

Влияние искусственного интеллекта на чувство субъектности при принятии решений



Таова Виктория Михайловна

аспирант кафедры психологии личности в системах управления ФОиР ВШГУ, Президентская академия, Москва, Российская Федерация
e-mail: victoria.taova@yandex.ru
ORCID ID: 0009-0003-0202-0582



Сидячева Наталья Владимировна

кандидат психологических наук, доцент, заведующий кафедрой социальной и педагогической психологии, ФГАОУ ВО «Государственный университет просвещения», Москва, Российская Федерация
e-mail: sidna@bk.ru
SPIN-код РИНЦ: 8064-3142
ORCID ID: 0000-0001-6454-964X

Аннотация

Статья посвящена анализу влияния искусственного интеллекта на чувство субъектности личности в процессах принятия решений. Цель – осмысление трансформации чувства субъектности в условиях взаимодействия человека и искусственного интеллекта и выявление факторов, влияющих на сохранение или ослабление переживания авторства. Методы исследования: теоретический анализ, систематизация и обобщение отечественных и зарубежных исследований в области психологии субъектности, когнитивной психологии и взаимодействия человека и цифровых технологий. Показано, что искусственный интеллект перераспределяет когнитивные функции, изменяя восприятие контроля и ответственности; проанализированы эффекты чрезмерного доверия алгоритму, снижения рефлексивности и размывания ответственности. Установлено, что степень трансформации субъектности зависит от прозрачности алгоритмов, объяснимости решений и возможности вмешательства пользователя. Теоретическая значимость исследования состоит в уточнении представлений о распределенной субъектности в цифровой среде, практическая – в обосновании принципов проектирования систем искусственного интеллекта, поддерживающих человеческую субъектность.

Ключевые слова

• субъектность • чувство субъектности • взаимодействие человека и цифровых технологий • искусственный интеллект • распределенная субъектность • принятие решений • ответственность •

Введение

Развитие искусственного интеллекта (ИИ) в последние годы привело к существенным изменениям в структуре взаимодействия человека и цифровых технологий. Интеллектуальные системы постепенно переходят от роли пассивных инструментов к роли активных участников когнитивной деятельности, способных анализировать информацию, формулировать рекомендации, прогнозировать результаты и участвовать в выработке решений [19; 20]. Особенно заметно это проявляется в системах навигации, управленческой аналитики, генеративных языковых моделях, медицинских экспертных системах и цифровых ассистентах.

Подобные изменения актуализируют вопрос о трансформации человеческой субъектности в условиях взаимодействия с ИИ. Если в традиционной модели управления технология выступала средством реализации намерений субъекта, то современные интеллектуальные системы начинают участвовать в формировании самого решения. В результате возникает ситуация распределенного принятия решений, при которой авторство действия субъективно разделяется между человеком и системой [8].

Ключевым психологическим понятием в данном контексте является чувство субъектности (sense of agency), или переживание себя как автора собственных действий и их последствий [24]. Классические исследования показывают, что чувство субъектности формируется на основе согласованности между намерением, действием и результатом [35]. Однако включение ИИ в процесс принятия решений нарушает данную последовательность: часть аналитических и оценочных функций передается системе, вследствие чего снижается определенность границ собственного вклада человека в итоговое решение [6].

Современные исследования взаимодействия человека и цифровых технологий демонстрируют, что интеллектуальные системы способны восприниматься пользователями как относительно автономные субъекты взаимодействия [34]. Особенно выраженным данный эффект становится в случаях, когда система демонстрирует признаки «намеренности»: уверенность формулировок, адаптивность поведения, способность объяснять действия или поддерживать диалог [23]. Это приводит к парадоксальной ситуации: чем более интеллектуальной и автономной воспринимается система, тем выше вероятность ослабления субъективного чувства контроля у пользователя.

Проблематика субъектности в системах с искусственным интеллектом имеет не только теоретическое, но и практическое значение. Ослабление чувства авторства решения может приводить к эффекту чрезмерного доверия алгоритму (over-reliance), снижению критичности мышления и размыванию ответственности [27]. В организационном и управленческом контексте это создает риски принятия решений без достаточной рефлексивной оценки их оснований.

Несмотря на возрастающее количество исследований в области взаимодействия человека и цифровых технологий, проблема трансформации субъектности в условиях взаимодействия с интеллектуальными системами остается недостаточно концептуализированной в психологической литературе. В частности,

требуется уточнение механизмов распределенной субъектности, факторов сохранения чувства контроля и роли дизайна взаимодействия в поддержке субъектной позиции личности.

Цель статьи заключается в теоретическом анализе влияния искусственного интеллекта на чувство субъектности личности в процессах принятия решений.

Для достижения поставленной цели необходимо:

- проанализировать теоретические основания анализа субъектности в системах с искусственным интеллектом;
- рассмотреть процессы трансформации субъектности при взаимодействии с искусственным интеллектом;
- выявить факторы, усиливающие или ослабляющие чувство субъектности пользователя.

Теоретические основания анализа субъектности в системах с искусственным интеллектом

Понятие субъектности традиционно занимает центральное место в отечественной психологии личности и деятельности. В работах С.Л. Рубинштейна, А.Н. Леонтьева и К.А. Абульхановой субъект рассматривался как активный источник деятельности, способный к целеполаганию, саморегуляции и осознанному контролю собственных действий [1; 3]. Современные отечественные исследования развивают понимание субъектности как системного качества, проявляющегося в способности к самодетерминации и авторству собственной активности [4]. В зарубежной психологии аналогичную функцию выполняет концепция *sense of agency*, описывающая субъективное переживание авторства действия [24].

Одной из наиболее влиятельных моделей объяснения субъектности является компараторная теория, согласно которой чувство контроля возникает в результате сопоставления ожидаемого и фактического результата действия [16]. Если прогнозируемый результат совпадает с реальным, действие воспринимается как собственное. Нарушение данной согласованности приводит к ослаблению чувства авторства.

Альтернативный подход представлен в теории кажущейся ментальной каузальности Д. Вегнера [35]. Согласно данной концепции, чувство субъектности является не непосредственным отражением контроля, а когнитивной интерпретацией причинной связи между намерением и действием. Человек воспринимает себя автором поступка, если мысль о действии предшествует ему, соответствует ему и исключает наличие других очевидных причин. В контексте взаимодействия с ИИ это особенно важно, поскольку интеллектуальная система может выступать как альтернативный источник действия и решения.

Современные исследования демонстрируют, что чувство субъектности имеет многоуровневую структуру и включает как сенсомоторный, так и рефлексивный компоненты [33]. Первый связан с непосредственным ощущением контроля над действием, второй — с осознанием ответственности и авторства. В системах с ИИ трансформация происходит преимущественно на рефлексивном уровне: пользователь продолжает инициировать действие, однако начинает воспринимать систему как значимого участника принятия решения.

Особое значение для анализа взаимодействия человека и цифровых технологий приобретает концепция распределенной субъектности. В рамках данного подхода субъектность рассматривается не как свойство исключительно отдельного индивида, а как результат взаимодействия человека, технологий и среды [13]. Подобная логика восходит к теориям распределенного познания Э. Хатчинса и акторно-сетевой теории Б. Латура, в рамках которых технические объекты рассматриваются как элементы сети, влияющие на структуру действия [21].

В условиях цифровизации субъектность все чаще приобретает кооперативный характер. Решение формируется в результате совместной работы человека и системы, при этом вклад каждого участника зачастую оказывается трудно определить. В литературе данный феномен описывается как совместная субъектность (shared agency) или распределенная субъектность (distributed agency) [8].

Исследования последних лет показывают, что восприятие субъектности, или агентности ИИ зависит не от объективной автономности системы, а от особенностей её поведения [13]. Пользователи склонны приписывать системе признаки субъекта при наличии следующих характеристик:

- адаптивности поведения,
- последовательности действий,
- способности к объяснению решений,
- использования естественного языка,
- демонстрации уверенности в ответах.

В результате возникает эффект социальной субъектности, при котором взаимодействие с ИИ начинает восприниматься не как работа с инструментом, а как коммуникация с партнером [34].

С психологической точки зрения данная трансформация представляет собой изменение структуры контроля. Если в классической модели субъект полностью связывает действие с собственным намерением, то в системах с ИИ происходит перераспределение когнитивных функций. Анализ информации, ранжирование альтернатив и прогнозирование последствий частично передаются алгоритму, тогда как за человеком сохраняется функция окончательного выбора [20].

Подобное перераспределение имеет двойственный характер. С одной стороны, снижается когнитивная нагрузка и возрастает эффективность принятия решений. С другой стороны, уменьшается степень субъективной вовлеченности человека в процесс выбора, что может ослаблять чувство авторства и ответственности.

Трансформация субъектности при взаимодействии с искусственным интеллектом

Одним из ключевых последствий внедрения ИИ в процессы принятия решений становится изменение субъективного восприятия контроля над действием. В отличие от традиционных цифровых инструментов интеллектуальные системы способны не только выполнять команды, но и участвовать в формировании выбора. Это создает ситуацию совместного конструирования решения, в которой пользователь перестает быть единственным источником действия [19].

Исследования взаимодействия человека и цифровых технологий показывают, что даже минимальное участие системы в формировании альтернатив способно влиять на восприятие авторства решения [6]. Если ИИ предлагает вариант действия, ранжирует информацию или формулирует рекомендацию, пользователь начинает воспринимать итоговый результат как совместный продукт взаимодействия.

Особенно отчетливо данный эффект проявляется при использовании генеративных языковых моделей. Пользователь нередко принимает предложенные формулировки с минимальными изменениями, вследствие чего размывается граница между собственным и сгенерированным содержанием. Это приводит к появлению феномена соавторства, при котором субъективное ощущение принадлежности результата становится неопределенным [25].

Важным проявлением трансформации субъектности является эффект чрезмерного доверия алгоритмической системе (over-reliance) [9]. Данный феномен связан с тенденцией воспринимать рекомендации ИИ как более объективные и

рациональные по сравнению с человеческими суждениями. В результате пользователь снижает уровень критической проверки информации и реже анализирует альтернативные варианты. Экспериментальные исследования показывают, что предоставление пользователю объяснений работы системы способно снижать уровень доверия к ней, однако этот эффект достигается лишь при условии, что объяснение действительно требует когнитивного вовлечения, а не воспринимается пассивно [33].

Исследования в области автоматизации принятия решений показывают, что высокая точность системы парадоксальным образом может усиливать когнитивную пассивность пользователя [7]. При длительном взаимодействии с эффективным алгоритмом формируется установка на доверие по умолчанию, вследствие чего человек начинает воспринимать ИИ как более компетентного участника процесса. Введение специальных «когнитивных стопоров» (cognitive forcing functions), побуждающих пользователя к самостоятельному анализу перед обращением к рекомендации системы, рассматривается как один из способов противодействия данной тенденции [9].

В организационном контексте это приводит к смещению локуса контроля. Руководитель или специалист формально сохраняет право окончательного решения, однако психологически ответственность начинает частично переноситься на систему [30]. Особенно заметно данный эффект проявляется в ситуациях высокой неопределенности и информационной перегрузки.

Отдельного внимания заслуживает проблема снижения рефлексивности. Передача части аналитических функций ИИ уменьшает необходимость самостоятельного анализа оснований выбора. Пользователь концентрируется на оценке готового результата, а не на процессе его формирования. Вследствие этого ослабляется метакогнитивный контроль и сокращается глубина осмысления решения [2].

В исследованиях когнитивной автоматизации данный эффект описывается как снижение внимательности и критического мониторинга при высокой автоматизации процессов (automation complacency) [27]. Человек начинает воспринимать систему как надежный источник решений и постепенно утрачивает привычку к самостоятельной проверке информации.

Дополнительной проблемой становится размывание ответственности. Если решение принимается в условиях взаимодействия человека и ИИ, возникает неопределенность относительно того, кто является автором этого решения [15]. В случае ошибки ответственность может субъективно распределяться между пользователем и системой, что снижает уровень личной вовлеченности.

Подобная ситуация особенно значима для сфер, связанных с высокими социальными рисками: медицины, управления персоналом, судебной аналитики, военных технологий. Использование алгоритмических рекомендаций в данных областях способно создавать иллюзию объективности, хотя сама система обучается на данных, содержащих ограничения и когнитивные искажения [26].

Вместе с тем современные исследования позволяют рассматривать происходящие изменения не как утрату субъектности, а как ее трансформацию. В условиях цифровой среды формируется новая модель распределенной субъектности, в которой контроль и авторство становятся результатом взаимодействия человека и технологии [8].

С этой точки зрения ИИ выступает не заменой субъекта, а когнитивным посредником, изменяющим структуру принятия решений. Человеческая субъектность сохраняется, однако приобретает иной характер: от автономного контроля происходит переход к координации и управлению совместным процессом действия.

Факторы, усиливающие или ослабляющие чувство субъектности

Степень трансформации субъектности при взаимодействии с ИИ во многом определяется особенностями дизайна системы и организацией взаимодействия пользователя с алгоритмом. Современные исследования показывают, что чувство субъектности не является фиксированной характеристикой, а зависит от условий деятельности и формы интерфейса [6].

Одним из наиболее значимых факторов выступает прозрачность работы системы. Если пользователь понимает логику формирования рекомендаций, чувство контроля сохраняется на более высоком уровне [31]. Напротив, непрозрачные алгоритмы, функционирующие как «черный ящик», усиливают ощущение отчуждения и снижают субъективную вовлеченность в принятие решения.

Важную роль играет объяснимость решений (explainability). Исследования демонстрируют, что возможность получения объяснения повышает доверие к системе без существенного снижения чувства авторства [12]. Пользователь воспринимает себя не пассивным исполнителем рекомендации, а активным участником анализа.

Существенным фактором также является возможность вмешательства в работу системы. Если ИИ допускает изменение параметров, корректировку рекомендаций или отказ от предложенного варианта, субъект сохраняет ощущение контроля над итоговым результатом [29]. Отсутствие подобной возможности усиливает переживание зависимости от алгоритма.

Значительное влияние оказывает и уровень антропоморфизма системы. Чем более «человекоподобным» представляется ИИ, тем выше вероятность его восприятия как самостоятельного субъекта [13]. Использование естественного языка, эмоциональных реакций, персонализированной коммуникации и социальных сигналов способствует формированию эффекта социальной субъектности.

При этом антропоморфизм ИИ имеет двойственный характер. С одной стороны, он повышает комфорт взаимодействия и доверие к системе. С другой стороны, усиливает тенденцию к приписыванию ИИ намерений, компетентности и ответственности, что может снижать чувство субъектности пользователя [34].

Отдельное значение имеет стиль представления рекомендаций. Системы, формулирующие ответы категорично и уверенно, воспринимаются как более компетентные, вследствие чего возрастает вероятность некритического принятия решений [23]. Напротив, указание уровня неопределенности и ограничений модели способствует сохранению критического отношения пользователя.

Важным психологическим фактором выступает уровень цифровой компетентности личности. Пользователи, обладающие более высоким пониманием принципов работы алгоритмов, демонстрируют меньшую склонность к чрезмерному доверию и сохраняют более устойчивое чувство контроля [22].

Кроме того, значимую роль играет контекст деятельности. В ситуациях высокой когнитивной нагрузки, дефицита времени или неопределенности люди чаще делегируют принятие решений ИИ [7]. Соответственно, вероятность ослабления субъектности возрастает в условиях стресса и информационной перегрузки.

Современные подходы к проектированию ИИ-систем все чаще ориентируются на концепцию человеко-ориентированного ИИ, предполагающую сохранение центральной роли человека в принятии решений [32]. В рамках данного подхода эффективность системы рассматривается не только через точность алгоритма, но и через степень поддержки человеческой автономии, ответственности и рефлексивности.

Заключение

Проведенный нами анализ отечественных и зарубежных исследований показывает, что внедрение ИИ приводит к качественному изменению структуры субъектности в системах принятия решений. Полученные данные позволяют говорить не о вытеснении человека из процесса управления, а о переходе к распределенной модели субъектности, в которой контроль, ответственность и авторство становятся результатом взаимодействия человека и технологии.

Сопоставление классических исследований субъектности с современными исследованиями взаимодействия человека и цифровых технологий позволяет сделать вывод о смещении акцента от индивидуального контроля к кооперативной модели действия. Если в традиционных концепциях субъектность связывалась преимущественно с автономным инициированием действия, то в цифровой среде существенную роль начинают играть внешние когнитивные посредники.

Вместе с тем результаты исследований демонстрируют противоречивый характер данного процесса. С одной стороны, ИИ снижает когнитивную нагрузку, повышает скорость анализа информации и расширяет возможности принятия решений. С другой стороны, чрезмерная автоматизация способна приводить к ослаблению рефлексивности и снижению чувства ответственности.

Особенно значимым представляется парадокс доверия к ИИ: повышение интеллектуальности и автономности системы одновременно увеличивает ее полезность и риск снижения человеческой субъектности. Чем более компетентным воспринимается алгоритм, тем выше вероятность делегирования ему когнитивного контроля.

Данный вывод имеет принципиальное значение для проектирования интеллектуальных систем. Традиционный подход, ориентированный исключительно на эффективность и точность алгоритма, оказывается недостаточным. Не менее важным становится учет психологических последствий взаимодействия, включая сохранение чувства контроля и авторства у пользователя.

В этой связи особое значение приобретает концепция объяснимого ИИ. Возможность понимания логики работы системы позволяет снизить риск отчуждения и поддерживать активную субъектную позицию личности. Аналогичным образом возможность вмешательства и корректировки решений способствует сохранению ответственности за результат.

В целом анализ современных исследований показывает, что взаимодействуя с интеллектуальными системами, человек начинает иначе воспринимать необходимость контроля и ответственности и переживать авторство действия. Можно говорить, что включение ИИ в процессы принятия решений приводит к существенной трансформации человеческой субъектности. Это выражается в том, что субъектность становится по существу распределенной: пользователь сохраняет участие в принятии решения, однако часть когнитивных функций и субъективного контроля переносится на систему, и в результате формируется модель совместной субъектности, в рамках которой решение воспринимается как продукт взаимодействия человека и технологии.

Ключевыми последствиями трансформации субъектности выступают чрезмерное доверие алгоритму, снижение рефлексивности, смещение локуса контроля вовне и размывание ответственности. При этом все эти эффекты нельзя однозначно назвать негативными, так как одновременно снижается когнитивная нагрузка и расширяются аналитические возможности пользователя.

При этом степень ослабления или сохранения субъектности зависит от особенностей дизайна взаимодействия человека и интеллектуальной системы. Поддержанию у пользователя чувства контроля и ответственности способствуют прозрачность алгоритмов, объяснимость решений, возможность вмешательства и ограничение избыточного антропоморфизма.

Таким образом, современное развитие искусственного интеллекта требует пересмотра традиционного понимания субъектности как исключительно индивидуального свойства. В цифровой среде субъектность все чаще формируется как распределенная система взаимодействия человека и технологии.

Практическое значение полученных выводов связано с необходимостью разработки человекоцентричных систем ИИ, ориентированных не только на эффективность, но и на поддержку человеческой автономии и рефлексивности. Перспективы дальнейших исследований связаны с анализом долгосрочного влияния ИИ на личностную регуляцию, структуру ответственности и механизмы формирования авторства в условиях цифрового взаимодействия.

Список литературы:

1. Абульханова К.А. Стратегия жизни. М.: Мысль, 1991. 299 с.
2. Знаков В.В. Субъектно-аналитический подход в психологии понимания // Психологические исследования. 2015. Т. 8. № 42. С. 12. DOI: 10.54359/ps.v8i42.536 EDN: TDXHCY
3. Рубинштейн С.Л. Основы общей психологии. СПб.: Питер, 2019. 713 с.
4. Сергиенко Е.А. Системно-субъектный подход: обоснование и перспектива // Психологический журнал. 2011. Т. 32. № 1. С. 120-132. EDN: NSIPQT
5. A meta-analysis of factors influencing the development of trust in automation / K.E. Schaefer [et al.] // Human Factors. 2016. Vol. 58. N 3. P. 377-400. DOI: 10.1177/0018720816634228
6. Automation technology and sense of control: A window on human agency / B. Berberian [et al.] // PLoS ONE. 2012. Vol. 7. N 3. URL: <https://pubmed.ncbi.nlm.nih.gov/22479528/> (дата обращения: 20.03.2026). DOI: 10.1371/journal.pone.0034075
7. Bainbridge L. Ironies of automation // Automatica. 1983. Vol. 19. N 6. P. 775-779. DOI: 10.1016/0005-1098(83)90046-8
8. Bratman M.E. Shared Agency: A Planning Theory of Acting Together. Oxford: Oxford University Press, 2014. 208 p.
9. Buçinca Z., Malaya M.B., Gajos K.Z. To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making // Proceedings of the ACM on Human-Computer Interaction. 2021. Vol. 5. CSCW1. Article 188. P. 1-21. DOI: 10.1145/3449287
10. Coeckelbergh M. AI Ethics. Cambridge, MA: MIT Press, 2020. 248 p.
11. De Jaegher H., Di Paolo E. Participatory sense-making // Phenomenology and the Cognitive Sciences. 2007. Vol. 6. N 4. P. 485-507. DOI: 10.1007/s11097-007-9076-9 EDN: VSAWVN
12. Ehsan U., Riedl M.O. Human-centered Explainable AI: Towards a Reflective Sociotechnical Approach // HCI International 2020 – Late Breaking Papers: Multimodality and Intelligence. Cham: Springer, 2020. P. 449-466. DOI: 10.1007/978-3-030-60117-1_33
13. Epley N., Waytz A., Cacioppo J.T. On seeing human: A three-factor theory of anthropomorphism // Psychological Review. 2007. Vol. 114. N 4. P. 864-886. DOI: 10.1037/0033-295X.114.4.864
14. Explanations can reduce overreliance on AI systems during decision-making / H. Vasconcelos [et al.] // Proceedings of the ACM on Human-Computer Interaction. 2023. Vol. 7. P. 1-38. DOI: 10.1145/3579605
15. Floridi L., Cowls J. A Unified Framework of Five Principles for AI in Society // Harvard Data Science Review. 2019. Vol. 1. N 1. DOI: 10.1162/99608f92.8cd550d1
16. Frith C.D., Blakemore S.-J., Wolpert D.M. Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action // Brain Research Reviews. 2000. Vol. 31. N 2-3. P. 357-363. DOI: 10.1016/S0165-0173(99)00052-1 EDN: AETHXL

17. Guidelines for Human-AI Interaction / S. Amershi [et al.] // Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019. Paper 3. P. 1-13. DOI: 10.1145/3290605.3300233
18. Haggard P. Sense of agency in the human brain // Nature Reviews Neuroscience. 2017. Vol. 18. N 4. P. 197-207. DOI: 10.1038/nrn.2017.14.
19. Hybrid Intelligence / D. Dellermann [et al.] // Business & Information Systems Engineering. 2019. Vol. 61. N 5. P. 637-643. DOI: 10.1007/s12599-019-00595-2 EDN: ZEGZJW
20. Jarrahi M.H. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making // Business Horizons. 2018. Vol. 61. N 4. P. 577-586. DOI: 10.1016/j.bushor.2018.03.007
21. Latour B. Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford: Oxford University Press, 2005. 320 p. DOI: 10.1093/oso/9780199256044.001.0001
22. Long D., Magerko B. What is AI Literacy? Competencies and Design Considerations // Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020. P. 1-16. DOI: 10.1145/3313831.3376727
23. Miller T. Explanation in artificial intelligence: Insights from the social sciences // Artificial Intelligence. 2019. Vol. 267. P. 1-38. DOI: 10.1016/j.artint.2018.07.007
24. Moore J.W. What is the sense of agency and why does it matter? // Frontiers in Psychology. 2016. Vol. 7. URL: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2016.01272/full> (дата обращения: 20.03.2026). DOI: 10.3389/fpsyg.2016.01272
25. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?/ E.M. Bender [et al.] // Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021. P. 610-623. DOI: 10.1145/3442188.3445922.
26. O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown Publishing Group, 2016. 259 p.
27. Parasuraman R., Manzey D.H. Complacency and bias in human use of automation: An attentional integration // Human Factors. 2010. Vol. 52. N 3. P. 381-410. DOI: 10.1177/0018720810376055
28. Parasuraman R., Riley V. Humans and automation: Use, misuse, disuse, abuse // Human Factors. 1997. Vol. 39. N 2. P. 230-253. DOI: 10.1518/001872097778543886
29. Power to the people: The role of humans in interactive machine learning / S. Amershi [et al.] // AI Magazine. 2014. Vol. 35. N 4. P. 105-120. DOI: 10.1609/aimag.v35i4.2513
30. Rahwan I. Society-in-the-loop: Programming the algorithmic social contract // Ethics and Information Technology. 2018. Vol. 20. N 1. P. 5-14. DOI: 10.1007/s10676-017-9430-8 EDN: MCMADR
31. Rai A. Explainable AI: From black box to glass box // Journal of the Academy of Marketing Science. 2020. Vol. 48. N 1. P. 137-141. DOI: 10.1007/s11747-019-00710-5 EDN: MNBEIB
32. Shneiderman B. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy // International Journal of Human-Computer Interaction. 2020. Vol. 36. N 6. P. 495-504. DOI: 10.1080/10447318.2020.1741118 EDN: ENPGDI
33. Synofzik M., Vosgerau G., Newen A. Beyond the comparator model: A multifactorial two-step account of agency // Consciousness and Cognition. 2008. Vol. 17. N 1. P. 219-239. DOI: 10.1016/j.concog.2007.03.010
34. Waytz A., Heafner J., Epley N. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle // Journal of Experimental Social Psychology. 2014. Vol. 52. P. 113-117. DOI: 10.1016/j.jesp.2014.01.005
35. Wegner D.M. The Illusion of Conscious Will. Cambridge, MA: MIT Press, 2002. 419 p. DOI: 10.7551/mitpress/3650.001.0001

The influence of artificial intelligence on the sense of agency in decision-making

Taova Victoria Mikhailovna

Postgraduate Student of Department of Personality Psychology in Management Systems, FEDM GSPM, Presidential Academy, Moscow, Russian Federation

e-mail: victoria.taova@yandex.ru

ORCID ID: 0009-0003-0202-0582

Sidyacheva Natalia Vladimirovna

Candidate of Psychological Sciences, Associate Professor, Head of the Department of Social and Pedagogical Psychology, Federal State Budgetary Educational Institution of Higher Education, Moscow, Russian Federation

e-mail: sidna@bk.ru

RSCI SPIN code: 8064-3142

ORCID ID: 0000-0001-6454-964X

Abstract

This article analyzes the impact of artificial intelligence on the human sense of agency during decision-making processes. The study aims to conceptualize the transformation of the sense of agency within human-AI interaction and to identify the factors that contribute to either preserving or diminishing the experience of authorship. The methodological framework encompasses theoretical analysis, systematization, and synthesis of domestic and international research in the psychology of agency, cognitive psychology, and human-computer interaction. The findings demonstrate that artificial intelligence redistributes cognitive functions, thereby altering the perception of control and accountability. Furthermore, the article analyzes the effects of automation bias, reduced reflexivity, and the erosion of accountability. The study establishes that the degree of agency transformation depends on algorithmic transparency, the explainability of decisions, and the opportunity for user intervention. The theoretical significance of this research lies in refining the conceptual understanding of distributed agency within digital environments, while its practical value stems from establishing design principles for artificial intelligence systems that support human agency.

Keywords

• agency • sense of agency • human-computer interaction • artificial intelligence • distributed agency • decision-making • accountability •

References:

1. Abul'khanova K.A. Life Strategy. Moscow: Mysl Publishers, 1991. 299 p.
2. Znakov V.V. The Subject-Analytic Approach in the Psychology of Understanding // Psychological Studies. 2015. Vol. 8. N 42. P. 12. DOI 10.54359/ps.v8i42.536 EDN: TDXHCY

3. Rubinshtein S.L. Fundamentals of General Psychology. Saint Petersburg: Piter Publishing House, 2019. 713 p.
4. Sergienko E.A. The System-Subject Approach: Substantiation and Prospects // Psychological Journal. 2011. Vol. 32. N 1. P. 120-132. EDN: NSIPQT
5. A meta-analysis of factors influencing the development of trust in automation / K.E. Schaefer [et al.] // Human Factors. 2016. Vol. 58. № 3. P. 377-400. DOI: 10.1177/0018720816634228
6. Automation technology and sense of control: A window on human agency / B. Berberian [et al.] // PLoS ONE. 2012. Vol. 7. № 3. URL: <https://pubmed.ncbi.nlm.nih.gov/22479528/> (accessed: 20.03.2026). DOI: 10.1371/journal.pone.0034075
7. Bainbridge L. Ironies of automation // Automatica. 1983. Vol. 19. N 6. P. 775-779. DOI: 10.1016/0005-1098(83)90046-8
8. Bratman M.E. Shared Agency: A Planning Theory of Acting Together. Oxford: Oxford University Press, 2014. 208 p.
9. Buçinca Z., Malaya M.B., Gajos K.Z. To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making // Proceedings of the ACM on Human-Computer Interaction. 2021. Vol. 5. CSCW1. Article 188. P. 1-21. DOI: 10.1145/3449287
10. Coeckelbergh M. AI Ethics. Cambridge, MA: MIT Press, 2020. 248 p.
11. De Jaegher H., Di Paolo E. Participatory sense-making // Phenomenology and the Cognitive Sciences. 2007. Vol. 6. N 4. P. 485-507. DOI: 10.1007/s11097-007-9076-9 EDN: VSAWVN
12. Ehsan U., Riedl M.O. Human-centered Explainable AI: Towards a Reflective Sociotechnical Approach // HCI International 2020 – Late Breaking Papers: Multimodality and Intelligence. Cham: Springer, 2020. P. 449-466. DOI: 10.1007/978-3-030-60117-1_33
13. Epley N., Waytz A., Cacioppo J.T. On seeing human: A three-factor theory of anthropomorphism // Psychological Review. 2007. Vol. 114. N 4. P. 864-886. DOI: 10.1037/0033-295X.114.4.864
14. Explanations can reduce overreliance on AI systems during decision-making / H. Vasconcelos [et al.] // Proceedings of the ACM on Human-Computer Interaction. 2023. Vol. 7. P. 1-38. DOI: 10.1145/3579605
15. Floridi L., Cowls J. A Unified Framework of Five Principles for AI in Society // Harvard Data Science Review. 2019. Vol. 1. N 1. DOI: 10.1162/99608f92.8cd550d1
16. Frith C.D., Blakemore S.-J., Wolpert D.M. Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action // Brain Research Reviews. 2000. Vol. 31. N 2-3. P. 357-363. DOI: 10.1016/S0165-0173(99)00052-1 EDN: AETHXL
17. Guidelines for Human-AI Interaction / S. Amershi [et al.] // Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019. Paper 3. P. 1-13. DOI: 10.1145/3290605.3300233
18. Haggard P. Sense of agency in the human brain // Nature Reviews Neuroscience. 2017. Vol. 18. N 4. P. 197-207. DOI: 10.1038/nrn.2017.14.
19. Hybrid Intelligence / D. Dellermann [et al.] // Business & Information Systems Engineering. 2019. Vol. 61. N 5. P. 637-643. DOI: 10.1007/s12599-019-00595-2 EDN: ZEGZJW
20. Jarrahi M.H. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making // Business Horizons. 2018. Vol. 61. N 4. P. 577-586. DOI: 10.1016/j.bushor.2018.03.007
21. Latour B. Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford: Oxford University Press, 2005. 320 p. DOI: 10.1093/oso/9780199256044.001.0001
22. Long D., Magerko B. What is AI Literacy? Competencies and Design Considerations // Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020. P. 1-16. DOI: 10.1145/3313831.3376727

23. Miller T. Explanation in artificial intelligence: Insights from the social sciences // Artificial Intelligence. 2019. Vol. 267. P. 1-38. DOI: 10.1016/j.artint.2018.07.007
24. Moore J.W. What is the sense of agency and why does it matter? // Frontiers in Psychology. 2016. Vol. 7. URL: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2016.01272/full> (дата обращения: 20.03.2026). DOI: 10.3389/fpsyg.2016.01272
25. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?/ E.M. Bender [et al.] // Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021. P. 610-623. DOI: 10.1145/3442188.3445922.
26. O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown Publishing Group, 2016. 259 p.
27. Parasuraman R., Manzey D.H. Complacency and bias in human use of automation: An attentional integration // Human Factors. 2010. Vol. 52. N 3. P. 381-410. DOI: 10.1177/0018720810376055
28. Parasuraman R., Riley V. Humans and automation: Use, misuse, disuse, abuse // Human Factors. 1997. Vol. 39. N 2. P. 230-253. DOI: 10.1518/001872097778543886
29. Power to the people: The role of humans in interactive machine learning / S. Amershi [et al.] // AI Magazine. 2014. Vol. 35. N 4. P. 105-120. DOI: 10.1609/aimag.v35i4.2513
30. Rahwan I. Society-in-the-loop: Programming the algorithmic social contract // Ethics and Information Technology. 2018. Vol. 20. N 1. P. 5-14. DOI: 10.1007/s10676-017-9430-8 EDN: MCMADR
31. Rai A. Explainable AI: From black box to glass box // Journal of the Academy of Marketing Science. 2020. Vol. 48. N 1. P. 137-141. DOI: 10.1007/s11747-019-00710-5 EDN: MNBEIB
32. Shneiderman B. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy // International Journal of Human-Computer Interaction. 2020. Vol. 36. N 6. P. 495-504. DOI: 10.1080/10447318.2020.1741118 EDN: ENPGDI
33. Synofzik M., Vosgerau G., Newen A. Beyond the comparator model: A multifactorial two-step account of agency // Consciousness and Cognition. 2008. Vol. 17. N 1. P. 219-239. DOI: 10.1016/j.concog.2007.03.010
34. Waytz A., Heafner J., Epley N. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle // Journal of Experimental Social Psychology. 2014. Vol. 52. P. 113-117. DOI: 10.1016/j.jesp.2014.01.005
35. Wegner D.M. The Illusion of Conscious Will. Cambridge, MA: MIT Press, 2002. 419 p. DOI: 10.7551/mitpress/3650.001.0001

Поступила в редакцию: 15.04.2026

Рецензия получена: 08.06.2026

Принята в печать: 20.06.2026