритмическими процессами, а сами действия дополняются или ограничиваются алгоритмами, которые формируют новый тип распределенного взаимодействия.

Однако изменение структуры взаимодействия человека и ИИ не отменяет необходимости анализа собственно человеческой агентности в новых условиях. Напротив, оно требует более точного понимания того, какие ее свойства сохраняются, какие трансформируются и какие новые формы ее проявления могут быть зафиксированы. Традиционные философские модели действия основаны на идеях автономии, рационального контроля и стабильности субъекта. На наш взгляд, они оказываются недостаточными для описания ситуаций, в которых действие является результатом совместного функционирования человека и алгоритмических компонентов ИИ, встроенных в инфраструктуру принятия решений.

Возникает необходимость выделить специфические свойства человеческой агентности в социотехнических системах с учетом ее ситуативности, обусловленности параметрами цифровой среды, включенности в сетевые процессы и способности к рефлексивной адаптации. Такое исследование важно как для этики в сфере ИИ, так и для уточнения теории ответственности, поскольку понимание природы человеческого действия определяет принципы распределения ответственности в среде, насыщенной ИИ.

Настоящая статья предлагает философский анализ цифровой среды, насыщенной ИИ, и рассматривает человеческую агентность не в оппозиции к ИИ, а как элемент гибридной системы. В работе рассматриваются теоретический контекст и методологические основания описания человеческого действия в новой социотехнической конфигурации.

Под социотехнической системой в статье подразумевается изменчивое, исторически обусловленное соединение социальных практик, норм, институтов, человеческого участия и материально-технических артефактов, которое образует структуру распределенной агентности и в котором техническое и социальное взаимно определяют формы действия, восприятия и ответственности.

Классические и современные модели агентности

Анализ свойств человеческой агентности в социотехнических системах, в которых человек действует совместно с системами искусственного интеллекта, предполагает опору на философские и междисциплинарные теории действия. В рамках классического подхода агентность связывается с фигурой индивидуального, автономного, рационального субъекта, способного к интенциональному действию, контролю над собственными импульсами и последовательной реализации планов. Однако появление распределенных систем познания, цифровых инфраструктур и адаптивных алгоритмов трансформировало контексты, в которых человеческая агентность проявляется и оценивается.

Задача выявления специфических свойств человеческой агентности в гибридных системах «человек — ИИ-инфраструктура» предполагает обращение к классическим моделям интенционального действия, современным концепциям расширенного и распределенного разума, а также к социотехническим подходам, рассматривающим агентность как результат сетевых взаимодействий.

Данный раздел включает две основные теоретические линии:

- классическую теорию интенции и объяснения действия;
- когнитивные теории расширенного и распределенного разума.

Вместе они образуют теоретическую рамку, позволяющую впоследствии выявить уникальные, неустранимые элементы человеческой агентности во всё более развитых и непрозрачных системах ИИ.

Классическая теория интенции и объяснения действия

Работа Элизабет Энском «Интенция» является одним из классических оснований современной философии действия (Anscombe, 1957). Действие трактуется не как физическое событие, а как явление, структурированное вопросом «зачем?». Интенциональность здесь выступает не как внутреннее психическое состояние, а как принцип, в соответствии с которым действие поддается пониманию в терминах практической рациональности. Агент действует, потому что способен соотносить собственные движения с определенной целью, знанием и практической ситуацией. Данная трактовка важна для понимания роли человека в гибридных системах: интенциональность не сводится к вычислительной обработке информации или оптимизации задачи, а выражает способность к смыслополаганию, формированию «практической перспективы», в рамках которой цель может быть осмыслена и подвергнута критике.

Дональд Дэвидсон дополнил этот подход положением, согласно которому действия объясняются причинами, то есть убеждениями и желаниями (Davidson, 1980). Рациональность у Дэвидсона не является абсолютной, однако

она структурирует объяснение действия. Если действие можно объяснить в терминах «он сделал это, потому что считал X и желал Y», такое действие является интенциональным.

Модель агента Дэвидсона рассматривает человека как источник причин собственного действия. Даже если причинные связи сложны, они сохраняют внутреннюю согласованность, обеспечивая возможность приписывания ответственности. Сходную концепцию автономии человеческого действия развивает Гарри Франкфурт. Он различает желания первого порядка (непосредственные импульсы) и желания второго порядка (желания относительно желаний). Автономия означает способность агента идентифицироваться со своими желаниями, а не просто испытывать их. Такое самоприсвоение желаний — ключевой аспект человеческой свободы воли и моральной ответственности.

В своей теории действия М. Братман учитывает, что значительное число действий человек осуществляет не в одиночку, а совместно с другими людьми (Bratman, 1980). В результате мы имеем дело не просто с суммой отдельных действий, выполняемых одновременно, а с их совокупностью, предполагающей особую координацию на уровне общего намерения (shared intention). Намерения участников взаимосвязаны и взаимозависимы: каждый участник осведомлен о намерениях других — тем самым формируется основа для взаимных ожиданий. Каждый из участников имеет намерение совершить совместное действие. Постулирование коллективного субъекта с едиными намерениями и действиями при этом не требуется. Агентность как способность действовать понимается Братманом как способность формировать и выполнять планы, а совместная агентность — как способность формировать и выполнять планы *совместно*. Именно акцент на планировании и ориентация на длительное совместное действие делают данную теорию не вполне подходящей для объяснения человеческой агентности в системах с технологическим компонентом. Наличие намерений трудно обоснованно приписать ИИ; кроме того, во многих случаях он ориентирован на немедленное влияние на поведение человека посредством алгоритмически обусловленного управления вниманием и действием. Симметричное планирование совместной деятельности вряд ли возможно при взаимодействии с ИИ.

Так или иначе, все эти теории подчеркивают, что автономия как специфическая характеристика человека-агента заключается в его способности поддерживать устойчивую архитектуру действия.

Насколько рассмотренные классические модели человеческого действия соответствуют современным ситуациям взаимодействия человека с ИИ-системами? Во-первых, значительная часть информации поступает к человеку посредством алгоритмической фильтрации. При этом изменяется не только содержание доступной информации, но и сама структура доступных ему целей. Во-вторых, ИИ может инициировать события, формирующие поведение

человека, при этом сам человек может не осознавать их влияния на собственную мотивацию. В таких случаях не всегда понятно, на каком уровне формируется интенция и где проходит граница между «моим желанием» и «предложенным системой вариантом поведения». Таким образом, классическая модель интенциональности с ее индивидуалистской основой оказывается под давлением новых условий, требующих рассмотрения агентности в контексте распределенных влияний и изменчивой структуры действия.

Теории расширенного и распределенного разума

Однако первоначально критике подвергается классическая теория агентности со стороны исследований процессов познания, а не принятия решений. Э. Кларк и Д. Чалмерс предложили радикальную идею: когнитивные процессы человека могут включать элементы внешней среды (Clark & Chalmers, 1998). Их знаменитый пример с записной книжкой Отто описывает ситуацию, в которой внешний объект выполняет ту же когнитивную функцию, что и внутренняя память. При этом характер используемой памяти — внешней или внутренней — не влияет на результат: если ее обладатель стремится достичь определенной цели, он достигает ее независимо от того, насколько автономна его когнитивная система.

В контексте исследуемого вопроса о человеческой агентности особый интерес представляет концепция Э. Хатчинса (Hutchins, 1995). В своих исследованиях морской навигации он показал, что когнитивные функции при решении сложных задач могут распределяться между людьми и артефактами. При этом система «человек — инструмент — инфраструктура» функционирует как единое целое. Представление о познании как о распределенном процессе позволяет описать не только решение человеком сложных профессиональных задач, но и его повседневную деятельность, значительная часть которой осуществляется в обусловленной ИИ цифровой среде, предназначенной для досуга, общения, игр, обучения и удовлетворения потребностей.

В современном цифровом мире внешние когнитивные структуры представлены не только простыми инструментами, но и предсказательными алгоритмами, цифровыми профилями, персонализированным доступом к информации и адаптивными интерфейсами (Зубофф, 2022). Современные сервисы на основе ИИ уже не просто хранят или обрабатывают данные. Они формируют контекст человеческих когнитивных операций. ИИ, особенно в форме больших языковых моделей, начиная с осени 2022 г. становится «когнитивным партнером», способным выполнять операции, которые ранее составляли неотъемлемую часть человеческого рассуждения. Информационные потоки перераспределяются между компонентами системы, изменяя конфигурацию познавательной активности. Границы индивидуального разума становятся более проницаемыми. Если когнитивные процессы распределены, то агентность становится функцией конфигурации системы, а не внутренним свой-

ством индивида. Это не отменяет человеческую агентность, но делает ее зависимой от качества взаимодействия человека с технологической средой.

Опираясь на опыт работы в исследовательском центре компании Xerox, Л. Сачмен в своей книге критикует идею о том, что действия полностью определяются внутренними планами (Сачмен, 2019). Она показывает, что действие — это результат постоянного взаимодействия человека с материальным и социальным контекстом окружающего мира. Интерфейс, навигационная система или система автоматической диагностики становятся участниками действия, определяющими, какие шаги возможны, а какие — нет. В условиях ИИ данная зависимость проявляется особенно отчетливо: алгоритмические системы создают динамический контекст, в котором человек вынужден переосмысливать свои интенции в процессе действия.

Сущность современной трансформации агентности еще до широкого распространения ИИ анализировалась в терминах акторно-сетевой теории (ANT) Б. Латура (2014). Он определяет агентность вне антропоцентрической перспективы как распределенную способность действовать, возникающую из сети взаимодействий между человеческими и нечеловеческими акторами. Агентность не связана с сущностными характеристиками; она производится как эффект выстраиваемых связей: «подсоединения первичны, акторы вторичны» (Латур, 2014, с. 303). Артефакты, наряду с людьми, обладают агентным воздействием, поскольку они определяют поведение, создают нормы, формируют связи, обеспечивают условия взаимодействия и задают последовательности событий. Интенциональность как сущностное свойство агента в классической теории утрачивает определяющее значение. ИИ — показательный пример такого актора. Мы не можем приписать ему интенциональность, однако он обладает реальной силой, способной определять траекторию действий человека.

Обзор классических и современных моделей агентности позволяет увидеть, что появление систем искусственного интеллекта не отменяет фундаментальных характеристик человеческого действия, но радикально меняет контекст, в котором оно реализуется. Классические философы действия (Энском, Дэвидсон) подчеркивали, что агентность опирается на интенциональность и связана со способностью объяснять действия в терминах причин, осознаваемых субъектом и принадлежащих ему самому. Теории расширенного и распределенного разума показали, что многие когнитивные функции фактически выходят за пределы индивидуального мозга, а акторно-сетевой подход сформировал понятийную рамку для рассмотрения агентности вне исключительно антропологического контекста, отказа от понимания интенциональности как ее сущностной характеристики и перехода к сетевой парадигме.

Анализ человеческого действия в современной цифровой среде, обусловленной ИИ, позволяет с учетом трансформации представлений об агентности обозначить «слабые места» классического понимания человеческой агент-

ности. Во-первых, когнитивные и практические функции, ранее считавшиеся исключительно человеческими, частично распределяются между человеческими и искусственными компонентами. Таким образом, границы между субъектом и технологической средой становятся более проницаемыми. Во-вторых, структура действия всё чаще определяется конфигурациями инфраструктур, а не только внутренними планами человека. В-третьих, интенции человека формируются в среде, в которой алгоритмы активно участвуют в производстве предпочтений, фокусировании внимания и формировании возможностей для действия (Floridi, 2011, 2013).

В этой ситуации вопрос о специфически человеческих свойствах агентности не может сводиться к идентификации качеств, которые ИИ пока не способен воспроизвести, поскольку развиваются как сама технология, так и способы ее интеграции в человеческую деятельность. Нас интересует не только поиск аспектов человеческой агентности, не затронутых ИИ-инфраструктурой, но и анализ ее реконфигурации в современных условиях. Далее мы последовательно рассмотрим эти два аспекта человеческой агентности.

Человеческая агентность в эпоху алгоритмов

Теории расширенного и распределенного разума показали, что познавательная деятельность не всегда осуществляется исключительно как «внутренний» процесс (Clark, 2008). Однако именно человек остается тем звеном, которое связывает различные когнитивные ресурсы в единое смысловое пространство. ИИ способен выполнять вычисления, делать прогнозы, формировать рекомендации, но он не является субъектом смыслополагания и постановки вопросов. Потребность знать, сомневаться и искать невозможно делегировать ИИ-агенту или группе ИИ-агентов. Вопрос всегда возникает из живого практического контекста существования человека, независимо от того, идет ли речь о разработке ученым дизайна эксперимента или о планировании отпуска с семьей. К решению этих задач можно привлекать ИИ-агентов, которые во многих случаях способны действовать быстрее и эффективнее человека (Gunkel, 2012). Однако возникает вопрос: что приведет их в действие, если не запрос, обусловленный определенной жизненной ситуацией? Предлагаемое ИИ-агентом решение человек может принять или отвергнуть, поскольку осознает собственное исходное смыслополагание. Все перечисленное может быть обобщено в понятии смысловой интеграции. Смысловая интеграция представляет собой ключевое специфическое свойство человеческой агентности, не сводимое ни к рациональному планированию, ни к обработке данных. Она предполагает способность видеть себя элементом совместного действия, осуществляемого коллективом агентов, и вместе с тем выходить за его пределы, охватывать конфигурацию целиком, оценивать и преобразовывать ее. ИИ-системы в настоящее время действуют в рамках заданных критериев оптимизации и не обладают способностью к «переосмыс-

лению» самой структуры, в которой они функционируют, и к рефлексии (Leiser, 2025).

Однако обратной стороной гибридизации процессов решения эпистемических задач становится утрата полной эпистемической автономии (Bucher, 2018). Традиционные философские представления — от картезианской модели познающего субъекта до постклассической концепции рационального агента Дэвидсона — предполагают, что субъект способен в той или иной степени контролировать свои основания для веры и действия. Даже доверяя другому, субъект знает, почему это делает; следуя рекомендациям эксперта, он понимает схему их обоснования.

В гибридных системах «человек-ИИ» эта модель нарушается. Современные системы машинного обучения, особенно системы, основанные на глубинных нейронных сетях, предоставляют человеку информацию и рекомендации, которые эпистемически непрозрачны: внутренние основания этих результатов либо трудно объяснить, либо они принципиально недоступны для человеческого понимания в силу масштаба моделей и сложности зависимостей (Eubanks, 2018). Сложность таких моделей превышает возможности их полного когнитивного освоения человеком.

Непрозрачность приводит к тому, что человек принимает решения, которые не может обосновать с опорой на традиционные процедуры рационального обоснования. Он действует на основании доверия к модели, а не на основании осознанной эпистемической позиции. Возникает своеобразная децентрализованная форма эпистемического действия: человек формирует убеждения через посредника, который сам не обладает убеждениями в философском смысле.

Ослабленная эпистемическая автономия не означает, что человек превращается в пассивный элемент. Как отмечалось выше, смысловая интеграция является сущностной характеристикой человеческой агентности, включенной в гибридную систему. Скорее, человек оказывается в положении агента, вынужденного распределять эпистемическое доверие между собственной позицией, системой ИИ и организационными структурами, ответственными за ее использование. Следовательно, эпистемическая автономия реализуется в сетевом, а не индивидуальном формате.

Смысловая интеграция позволяет человеку как адаптироваться к контексту технологически опосредованных решений, например повышая уровень ИИ-грамотности для более эффективного взаимодействия с ИИ, так и изменять этот контекст. Человек способен выходить за рамки заданных алгоритмических структур и формировать новые структуры. Так, исследовательская деятельность инженеров приводит к созданию новых ИИ-архитектур. Человек может противостоять алгоритмическим сценариям, однако речь идет не о внешнем, наивном сопротивлении системе (это соответствовало бы инструментальному подходу к ИИ, в рамках которого агентностью в принципе

обладает только человек). Алгоритмы опираются на предположение о некоторой инертности и стабильности предпочтений, поскольку исходят из представления об эффективности и комфорте как о ключевых характеристиках человеческого опыта, поддающихся параметризации. Человек же может отключать персонализацию в разных приложениях, если замечает, что рекомендации ИИ начинают формировать его опыт. Тем самым становится возможным различение «моих» намерений и статистически сгенерированных предложений. Человек может действовать непоследовательно или иррационально с точки зрения ИИ-системы, выводя свою цифровую активность за рамки ИИ-оптимизации. Так, намеренный выбор на маркетплейсе предложения, не являющегося оптимальным с точки зрения соотношения «цена — качество» или «цена — срок доставки», идет вразрез с «ожиданиями» алгоритмов. В гибридных системах, включающих ИИ, эта способность становится особенно значимой. ИИ может создавать контекст и задавать возможности действия, но не может самостоятельно пересмотреть саму структуру взаимодействия. Алгоритмы оптимизируют заданную структуру, но не пересматривают ее; адаптируются, но не поднимаются на уровень метаанализа.

В гибридных системах контроль над действием перестает быть единым и локализованным. В классических моделях человеческой агентности, от Энском до Франкфурта, контроль является необходимым условием для того, чтобы действие считалось принадлежащим субъекту. Контроль выражается в способности агента направлять свои поступки в соответствии со своими желаниями и намерениями.

Однако в гибридных системах контроль оказывается распределенным по уровням. Человек задает цель: например, получить оптимальную стратегию инвестирования или быстро обработать большой массив данных. ИИ осуществляет оптимизацию, преобразуя цель в операционные шаги, которые человек не может полностью отследить. Алгоритмическая инфраструктура ограничивает пространство решений, предлагая готовые варианты. Организация или разработчики контролируют параметры модели, обновления, ограничения и допустимые операции. В результате контроль становится многоуровневой архитектурой, в которой человеческий вклад формирует общее направление, но детали его осуществления оказываются вынесены за пределы индивидуального управления.

Это не означает утрату агентности; скорее, речь идет о переходе к контекстуально распределенному контролю, при котором действие формируется совокупностью факторов, каждый из которых является необходимым, но ни один не является достаточным. Человеческая агентность, таким образом, приобретает координирующий, а не директивный характер.

В рамках координирующей агентности человека реализуется его способность к нормативной рефлексии. Несмотря на то, что ИИ способен формировать предпочтения пользователей, изменять их информационную среду

и создавать новые траектории поведения, человек остается единственным компонентом системы, способным к оценке собственных желаний, планов и действий — к оценке второго порядка.

Идея Г. Франкфурта о желаниях второго порядка здесь сохраняет свою значимость (Frankfurt, 1969). Человек *способен* не только испытывать предпочтения, но и интерпретировать их происхождение, критически оценивать, являются ли эти желания результатом осознанного выбора или следствием алгоритмических манипуляций, а также изменять свои мотивационные структуры, сопротивляясь цифровому влиянию или принимая его.

Именно способность оставаться субъектом собственных действий в ситуации, когда многие мотивации подвергаются внешнему воздействию, имеет фундаментальное значение для понимания специфически человеческого элемента по своей сути гибридной системы.

ИИ, каким бы автономным он ни был в операциональном смысле, не обладает способностью к нормативной рефлексии (Hildebrandt, 2015). Формы такой рефлексии алгоритмически задаются в нем людьми для более полного соответствия его поведения нормативным ожиданиям человека.

Таким образом, человеческая агентность сохраняет уникальные свойства именно в сфере нормативной саморефлексии: человек может не только действовать, но и оценивать ценность, допустимость и смысл как собственных действий, так и всей системы, частью которой он является.

В настоящее время речь идет прежде всего о гибридных системах «человек — ИИ» как о социотехнических конфигурациях, в которых человеческая агентность и агентность алгоритмических систем переплетены и совместно конституируют практики и решения. В таких системах ИИ не выступает только как нейтральный инструмент; он становится активным медиатором, участником когнитивных процессов и носителем процедурных норм (Hoskins, 2017). Если в классическом инструментализме человек действует, опираясь на собственное знание и рассуждение, то в гибридной системе эпистемическая база действия частично переносится в область данных и моделей (Cheney-Lippold, 2017). Решения базируются на выводах модели, которые не всегда совпадают с человеческой интерпретацией ситуации. Более того, алгоритмы могут адаптироваться, обучаться и менять свое поведение в ответ на действия человека и изменения среды. В результате взаимодействие становится двунаправленным и эмерджентным: поведение системы не редуцируется к сумме намерений участников. В гибридной агентности границы становятся размытыми: какие действия можно приписать человеку, а какие — системе, зависит от протоколов взаимодействия, видимости вмешательства и институционального контекста (Хейлз, 2025). Однако не всегда удается адекватно реконструировать функциональное распределение агентности между акторами системы. Так, в инструменталистской парадигме ответственность возлагается на пользователя системы,

поскольку именно он определяет цели и оценивает последствия ее использования. В гибридной парадигме проектировщики системы, владельцы данных, операторы и регуляторы, сама система ИИ и ее пользователи — все выступают источниками нормативной значимости, поскольку они формируют архитектуру, метрики и условия обратной связи (List & Pettit, 2011).

Однако одной из наиболее значимых, на наш взгляд, черт гибридизации человеческой деятельности является трансформация фундаментальной характеристики действия — интенциональности. Классические концепции интенциональности предполагают, что действие агента определяется его собственными намерениями. Энском подчеркивает, что действие интенционально, если оно может быть объяснено в терминах «для чего» субъект делает то, что делает. Дэвидсон связывает интенциональность с причинными отношениями между убеждениями, желаниями и поступками. Но в гибридных системах человек действует посредством ИИ. Это приводит к тому, что интенциональность не исчезает, но расширяется и включает результаты, которые человек не мог бы произвести самостоятельно. Можно утверждать, что действие в гибридной формации включает:

- человеческое намерение («я хочу оптимизировать маршрут»);
- алгоритмическое решение («модель выбрала вариант маршрута»);
- человеческое принятие этого решения («я следую этому маршруту»).

Таким образом, интенциональность приобретает составной характер и включает человеческие и нечеловеческие компоненты. Человек не просто использует инструмент. Он делегирует ему часть интенциональной структуры. В результате человеческое действие теперь может включать элементы, не являющиеся продуктом человеческого разума, но тем не менее входящие в структуру поступка.

Вместо исключительно человеческой интенциональности мы получаем гибридную интенциональность, в которой человеческие намерения «встраиваются» в алгоритмически сгенерированные планы и решения. Это требует пересмотра традиционных понятий ответственности, свободы воли и агентного статуса человека.

Заключение

Проведенный анализ показал, что распространение систем искусственного интеллекта приводит не к исчезновению человеческой агентности, а к ее глубокой реконфигурации в рамках гибридных социотехнических систем. Классические философские модели агентности, основанные на интенциональности, автономии и индивидуальном контроле, сохраняют эвристическую ценность, однако оказываются недостаточными для описания действий, осуществляемых в условиях алгоритмического посредничества. Человеческое

действие всё чаще формируется как результат взаимодействия с непрозрачными, адаптивными и нормативно нагруженными ИИ-инфраструктурами, что требует расширения понятийного аппарата философии действия.

В статье показано, что специфические свойства человеческой агентности в гибридных системах проявляются не в исключительном контроле или эпистемической самодостаточности, а в способности к смысловой интеграции, нормативной рефлексии и метаоценке конфигурации действия в целом. Ослабление эпистемической автономии и распределение контроля между человеком и ИИ не означают утраты агентности, а переводят ее в контекстуально распределенный режим, в котором человеческий вклад заключается в постановке целей, интерпретации результатов и критической оценке источников собственных предпочтений и решений.

Особое значение приобретает переопределение интенциональности: человеческие намерения встраиваются в алгоритмически сгенерированные планы и решения, образуя гибридную интенциональную структуру. Это позволяет говорить о специфической форме человеческой агентности, не сводимой ни к инструментальному использованию технологии, ни к симметричной совместной агентности человека и ИИ. Человек остается единственным элементом системы, способным к нормативному суждению второго порядка, пересмотру собственных мотиваций и осмысленному сопротивлению алгоритмически заданным рамкам.

Таким образом, отказ от жесткого инструментализма и антропоцентризма позволяет более адекватно описать реальность современных социотехнических систем. Человеческая агентность в них проявляется как динамическая, уязвимая, но принципиально рефлексивная форма участия в распределенных процессах осуществления действий и принятия решений.

Заявления авторов

Согласие с политиками и положениями сетевого издания

Автор данной рукописи был уведомлен и согласен с действующими политиками и положениями сетевого издания, изложенными на веб-сайте.

Оригинальность и эксклюзивность рукописи

Автор подтверждает, что рукопись является оригинальной, ранее не публиковалась ни полностью, ни частично и не находится на рассмотрении для публикации в других изданиях.

Авторство и вклад (CRediT)

Проведение исследования, концептуализация и написание статьи — Чирва Д. В.



Политика в отношении препринтов и предыдущих версий

Автор подтверждает наличие препринта/рабочей версии.

Конфликт интересов

Конфликтов нет.

Финансирование

В статье представлены результаты исследований по проекту РНФ 25-18-00208 «Экзистенциальный опыт в цифровой среде: “бытие к цифре”, онтология виртуального и человеческое Я», выполненных в НИУ ВШЭ в 2025 г.

The article presents the results of research conducted within the framework of the Russian Science Foundation project No. 25-18-00208, “Existential Experience in the Digital Environment: ‘Being-toward-the-Digital,’ the Ontology of the Virtual, and the Human Self,” carried out at HSE University in 2025.

Использование генеративного ИИ и вспомогательных инструментов

Инструменты искусственного интеллекта не указаны в качестве авторов; любой вклад (редактирование текста, перевод, иллюстрации, код) описан в разделе «Методы»; точность, оригинальность и соблюдение прав проверены.

Разрешения на использование (тексты, изображения, карты, иллюстрации)

Все материалы, защищённые авторским правом, воспроизводятся с разрешения правообладателя; соблюдаются условия лицензии (включая Creative Commons).

Благодарности и вклады без авторства

Автор данной работы выражает благодарность проф. И. Б. Микиртумову за поддержку в ходе проведения исследования.

Список литературы

Anscombe, G. E. M. (1957). *Intention*. Cornell University Press.

Bratman, M. E. (1993). Shared Intention. *Ethics*, 104(1), 97–113. https://doi.org/10.1086/293577

Bucher, T. (2018). *If ... Then: Algorithmic Power and Politics*. Oxford University Press. https://doi.org/10.1093/oso/9780190493028.001.0001

Cheney-Lippold, J. (2017). *We are Data: Algorithms and the Making of Our Digital Selves*. New York University Press.

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195333213.001.0001

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Davidson, D. (1980). *Essays on Actions and Events*. Oxford University Press.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. Macmillan + ORM.

Floridi, L. (2011). *The Philosophy of Information*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199232383.001.0001

Floridi, L. (2013). *The Ethics of Information*. OUP Oxford. https://doi.org/10.1093/acprof:oso/9780199641321.001.0001

Frankfurt, H. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, 66(23), 829–839. https://doi.org/10.2307/2023833

Gunkel, D. J. (2012). *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. MIT Press. https://doi.org/10.7551/mitpress/8975.001.0001

Hildebrandt, M. (2015). *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*. Edward Elgar Publishing. https://doi.org/10.4337/9781849808774

Hoskins, A. (2017). *Digital Memory Studies: Media Pasts in Transition*. Routledge. https://doi.org/10.4324/9781315637235

Hutchins, E. (1995). *Cognition in the Wild*. MIT Press. https://doi.org/10.7551/mitpress/1881.001.0001

Leiser, D. (2025). *The Inscrutability of AI Systems – A Constructivist Perspective* (SSRN Scholarly Paper No. 5296298). Social Science Research Network. https://doi.org/10.2139/ssrn.5296298

List, C., & Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press.

Зубофф, Ш. (2022). *Эпоха надзорного капитализма: Борьба за будущее человечества на новых рубежах власти*. Издательство Института Гайдара.

Латур, Б. (2014). *Пересборка социального: Введение в акторно-сетевую теорию*. Издательский дом Высшей школы экономики.

Сачмен, Л. (2019). *Реконфигурации отношений человек—Машина: Планы и ситуативные действия*. Элементарные формы.

Хейлз, Н. К. (2025). *Как мы стали постлюдьми: Виртуальные тела в кибернетике, литературе и информатике*. V-A-C press.

References

Anscombe, G. E. M. (1957). *Intention*. Cornell University Press.

Bratman, M. E. (1993). Shared Intention. *Ethics*, 104(1), 97–113. https://doi.org/10.1086/293577

Bucher, T. (2018). *If ... Then: Algorithmic Power and Politics*. Oxford University Press. https://doi.org/10.1093/oso/9780190493028.001.0001

Cheney-Lippold, J. (2017). *We are Data: Algorithms and the Making of Our Digital Selves*. New York University Press.

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195333213.001.0001

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Davidson, D. (1980). *Essays on Actions and Events*. Oxford University Press.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. Macmillan + ORM.

Floridi, L. (2011). *The Philosophy of Information*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199232383.001.0001

Floridi, L. (2013). *The Ethics of Information*. OUP Oxford. https://doi.org/10.1093/acprof:oso/9780199641321.001.0001

Frankfurt, H. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, 66(23), 829–839. https://doi.org/10.2307/2023833

Gunkel, D. J. (2012). *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. MIT Press. https://doi.org/10.7551/mitpress/8975.001.0001

Hayles, N. K. (2025). *How we Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. V-A-C Press. (In Russian).

Hildebrandt, M. (2015). *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*. Edward Elgar Publishing. https://doi.org/10.4337/9781849808774

Hoskins, A. (2017). *Digital Memory Studies: Media Pasts in Transition*. Routledge. https://doi.org/10.4324/9781315637235

Hutchins, E. (1995). *Cognition in the Wild*. MIT Press. https://doi.org/10.7551/mitpress/1881.001.0001

Latour, B. (2014). *Reassembling the Social: An Introduction to Actor-Network-Theory*. HSE Press. (In Russian).

Leiser, D. (2025). *The Inscrutability of AI Systems – A Constructivist Perspective* (SSRN Scholarly Paper No. 5296298). Social Science Research Network. https://doi.org/10.2139/ssrn.5296298

List, C., & Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press.

Suchman, L. (2019). *Human-Machine Reconfigurations: Plans and Situated Actions*. Elementarnye formy. (In Russian).

Zuboff, S. (2022). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Gaidar Institut Press. (In Russian).