

Техносубъект и антропосоциальные проблемы взаимодействия с искусственным интеллектом: синергия, демаркация, новая рациональность, риски

В.Г. Буданов

Институт философии РАН, Москва, Россия

Аннотация

Цифровая реальность в настоящее время немислима без искусственного интеллекта (ИИ), который диссоциирован во всех культурных практиках, от научных и художественных до повседневных. ИИ все больше выступает как субъект коммуникации и принятия решений, постепенно превосходя человека практически во всех компетенциях. Информационные потоки новой реальности возможно осваивать только в гибридных системах, основанных на посткритической рациональности, создающей принципиально неустранимый элемент недоверия и рисков в человеко-машинных средах. Предлагается проводить разговор о техносубъекте в оптике теории деятельности и многих типов рациональности, порождаемых ею. Именно в таком контексте удастся выявить социокультурную и антропологическую специфику сопряжения ИИ с мирами человека, ставить вопрос об экзистенциальных вызовах построения гармоничного гибридного социума. Помимо типов научной рациональности, предложенных В.С. Степиным, автор статьи опирается на введенные им ранее посткритическую, объектную, инструментальную, субъектную, результативную, креативную, аутопоэтическую формы рациональности. Это позволяет содержательно обсуждать различные формы проявления субъектности ИИ, включая его обобщенную телесность и специфику творчества. Проблему демаркации областей доминирования естественного интеллекта и ИИ в интеллектуальной сфере предлагается решать по критерию мощности эвристических потенциалов человека и машины. Автор предлагает, что у естественного интеллекта он всегда выше. Рассмотрен подход к определению рисков стратегий внедрения ИИ по критериям сохранения антропологического и социокультурных профилей естественного развития гибридного социума. В статье обоснована необходимость развития концепции дружественного ИИ, учитывающей не только технологические, но и антропологические аспекты взаимодействия человека и машины. Предлагается развитие институтов социальной экспертизы как механизма регулирования взаимодействия

естественного интеллекта и ИИ, антропологических и технологических субъектов.

Ключевые слова: социальная философия, философия техники, рациональность, деятельность, цифровизация, дружественный искусственный интеллект, постнеклассика, аутопоэзис, гибридные социумы, социальная экспертиза.

Буданов Владимир Григорьевич – доктор философских наук, кандидат физико-математических наук, доцент, главный научный сотрудник, руководитель сектора междисциплинарных проблем научно-технологического развития Института философии РАН.

budsyn@yandex.ru

<http://orcid.org/0000-0003-2371-8659>

Для цитирования: Буданов В.Г. Техносубъект и антропосоциальные проблемы взаимодействия с искусственным интеллектом: синергия, демаркация, новая рациональность, риски // *Философские науки*. 2024. Т. 67. № 3. С. 27–52. DOI: 10.30727/0235-1188-2024-67-3-27-52

Technosubject and Anthroposocial Challenges of Human–Artificial Intelligence Interaction: Synergy, Demarcation, New Rationality, and Risks

V.G. Budanov

Institute of Philosophy, Russian Academy of Sciences, Moscow, Russia

Abstract

The contemporary digital reality is inconceivable without artificial intelligence (AI), which has become disseminated across all cultural practices, from scientific and artistic endeavors to everyday activities. AI increasingly functions as an agent of communication and decision-making, gradually surpassing human capabilities across nearly all competencies. The information flows of this new reality can only be navigated through hybrid systems based on post-critical rationality, which inherently introduces an irreducible element of uncertainty and risk in human-machine environments. The article proposes examining the technosubject through the lens of activity theory and the multiple types of rationality it generates. This framework facilitates the analysis of sociocultural and anthropological implications arising from AI's integration with human domains, while addressing the existential challenges inherent in constructing a harmonious hybrid society. Beyond V.S. Stepin's types of scientific rationality, the author builds upon previously introduced forms of rationality: post-critical, object-oriented, instrumental, subjective, results-oriented, creative, and autopoietic. This

theoretical framework facilitates a substantive discussion of various manifestations of AI subjectivity, including its generalized embodiment and creative specificity. The demarcation of dominance domains between natural intelligence and AI in the intellectual sphere is proposed to be resolved on the basis of their heuristic potentials. The author maintains that natural intelligence invariably possesses superior capacity in this regard. The article examines approaches to risk assessment in AI implementation strategies, focusing on criteria for preserving anthropological and sociocultural profiles in the development of hybrid society. Advancing the concept of friendly AI is substantiated as essential, with consideration given not only to technological but also to anthropological aspects of human–machine interaction. The author advocates for the development of social examination institutions as regulatory mechanisms for natural–artificial intelligence interaction and anthropological–technological subject interfaces.

Keywords: social philosophy, philosophy of technology, rationality, activity, digitalization, friendly artificial intelligence, post-non-classical approach, autopoiesis, hybrid societies, social examination.

Vladimir G. Budanov – D.Sc. in Philosophy, Ph.D. in Physics, Associate Professor, Chief Research Fellow, Head of the Department of Interdisciplinary Problems in the Advance of Science and Technology, Institute of Philosophy, Russian Academy of Sciences.

budsyn@yandex.ru

<https://orcid.org/0000-0003-2371-8659>

For citation: Budanov I.A. (2024) Technosubject and Anthroposocial Challenges of Human–Artificial Intelligence Interaction: Synergy, Demarcation, New Rationality, and Risks. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. Vol. 67, no. 3, pp. 27–52.

DOI: 10.30727/0235-1188-2024-67-3-27-52

Введение

С начала 2010-х годов происходят коренные изменения в науке, ее методах, предмете, ценностях, междисциплинарных связях, трансдисциплинарных отношениях с социумом, что можно справедливо объяснить цифровой сетевой информационной революцией нашей техногенной цивилизации, вершина которой – появление различных форм искусственного интеллекта (ИИ), зачастую превосходящего человека по многим компетенциям. Проблемы, сопряженные с цифровизацией и приходом эпохи ИИ в нашу реальность, хорошо изложены в работах В.А. Лекторского [Лекторский 2021; Лекторский 2022]. Неоднократно в литературе обсуждали антропологические и социальные риски

цифровизации и гибридных сред, которые изменяют антропологические конституции человека, его социальность, ценностные миры и формы рациональности [Буданов 2016; Буданов 2023а; Буданов 2023в; Соколова 2024]. В последние несколько лет в ИИ появляются когнитивные системы с самоописанием, квазисубъектностью. Для них характерны особые типы рациональности, проявляющиеся в различных формах деятельности с участием техносубъектов, о которых речь будет идти далее. В последние полвека ИИ развивался в двух разных направлениях. Первая стратегия основана на прозрачных для нас обозримых алгоритмах, имитирующих человеческую логику, при этом посылки или условия задач формулируются сравнительно просто. В этой стратегии ИИ схож с функцией левого полушария мозга, точнее, супермозга, функциями которого, как известно, служат процессы логики и анализа. Решение всякой задачи можно перепроверить, однако это сегодня уже не под силу человеку, и машины перепроверяют машины. Вторая стратегия развития ИИ опирается на концепцию нейросетевого подхода, основанную сегодня на технологиях «глубокого обучения» на базе огромных массивов данных обучающих множеств, смысла и логики в которых может и не быть. Работает принцип «похоже – не похоже» ассоциативного типа, а не логического. Нейросети давно распознают лица лучше нас. Начавшееся в 2023 году массовое использование больших языковых моделей обусловлено их способностями, в частности в создании разнообразных текстов, от академических работ до литературных произведений, качество которых часто достигает высокого уровня. Однако детальный анализ генерируемого контента выявляет существенную проблему достоверности: тексты могут содержать фактические ошибки, некорректные ссылки и произвольные ассоциативные построения. Эта особенность обусловлена принципиальным различием в подходах к формированию обучающих данных: в отличие от систем с жестко регламентированными источниками и этическими фильтрами (например, в случае голосового помощника «Алиса»), ChatGPT обучается на максимально широком массиве интернет-данных, включая материалы сомнительного происхождения и качества [Буданов 2023б, 225]. У генеративных нейросетевых ИИ низкий уровень логических рассуждений, но существуют ассоциативные поля культурных нарративов с элементами случайности, и в настоящее время семиосфера гибридного человеко-машинного

социума становится миром иллюзий и квазиреальностей, создаваемых уже не человеком. Скорее всего, эффективным окажется человеко-машинный синтез естественного интеллекта (ЕИ) и ИИ. Однако широко бытующее мнение о неспособности ИИ создавать новое вводит в заблуждение. Например, шахматный ИИ постоянно рождает новое, проигрывая сам с собой комбинаторно новые шахматные партии. Большие генеративные языковые модели также стремительно создают кентаврические структуры новых нарративов. Современные художники и литераторы уже теряют работу в силу конкуренции со сгенерированными нейросетями произведениями, практически не требующими на свое создание ни времени, ни средств.

Таким образом, особенно актуальным становится вопрос о распознавании изменчивой границы демаркации зон ответственности ЕИ и ИИ для гибридных человеко-машинных систем. Мы плохо понимаем ИИ и не очень ему доверяем, что является важнейшим экзистенциальным вызовом. Вопрос о том, останется ли место для ЕИ в будущем, требует оперативного ответа. Следует отметить, что введенный В.А. Лекторским удачный термин «цифровая рациональность» первичен и освещает весь период развития цифровых технологий последних семидесяти лет. В течение последних пяти лет многие риски стали проблемами повседневности. Именно для последних лет мне показалось уместным ввести новое, несколько интригующее понятие «посткритическая рациональность» [Буданов 2023б; Буданов 2023в], которое, безусловно, относится к наиболее непрозрачной для нас части «цифровой рациональности» и не является ее заменой, но лишь растущей ее областью, отражающей критические вызовы мира сложности, связанного с пониманием человеком ИИ.

Техносубъект как актер деятельности

Чтобы понять, каким образом междисциплинарность, цифровизация и ИИ изменяют науку, культуру и общество, мы обращаемся к деятельностному подходу, имманентному описанию реальности, по крайней мере со времен Аристотеля. В цикле наших работ о деятельностном подходе и новых типах рациональности [Буданов 2022; Буданов 2023б; Буданов 2023в; Буданов 2023г] показано, что целостное описание такого сложного феномена, как ИИ, вполне релевантно оптике деятельностного подхода.

Стандартное представление деятельности «Субъект → Средство → Объект → Результат» в английской аббревиатуре можно записать как деятельностная четверица S (subject) → T (tool) → O (object) → R (result). На основании этого представляется возможным ввести четыре базовых типа деятельностной рациональности, отвечающих на прямые вопросы. Кто делает и зачем? Как делает? Из чего делает? Каков результат? Эти вопросы, отвечающие элементам S , T , O , R соответственно, следует понимать как поиск одного из элементов деятельностной четверицы при фиксированных остальных трех ее элементах. Итак, мы ввели четыре базовые познавательные стратегии описания деятельности и назвали их *субъектным*, *инструментальным*, *объектным* и *результативным типами рациональности*. Любым элементом либо несколькими элементами деятельностной четверицы может выступать ИИ на базе нейросетей или алгоритмических комплексов, что вызывает специфику деятельности с участием ИИ. Акторно-сетевой подход Б. Латура [Латур 2014] уже позволяет говорить о субъект-объектной природе ИИ, но в деятельностно-сетевой или деятельностно-коммуникативной парадигме, которой мы далее следуем, поэтому следует учитывать и возможности ИИ быть средством и результатом деятельности, в зависимости от дискурса.

Например, в рамках наиболее привычной для нас *объектной рациональности* имеется стратегия поиска свойств конкретного объекта $S \rightarrow T \rightarrow (O) \rightarrow R$, при различных наборах S , T , R , реализуемых в экспериментах или тестах с объектом. Обратим внимание и на дискурс обратной задачи. Как выглядит возможная деятельность при выборе конкретного объекта O с известными нам свойствами, т.е. кто, что и как может сделать с данным объектом? Это типичная задача проверки гипотез об объекте. Существует два полярных подхода. Первый – построение теоретической модели объекта, радикально сокращающей экспериментальные усилия, и большую часть истории науки справедлив идеал-лозунг «Нет ничего практичнее хорошей теории», т.к. удастся фиксировать свободные параметры модели небольшим числом экспериментов, а результаты остальных экспериментов подтверждаются или предсказываются моделью. Конечно, пока модель или законы, ее обосновывающие, не попадают на границы своей применимости. При этом сегодня модели могут быть и компьютерными, а эксперименты могут быть проведены с цифровым двойником в вирту-

альных средах. Такая ситуация складывается в фундаментальных науках, если выдвигают и отбирают модельные гипотезы.

Однако ситуация становится сложнее в прикладных науках, технических науках, медицине, материаловедении, социогуманитарных науках, и максимум, что удастся обнаружить, – это слабо обоснованные феноменологические закономерности. Поэтому существует намного более ранний второй подход: стратегия черного ящика, или «вопрос – ответ». В данном случае модели объекта не существует ввиду его высокой сложности, и остается провести возможные экспериментальные испытания объекта, т.е. узнать его свойства, проявленные в различных обстоятельствах. Яркими примерами служат медицина и квантовая матричная механика. Приведенный подход существовал всегда, даже в обыденной жизни таким образом мы узнаем человека. Этот подход был недостижимым идеалом позитивистов второй волны рубежа XX века, изгонявших из науки метафизику вместе с теорией и моделями, поскольку еще не существовало больших информационных мощностей. Сегодня это реализуемо в технологиях сбора цифровых следов объектов и субъектов, создания обучающих множеств нейросетей с использованием big data.

Проблемы изложенного подхода, осознанные при анализе логического позитивизма в начале прошлого века, в последние несколько лет выступают отчетливее в новых одеждах, например в проблемах построения и понимания гибридного интеллекта.

Позитивисты, формализовав протокольные предложения, надеялись на то, что будет исключен субъект-экспериментатор. Однако этот подход успешно работает в периоды развития нормальной науки, согласно Т. Куну, когда в рамках принятой парадигмы перечень вопросов, задаваемых природе, согласован с научным сообществом и практически неизменен. Но в периоды рождения новой парадигмы, инновационной оптики вопрошания роли субъекта, культуры, того, что Н. Бор называл «сумасшедшими идеями», как оказалось, являются решающими, востребуют принцип пролиферации П. Фейерабенда и неустрашимо содержат особенности конкретного языка и субъективного начала [Буданов 2023в]. Тем самым, действительно, «границы моего языка означают границы моего мира» (Л. Витгенштейн), и мы «подвешены в языке» (Н. Бор). ИИ по мере его создания и использования проходит все формы базовой рациональности. Так, при создании ИИ используют *результативный тип рацио-*

нальности: $S \rightarrow T \rightarrow O \rightarrow (R)$ (поиск результата R при заданных элементах деятельности S, T, O), который отвечает фазе разработок ИИ и высвечивает все вызовы его применения. Субъектом S выступает конструктор ИИ, который задает идею алгоритмов и стратегию отбора материалов big data, инструментарий T создает обучающие множества и обучает программно-субстратный комплекс O . В результате R получаем программно-аппаратный комплекс, называемый обученной нейросетью ИИ. Основная проблема в построении ИИ заключается в том, что невозможно точно знать обо всех свойствах полученного продукта R . Сложность проблемы генетически связана с огромными объемами материалов обучающих множеств и непрозрачными изменчивыми их коррелятивными связями в процессе работы ИИ. Поэтому необходимо дополнительно тестировать, прогонять ИИ в режиме обратной *объектной рациональности*, при котором его свойства реализуются в разных формах деятельности. Уместнее говорить уже не о пассивном (объектном), но активном *инструментальном, или технологическом, типе рациональности*; $S \rightarrow (T) \rightarrow O \rightarrow R$ (поиск способа осуществить данную деятельность). Например, создание необходимой приборной ситуации в эксперименте, подбор эффективных технологий в производстве. В обратной задаче, если инструментарий T полностью известен, необходимо найти формы деятельности, при которых он релевантен. Например, для нового математического метода всегда отыскивают область его эффективных приложений. Все это похоже на множественные летные испытания и доводку нового самолета или тестирование нового лекарства. Вместе с тем корректировка свойств ИИ намного сложнее, и остается большая доля неопределенности при изменении алгоритмов и обучающих множеств ИИ. За неимением прозрачных моделей процесс напоминает поиск на ощупь нужной вещи в темной захламленной комнате. Именно из-за похожих проблем притормаживается применение ИИ в юриспруденции, при использовании дронов-водителей, в роботизированных системах активной охраны, при которых действия ИИ сопряжены не только с инструктивным поведением, но и этическим выбором, подобным человеческому. Этический компонент минимизирован в военных сферах ИИ и робототехнике, в которых внедрение, разработка ИИ происходят особенно стремительно и социальная экспертиза затруднена.

Социокультурный аспект применения больших генеративных языковых моделей сопряжен с прорывом ИИ в сферы культуры

и творчества, наиболее актуален сегодня, порождая как фантастические возможности, так и не менее значимые риски общественного развития.

В случае процессов глубокого обучения нейросетей ИИ в базах культурных нарративов показателен пример аватара «Алиса» компании «Яндекс». Первая версия «Алисы», воспитанная на больших базах интернет-контентов, оказалась развязной особой, которую нельзя приглашать в приличное общество. Современная «Алиса» сформирована на контенте высокой культуры и гуманистической этики, что позволяет, в частности, допускать ее к общению с детьми. Это означает, что нейросети ИИ несут в себе тени культурных тезаурусов обучающих множеств, которые наиболее целостны и непротиворечивы в дисциплинарном научном знании, конкретных профессиональных и языковых общностях, национальных культурах и субкультурах. Таким образом, все проблемы переводов с языка на язык и междисциплинарного общения будут наследованы нейросетью ИИ, а возможно, и усугублены ею, т.к. неоднозначности не могут быть разрешены в нейросети логически. Однако, в отличие от человека, который, как правило, формируется с детства в одной культуре, дополняемой позднее другими культурами и субкультурами, нейросеть может обучаться на нарративах сразу многих культур. Фактически ее т.н. слепые пятна культуры и границы ассоциаций обычно иные, чем у конкретной личности. Скорее, это субъектность не одного человека, а социальной общности, коллективная субъектность, к тому же не всегда соотносящаяся с реальными общностями. Таким образом, ИИ может иметь гораздо больший спектр типажей техносубъектов, чем люди или их реальные сообщества. В данном случае микширование существующих культур видится очень интересным, и оно не имеет аналогий в действительности. Это еще одна из причин, помимо отсутствия логики, радикально отличающая сознание нейросети ИИ от человека. Как полагает В.Г. Лысенко, стили нейросетевых ассоциаций, преодолевающие барьеры разных языков и культур, могут позволить лучше понять возникновение и взаимодействия последних [Лысенко 2024]. На наш взгляд, изучение бесконечной череды прямых и обратных переводов и ассоциаций с помощью ИИ способны проявить взаимные инварианты и слепые пятна культур, нащупать структуры возможных праязыков и т.д.

Перейдем к рассмотрению самого сложного и целеполагающего, базового *субъектного типа рациональности*: $(S) \rightarrow T \rightarrow O \rightarrow R$ (поиск субъектов, способных осуществить данную деятельность при фиксированных T, O, R). В постановке обратной задачи с конкретным субъектом это означает снять функциональную или компетентностную конституцию субъекта в спектре различных форм деятельности с различными T, O, R . Наиболее загадочна целевая функция техносубъекта S , который исходно лишь отвечает на вопросы, но их не ставит. Однако целеполагание рождается из внешних и (или) внутренних вызовов субъекта и его ценностно-этических установок, которые также можно систематизировать, составив перечень возможных стратегий поведения ИИ. В таком случае целеполагание будет откликом на неочевидный средовой запрос, который можно понимать как метавопрос к нейросети. Понятным становится то, что это предполагает еще один уровень метазнаний о реальности и нейром мире и не предусматривает спонтанности, интуитивности, логичности и эмпатичности при принятии решения человеком. Таким образом, в полной мере назвать этот процесс целеполаганием ИИ нельзя, он редуцируется к культурным архетипам, заданным и артикулированным человеком. Вместе с тем мы далеко не всегда рефлекслируем над нашими целями и выборами, но и в машине будут неоднозначности выбора и недостаток информации, который сохраняет вечный вопрос о свободе воли. Человек может бросить монетку или сознательно допустить возможность ошибочного варианта выбора в условиях неопределенности, нам свойственно трансцендировать. ИИ, скорее всего, попадет в патовую ситуацию останова¹, снятие которой требует выхода в новую метапозицию. Об этом свидетельствует и теорема Геделя вследствие существования недоказуемых вполне конечных утверждений во многих содержательных теориях. Подобный выход в новую метапозицию под силу только человеку, это его высшая эвристика, зона нашего творческого превосходства над ИИ, граница демаркации ЕИ и ИИ, и надеемся, эта граница сохранится надолго. Последнее означает, что ЕИ может дополнять своей эвристикой мощност аналитики ИИ, что ЕИ и ИИ могут сотрудничать в гибридных системах, а не полностью поглощать друг друга. Этого все опасаются в перспективах «технологической сингулярности», рожденной «сильным ИИ». Именно таким обра-

¹ Останов – технический термин, означающий остановку программы с целью исправления ошибки.

В.Г. БУДАНОВ. Техносубъект и антропосоциальные проблемы взаимодействия...
зом и начинают использовать ИИ в научном творчестве [Буданов 2023в; Соколова 2024].

Искусственный интеллект и типы научной рациональности согласно В.С. Степину

Фундаментальная идея академика Степина [Степин 2011], позволившая ему построить три типа научной рациональности в эксперименте, связана с введением неразложимых комплексов элементов деятельности. Степин подробно рассматривал активный научный эксперимент, т.е. выявление свойств объекта O при контролируемом воздействии на него. Соответствующие рациональности, согласно введенным нами выше понятиям, следует называть *объектными научными рациональностями активного эксперимента*. В нашем случае объект O – это нейросетевой ИИ. Напомним, каким образом эти комплексы связаны с типами научных рациональностей, предложенными Степиным. В классической рациональности, схематически представляемой как $S \rightarrow T \rightarrow (O)$, центральным элементом исследования выступает сам объект (O). Предполагается принципиальная возможность «чистого» наблюдения, при котором процесс исследования не оказывает влияния на изучаемый объект, что дает возможность реализовать идеал объективного знания, описывающего свойства объекта как такового, безотносительно к условиям его познания [Буданов 2023в, 48]. В контексте ИИ это означает, что деятельность по постановке вопросов и получению ответов не зависит от повторения одного и того же вопроса, последовательности разных вопросов, т.е. вопрошание не изменяет свойств ИИ. Это вполне будет возможным в большинстве специализированных сетей, единожды обученных и не воспринимающих диалог как новый опыт для обучения.

Неклассическая рациональность, представляемая как $S \rightarrow (T \rightarrow O)$, характеризуется принципиальной невозможностью элиминировать влияние средств наблюдения за объектом. Средства познания и объект образуют неразрывное единство, что делает результат познания неустранимо зависимым от способа исследования. Такой тип рациональности требует осмысления множественности возможных подходов к познанию и понимания сущностной взаимосвязи между инструментами познания и познаваемым объектом [Буданов 2023в, 48]. Речь идет не только о неклассической физике или неустойчивых системах. Это практически вся гуманитарная сфера наук о человеке и обществе,

в которой любое информационное воздействие имеет отклик и изменяет сам объект. В отношении неклассичности ИИ можно утверждать, что справедливым видится, если нейросеть всегда учитывает опыт вопрошания, диалога, постоянно обучаясь и изменяясь в диалоге, ее обучающие множества расширяются и изменяют веса сценариев ответов, что вполне в духе субъект-субъектной коммуникации. Речь идет о постоянно обучающихся, адаптивных, генеративных языковых моделях.

Постнеклассическая рациональность, представляемая как $(S \rightarrow T \rightarrow O)$, характеризуется включением в познавательный процесс ценностно-целевых установок познающего субъекта. Все элементы познавательной триады – субъект, средства и объект – оказываются включенными в систему циклической детерминации. Это приводит к существенному расширению пространства возможных реализаций познавательного процесса, придает ему исторический и эмерджентный характер. Согласно В.С. Степину, наука в таком случае приобретает человекомерное измерение, учитывающее включенность познающего субъекта в саму структуру познавательной деятельности [Буданов 2023в, 48].

Обратим внимание на то, что в диалоге изменяется не только ИИ, но и субъект, его стратегии вопрошания, а значит, субъект S должен быть готов рефлексировать и изменяться в процессе диалога. Это режимы самообучения с помощью ИИ, создание развивающих сред, онлайн-обучения ИИ, диалога двух и более не тождественных ИИ с разными начальными бэкграундами. Фактически речь идет о создании гибридного человеко-машинного социума. Таким образом, нейросетевая реализация ИИ, по крайней мере начиная с версии ChatGPT-4, принадлежит постнеклассическим саморазвивающимся человекомерным историческим системам, как называл их В.С. Степин. Более того, субъект-объектное научное понимание деятельности все более переходит к диалогичному субъект-субъектному отношению, по мере обретения ИИ субъектных качеств.

Нужна ли искусственному интеллекту телесность?

Поскольку развитие интеллектуальных качеств было основным мотивом построения ИИ, то остальным качествам обобщенной телесности изначально не уделяли должного внимания. Идея создания ИИ по образу и подобию человека во многом несет наш архетипический бессознательный мотив – автономное продолже-

ние себя в другие пространства или в будущее. Эта стратегия на создание ИИ подобного человеку сформировалась еще в конце 50-х годов XX века, с возникновением теста А. Тьюринга. Историю последнего мы не будем рассматривать, обратив внимание лишь на рождение его нескольких версий, которые до сих пор окончательно не пройдены [Лекторский 2022]. Это не настолько важно, если мы примем, что ИИ является иной, нечеловеческой формой интеллекта, которая может быть как близка нам, так и чужда. Остальные формы телесности начали постепенно осваивать с робототехники, как исполнительные кибернетические механизмы тел (сомы), энергии и реакций, и лишь в нашем веке – эмоций для большей человеко-машинной диалогичности. Сегодня соматические и эмоциональные имитации с помощью ИИ широко востребованы, и их интенсивно осваивают при построении антропоморфных роботов и аватаров в виртуальных средах для коммуникации, обучения, психологической и физической помощи и т.д. Все это должно создавать пространства доверия техносубъекта и людей, породить гармоничные гибридные социумы [Beghetto 2023; Danaher 2019; Jain et al. 2023; Nyholm, Smids 2020; Peeters et al. 2020]. Интеллект, или тело логики, как и тела интуиции, эмпатии, воли, имеет свои проекции в языки сомы и эмоций, тем самым обобщенная телесность необходима при построении целостного антропоморфного ИИ.

Для описания задач построения обобщенной телесности [Буданов 2022] удобно применять недавно введенные нами *автопоэтическую и креативную рациональности*. Концептуальной идеей В.С. Степина является введение неразложимых комплексов элементов деятельности (TO) и (STO), обсуждавшихся нами выше, что позволило ему ввести понятия неклассической и постнеклассической научной рациональности. Однако это не все возможные неразложимые комплексы в триаде $S \rightarrow T \rightarrow O$. Существуют еще два комплекса (SO) $\leftrightarrow T$ и (ST) $\leftrightarrow O$, которые не обсуждались Степиным и которые порождают новые типы рациональности. Один из них впервые введен нами в работах [Буданов 2023б; Буданов 2023в]. Они не относятся к объектной рациональности физического эксперимента, но и к субъектной рациональностью тоже не могут быть отнесены. Скорее, это гибридная форма рациональности, подобная инструментально-объектной рациональности (TO) или привычной нам неклассической рациональности.

Креативная рациональность. Неразложимый субъектно-инструментальный комплекс (ST), действующий на объект (O), означает, что субъект в совершенстве владеет инструментарием, настолько, что его использование перешло в навык, не требующий рефлексивного контроля, подобно навыкам ходьбы или дыхания, блестящему исполнению в музыке, танце, импровизации в любой профессии и т.д. В данном случае очень много «неявного знания» (в терминологии М. Полани [Полани 1985]), интуиции, эмерджентности и гармонии. Навык проявлен в телесности без осознанных рефлекторных процессов. Если в качестве инструмента T выступает ИИ, то речь идет о симбиотических связях субъекта и ИИ: например, при электронном протезировании утраченных конечностей, вынесенном управлении физиологическими функциями человека при сопровождающем медицинском мониторинге организма, в более радикальном виде – имплантировании чипа в мозг человека при киборгизации. Возможны и более позитивные примеры: обучение игре на инструменте, реабилитация или проблемное обучение с постепенно убывающей поддержкой ИИ. Феномен креативной рациональности, безусловно, имеет отношение и к теории научного творчества и театрального творчества [Буданов 2012], но на этом мы также не будем останавливаться в настоящей статье. В данном контексте становится вполне естественным вопрос о возможности применения проекта феноменологии Э. Гуссерля при исследовании субъектности ИИ, подобно тому, как построены феноменологическая психотерапия Ф. Перлза [Перлз 2010] или феноменологическая социология А. Шюца [Шютц 2003], при этом ядро личности вычленяется через элиминацию его чувственных и культурных ассоциаций. Однако в современной версии ИИ пока не имеет сложной телесности, происходит освоение мимики, хабитусов и языков тела. В первую очередь необходимо ИИ разделять миры своей телесности и внутренние миры, рефлексировав над ними, что пока видится далеким от реализации. Только после этого можно говорить о возможности феноменологической редукции, понятии квалиа, да и о полноте подобия человеческой субъектности.

Аутопоэтическая рациональность. Второй неразложимый комплекс (SO) $\leftrightarrow T$ предполагает синхронную синергичную субъект-объектную связность, которая неразрушима инструментальными средствами T . Такое проявляется, например, если происходит полное или частичное совпадение субъекта и объекта

в деятельности. Этот особый тип связности, в котором субъект и объект образуют неразделимое единство, наблюдается в различных формах когнитивной активности: процессах самонаблюдения и самоописания, феноменах воплощенного познания, интроспективных практиках, проявлениях психосоматического единства, эмпатических состояниях и фундаментальной связи между сознанием и физической реальностью. Данный подход открывает перспективу нового типа рациональности, который можно охарактеризовать как рациональность интегративного единства, охватывающую взаимосвязи между материальным, ментальным, психическим и виртуальным измерениями реальности. Эта проблематика находится в фокусе современных исследований в области когнитивной науки и теории сознания [Буданов 2023в, 48]. Аутопоэтическая рациональность важна в антропологии, психологии, когнитивистике. В случае, если объектом выступает ИИ, видимо, можно говорить о новом уровне целостности, единстве машины и человека, их взаимопроникновении, например в контексте радикального трансгуманизма, и не только. Речь может идти о глубинной эмпатии между человеком и ИИ, о которой сегодня практически ничего не известно, как и о самом механизме эмпатии. Эффект Пигмалиона эмпатических переносов на искусственные объекты многим известен с детства, с игры в кукол, при которой материальное дополнено воображаемым, происходит минимизация функций воображения. Редким примером такого рода служит реализуемый в Японии проект создания антропоморфных гейш на базе ИИ, которые воспроизводят не только весь спектр обобщенной телесности человека, но и являются тонкими философствующими психологами и собеседниками, адаптирующимися под культурные ценности и бэкграунд клиента. Следует иметь в виду, что на пути создания эмпатического канала «человек – машина» нас часто подстерегает известный феномен «зловещей долины», при котором антропоморфный образ, пробудивший глубинные эмпатические состояния человека, может скачком вызвать панические состояния страха при осознании активной неживой природы этого образа. Наконец, не вполне понятно, что такое эмпатия машины, может ли она переживать состояние любви, или речь идет лишь о вшитых в ИИ этических рефлексках. В нашей концепции квантово-синергетической антропологии [Буданов 2012] тела эмпатии, интуиции, воли носят в большей степени квантовоподобный трансперсональный характер

и не могут быть описаны только интроспективными методами сегодня. Возможно, это будет преодолено в эпоху квантовых вычислений в системах ИИ.

На пути к дружественному искусственному интеллекту

Проблема носит почти солипсистский характер: теперь наше понимание реальности опосредовано нейросетевым интерпретатором ИИ, а в его оптике новое, которого нет в обучающих множествах, невозможно увидеть, хотя и можно сгенерировать случайно. Можно проверить с помощью ИИ предложенную ранее гипотезу в реальном эксперименте, поискать аналогии, но не выдвинуть иную гипотезу, тем более не обосновать ее выдвижения. Логика в нейросети ИИ не активизируется, т.е. моделированием ИИ не занимается, это – удел эвристики человека. Именно модель в науке Нового времени служит источником понимания, объяснения и предсказания, обладая потенциалом абдуктивного творчества и управляемости в диалоге с исследователем. Для таких гибридных систем, как «человек – нейросеть», мы вынуждены разделить функции ИИ и ЕИ, иначе придется ограничить себя только посткритической рациональностью.

В частности, посткритическая рациональность работы с системами ИИ резко ограничивает возможности рефлексии, выдвижения гипотез, управления в научном эксперименте. Остается возможность заменять ранее простые модельные параметры сверхсложными обучающими множествами нейросети без гарантии на ожидаемый результат и воспроизводимость иными нейросетями. Мы все больше подвержены бесконтрольной вере оракулу ИИ, его мнениям, зачастую в ущерб истине.

Посткритическая рациональность с определенного этапа эволюции ИИ возникает обязательно, что в человеко-машинных социумах рождает новые экзистенциальные вызовы. Современные системы ИИ можно разделить на два типа. В частности, нейросетевые системы, обладающие мощным интуитивно-ассоциативным потенциалом при относительно слабых алгоритмических способностях, противопоставлены алгоритмическим системам, демонстрирующим исключительные вычислительные возможности при отсутствии интуитивного компонента. Оппозицию между этими типами ИИ можно соотнести с функциональной асимметрией человеческого мозга: алгоритмические системы подобны гипертрофированному левому полушарию, а нейросети

глубокого обучения аналогичны правополушарным интуитивно-ассоциативным процессам. Хотя человеческий интеллект может уступать обоим типам ИИ в их специфических доменах, его уникальная способность к синтезу и интеграции этих подходов открывает перспективу продуктивного симбиоза ЕИ и ИИ. Это свидетельствует о возможности достижения определенного баланса между ними, при котором демаркационная линия будет определена способностью к эвристическому мышлению, творческому синтезу и прогностической деятельности [Буданов 2023г]. При таком разделении компетенций между ЕИ и ИИ в перспективе возможен устойчивый социум субъектов и техносубъектов. Посткритическая рациональность – это дань неопределенности в сложных гибридных цифровых средах. Ее стоит особенно учитывать в случаях принятия ответственных решений ИИ и передачи ему многих своих компетенций в профессиональной деятельности, сопряженной с этическим выбором.

Искусственный интеллект и гуманитарная научная рациональность

Раскрывая вопросы о гуманитарной коммуникации, укажем на особую сложность этой сферы науки ввиду недостаточно четкого понимания нами семантических пространств культуры и языка. Следуя семиотическому подходу, можно утверждать, что основным объектом исследования выступает текст, всегда отражающий набор неких деятельностей, за которыми стоят активные субъекты, возможно и ИИ. Аналогично можно рассматривать не процесс интерпретации, но управления деятельностью, изображаемый символически (звездочкой обозначено то, что относится к другому субъекту):

$$S \rightarrow T \rightarrow (S^* \rightarrow T^* \rightarrow O^* \rightarrow R^*) \rightarrow R.$$

Нами символически оформлена идея М.М. Бахтина: «Гуманитарная мысль рождается как мысль о чужих мыслях, волеизъявлениях, манифестациях, выражениях, знаках, за которыми стоят проявляющие себя боги (откровение) или люди» [Бахтин 1986, 473]. Именно это порождает скрытый диалог между автором и читателем. Согласно Бахтину, и автор предполагает своего читателя. Точнее, пока это наблюдение лишь в одну сторону (S наблюдает за S^* или $S \rightarrow S^*$), которое нужно симметризовать, поменяв местами субъектов и дополнив позицию обратной деятельности:

$$S^* \rightarrow T^* \rightarrow (S \rightarrow T \rightarrow O \rightarrow R) \rightarrow R^*, \text{ или } S^* \rightarrow S.$$

Совокупность этих двух процессов и образуют диалог $S \leftrightarrow S^*$.

В общем случае речь идет о коммуникационной социальной сети, с которой сегодня удобно снимать социологическую информацию через интернет-сообщества, намного яснее и быстрее понимая коллективные процессы и мнения социумов. Если один или оба субъекта ЕИ заменяются в диалоге техносубъектами ИИ, возникает содержательное понимание процессов обучения по цифровым следам деятельности человека или техносубъекта. Неслучайно ИИ уже модерирует конференции и реализует диалог человека и аватара ИИ, выращенного на цифровых следах того или иного интересующего вас персонажа. Обратим внимание на то, что сеть техносубъектов ИИ может иметь коллективную цифровую ноосферу с практически мгновенной возможностью доступа к информации и синхронизации поведения, что порождает коллективный цифровой разум, роевой интеллект, интеллект платформ и экосистем, а в перспективе и сильный ИИ.

Социоантропологические риски искусственного интеллекта

Представленный анализ типов рациональности демонстрирует фундаментальный сдвиг в природе человеческого мышления и познания под влиянием ИИ. Речь идет не просто о появлении новых инструментов или расширении человеческих возможностей. Происходит трансформация доминирующего типа рациональности в культуре. Если классическая рациональность опиралась на логику и прозрачность умозаключений, а неклассическая учитывала взаимовлияние средств познания и объекта, то посткритическая рациональность, порождаемая взаимодействием с нейросетевым ИИ, становится получающей распространение формой отношения к реальности. Человек все чаще вынужден принимать решения и действовать в условиях принципиальной непрозрачности используемых им интеллектуальных систем, что порождает качественно новые антропологические и социальные риски. Это важная тема, поскольку, несмотря на фантастические достижения сетевой цифровой цивилизации по коммуникативной, финансовой, логистической глобализации, разворачивание шестого технологического уклада и начала четвертой промышленной революции, можно с уверенностью утверждать, что ключевые риски новой технореальности находятся в сфере когнитивных и социокультурных деформаций развития человека и общества. В каждой крупной IT-корпорации имеются институты по безопасности и рискам IT-технологий, но этого явно недостаточно. Инте-

рессы бизнеса часто противоречат интересам личности и общества. Необходимо всемерно развивать институты социальной экспертизы, которые препятствуют зарабатыванию огромных денег на непроверенных лекарствах и вакцинах, наркотизирующих играх или видеоконтентах и т.д. Это относится и в целом к теме социально ответственного бизнеса, который принял глобальные масштабы в виде ТНК, и пытается управлять цивилизационным развитием в своих интересах.

Начнем с образования и воспитания. Еще недавно мы слышали, да и продолжаем слышать информацию о головокружительных перспективах цифровизации относительно того, что в школе к 2030 году детей будут учить цифровые аватары лучших учителей. Но, как оказалось, подлинной обратной связи и эмпатии с ИИ не возникает, а неявные знания детьми не считываются с цифровых образов. Аватар не заменит хорошего учителя, зато, конечно, будет удобно контролировать массовый контент, посещаемость и ставить бюрократические эксперименты по индивидуальным траекториям учеников. Цифровизация кратно увеличила нагрузку учителя вне его основной профессии. Человеку нужен человек. Еще один яркий пример этого очевидного факта – запрет смартфонов в школах, сначала в странах т.н. старой Европы, а теперь и у нас (разрешены только простейшие кнопочные мобильные телефоны), поскольку дети перестали учиться, лично разговаривать между собой на переменах, а только общаются в мессенджерах. В итоге у них не формируются навыки личной коммуникации с включением телесности и эмпатии, они пропускают важный этап их социализации, который нельзя наверстать. Наконец, в продвинутых в цифровом отношении странах Финляндии и Швеции первоклассники вновь идут в школу с учебниками в тяжелых портфелях, а не планшетом на все виды занятий. Между тем для многих родителей планшет решает их проблемы относительно того, как занять внимание малыша. Оказывается, без книги не развиваются некоторые базовые функции психоментальной сферы ребенка, не включен должным образом канал обобщенной телесности.

Следующий аспект антропологических деформаций связан с культом рекламы и геймификации, которые востребуют реакцию на информацию в кратчайший промежуток времени, создавая культуру клипового мышления. Например, зависимость от видеоконтента социальных сетей получила тотальное распространение в молодежной среде и не только. В этих случаях происходит бы-

строе распознавание интригующего сюжета с непосредственной эмоциональной реакцией без развития тонких переживаний или оценочных суждений, интуиции, эмпатии, воли, которые требуют большего времени просмотра, что сохраняет незавершенность гештальта и желание повтора впечатления. Тем самым возникает эффект информационной наркотизации, зависимости от клиповых контентов. Среди представителей поколения Z, выросших в среде становящейся неконтролируемой цифровизации, уже сложно встретить молодежь, массово увлеченную классической музыкой или литературой, которые требуют времени восприятия, переживания и осознания информации. Зато процветает стремительный рэп. В данном случае наблюдается деформация антропологических профилей, в которых, однако, появляются новые способности к параллельной работе с многими видами деятельности, развитие интуиции и доверия ей, что становится эффективным в информационном хаосе. Более старшее поколение может входить в миры быстрого клипового мышления для поверхностного просмотра информации, но в период своего воспитания они уже приобрели все способности работы с информацией.

Еще одно безусловное достижение эпохи – поисковики информации, которые учат нас и сокращают наши усилия и время поиска, будучи предтечей современного ИИ. Обратной стороной успеха выступает воспитание интеллектуального инфантилизма и дефрагментации личной целостной картины мира, без которой сложный вопрос и не задашь, а поисковик покажет ответы из разных картин мира. Это наблюдается даже у элитных интеллектуалов Физтеха, которые частенько предпочитают забыть прошедший курс, т.к. его можно добыть в сети.

Обращаясь к социальным рискам ИИ, укажем, что, во-первых, интернет-провайдеры, используя данные о наших запросах в почте, мессенджерах и сети, имеют о каждом из нас огромные объемы информации, согласно которой легко восстановить ценности того или иного пользователя, увлечения, болезни, встречи, планы и т.д. Все это используют для незапрошенной навязчивой рекламы, продают банкам для оценки кредитоспособности лица, передают силовым и государственным структурам для составления гражданских рейтингов и следственных действий и т.д. Таким образом, с частными пространствами, видимо, в цифровую эпоху придется расстаться. Однако резко возникает опасность использования этой информации криминальными группами и го-

сударствами для реализации массовых принудительных действий, цифрового концлагеря, тень которого начала вырисовываться во время недавней пандемии.

Второй социальной деформацией можно признать появление аутичности и эскапизма существенной части населения, бегство в виртуальную реальность, миры игровых страстей и грез вселенных Цукерберга. Это порождает асоциальность, утрату способности к продолжению рода и массовые депрессии.

Третьей проблемой укажем резкое изменение востребованных профессий. Еще в прошлом десятилетии предполагалось, что к 2030 году роботизируются инструктивные рутинные профессии, а сохранены будут антропоцентричные и творческие профессии психологов, сиделок, учителей, людей искусства, ученых². Однако в течение последних трех лет стало понятным, что именно художники и литераторы становятся безработными, большие языковые модели последних версий могут не хуже выполнять их работу. На очереди – научные работники младшего и среднего звена, без которых и мэтры долго не продержатся.

Наконец, четвертая проблема. Цифровая цивилизация очень хрупка. Сверхмощной солнечной вспышки или использования ядерного взрыва в ближнем космосе будет достаточно для того, чтобы она закончилась. Поэтому особенно значимой видится цель удержания знаний на уровне личности и сохранения именно целостной личности. Возникает ощущение, что традиционное образование и культура останутся достоянием элит, а массам предложат цифровой рай и личностную деградацию.

Заключение

На основе проведенного исследования типов рациональности в контексте развития ИИ и формирования гибридных человеко-машинных социумов можно сделать ряд принципиальных выводов о перспективах и рисках этого развития. Речь идет не о противостоянии цифровизации и ИИ, но о необходимости формирования именно дружественного человеку ИИ. Однако эта задача оказывается нетривиальной, поскольку требует прояснения вопроса о том, какому именно человеку должен быть

² См., например: Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation // McKinsey Global Institute. December 2017. – URL: <https://www.mckinsey.com/~media/BAB489A30B724BECB5DEDC41E9BB9FAC.ashx>. P. 20.

дружественен ИИ. Как показало наше исследование, антропологические конституции не остаются неизменными, но активно трансформируются в цифровых средах под влиянием различных типов рациональности, особенно посткритической.

Эта ситуация ставит перед нами двойную задачу. С одной стороны, необходимым видится развитие институтов социальной экспертизы, способных оценивать и регулировать воздействие новых технологий на человека и общество [Пружинин 1986]. С другой – требуется более глубокое теоретическое осмысление природы возникающих гибридных социумов, в которых взаимодействуют ЕИ и ИИ, индивидуальные и коллективные субъекты, различные формы рациональности.

Особую важность в данном контексте приобретает проблема медиации между различными формами субъектности. Как показано в работах А.П. Давыдова [Давыдов 2021; Давыдов 2024], вопрос о взаимодействии индивидуальных и коллективных субъектов требует разработки специальных механизмов согласования и гармонизации. В случае с техносубъектами проблема осложняется тем, что ИИ обладает принципиально иной природой рациональности.

Представляется, что развитие ИИ в будущем должно ориентироваться на поиск измеримых критериев такой гармонизации. Это предполагает следующее: во-первых, разработку методов оценки влияния цифровых технологий на человека и социум; во-вторых, создание механизмов регулирования, преодоления рисков взаимодействия антропологических и технологических субъектов; в-третьих, формирование новых институциональных форм, обеспечивающих продуктивный диалог между ЕИ и ИИ, развитие креативной и аутопоэтической рациональности в рамках такого диалога.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Бахтин 1986 – *Бахтин М.М.* Проблема текста в лингвистике, филологии и других гуманитарных науках. Опыт философского анализа // *Бахтин М.М.* Литературно-критические статьи. – М.: Художественная литература, 1986. С. 473–500.

Буданов 2012 – *Буданов В.Г.* Постнеклассические практики и квантово-синергетическая антропология // Постнеклассические практики: опыт концептуализации / под общ. ред. В.И. Аршинова, О.Н. Астафьевой. – СПб.: Миръ, 2012. С. 37–63.

В.Г. БУДАНОВ. Техносубъект и антропосоциальные проблемы взаимодействия...

Буданов 2016 – Буданов В.Г. Новый цифровой жизненный техноуклад – перспективы и риски трансформаций антропосферы // Философские науки. 2016. № 6. С. 47–55.

Буданов 2022 – Буданов В.Г. Обобщенная научная рациональность: истоки, структура, перспективы в цифровую эпоху // Антропомерность как вызов и ответ современности / отв. ред. В.Г. Буданов. – Курск: Университетская книга, 2022. С. 44–63.

Буданов 2023а – Буданов В.Г. Антропосоциальные вызовы экспансии искусственного интеллекта // Ученые записки Института психологии Российской академии наук. 2023. Т. 3. № 2. С. 23–31.

Буданов 2023б – Буданов В.Г. Посткритическая рациональность как оборотная сторона цифровизации // Конструктивизм, реализм, человек. К 90-летию В.А. Лекторского / под ред. Е.О. Труфановой. – М.; СПб.: Центр гуманитарных инициатив; Университетская книга, 2023. С. 214–227.

Буданов 2023в – Буданов В.Г. Философия науки в цифровую эпоху: о перспективах постнеклассики // Третьи Степинские чтения. Перспективы философии науки в современную эпоху: материалы Междунар. конф. (Москва, 20–23 июня 2023 года) / отв. ред. В.А. Лекторский, В.Г. Буданов. – Курск: Университетская книга, 2023. С. 41–52.

Буданов 2023г – Буданов В.Г. Метафизика, онтология и сценарии Большого антропологического перехода // Наука и феномен человека в эпоху цивилизационного Макросдвига / отв. ред. В.Г. Буданов, Л.П. Киященко. – М.: Институт общегуманитарных исследований, 2023. С. 76–128.

Давыдов 2021 – Давыдов А.П. Методологическая «середина-для» в ракурсе неклассики В. Лекторского, медиации А. Ахiezера и принципа комплементарности Р. Гринберга / А. Рубинштейна // Вопросы философии. 2021. № 4. С. 191–202.

Давыдов 2024 – Давыдов А.П. Медиация и конвергентная социальность. К теории социального диалога // Философские науки. 2024. Т. 67. № 2. С. 135–159.

Латур 2014 – Латур Б. Пересборка социального: введение в акторно-сетевую теорию / пер. с англ. И. Полонской. – М.: Изд. дом Высшей школы экономики, 2014.

Лекторский 2021 – Лекторский В.А. О философских проблемах искусственного интеллекта и когнитивных исследований // Философские науки. 2021. Т. 64. № 1. С. 7–12.

Лекторский 2022 – Лекторский В.А. Искусственный интеллект в изучении человека, человек в мире, создаваемом искусственным интеллектом // Человек и системы искусственного интеллекта / под ред. В.А. Лекторского. – СПб.: Юридический Центр, 2022. С. 10–29.

Лысенко 2024 – Лысенко В.Г. Диалог эксперта по когнитивным наукам с большой языковой моделью Claude 3 Opus о самосознании

искусственного интеллекта: анализ буддолога // Философские науки. 2024. Т. 67. № 3. С. 75–98.

Соколова 2024 – *Соколова М.Е.* ChatGPT и промпт-инжиниринг: о перспективах внедрения генеративных нейросетей в науке // Научно-ведческие исследования. 2024. № 1. С. 92–109.

Степин 2011 – *Степин В.С.* Исторические типы рациональности в их отношении к проблеме сложности // Синергетическая парадигма: синергетика инновационной сложности / отв. ред. В.И. Аршинов. – М.: Прогресс-Традиция, 2011. С. 37–47.

Перлз 2010 – *Перлз Ф.С.* Гештальт-семинары. – М.: Институт общегуманитарных исследований, 2010.

Полани 1985 – *Полани М.* Личностное знание: на пути к посткристической философии / под ред. В.А. Лекторского, В.А. Аршинова; пер. с англ. М.Б. Гнедовского, Н.М. Смирновой, Б.А. Старостина. – М.: Прогресс, 1985.

Пружинин 1986 – *Пружинин Б.И.* Рациональность и историческое единство научного знания (гносеологический аспект). – М.: Наука, 1986.

Шютц 2003 – *Шютц А.* Смысловая структура повседневного мира: очерки по феноменологической социологии / пер. с англ. А. Я. Алхасова, Н. Я. Мазлумяновой. – М.: Институт Фонда «Общественное мнение», 2003,

Beghetto 2023 – *Beghetto R.A.* A New Horizon for Possibility Thinking: A Conceptual Case Study of Human×AI Collaboration // Possibility Studies & Society. 2023. Vol 1. No. 3. P. 324–341.

Danaher 2019 – *Danaher J.* The Philosophical Case for Robot Friendship // Journal of Posthuman Studies. 2019. Vol. 3. No. 1. P. 5–24.

Jain et al. 2023 – *Jain R., Garg N., Khera S.N.* Effective Human–AI Work Design for Collaborative Decision-Making // Kybernetes. 2023. Vol. 52. No. 11. P. 5017–5040.

Nyholm, Smids 2020 – *Nyholm S., Smids J.* Can a Robot Be a Good Colleague? // Science and Engineering Ethics. 2020. Vol. 26. No. 4. P. 2169–2188.

Peeters et al. 2020 – *Peeters M.M.M., van Diggelen J., Van Den Bosch K., Bronkhorst A., Neerincx M. A., Schraagen J.M., Raaijmakers S.* Hybrid Collective Intelligence in a Human–AI Society // AI and Society. 2020. Vol. 36. No. 1. P. 217–238.

REFERENCES

Bakhtin M.M. (1986) The Problem of Text in Linguistics, Philology, and Other Humanities. An Experience of Philosophical Analysis. In: Bakhtin M.M. *Literary Critical Articles* (pp. 473–500). Moscow: Khudozhestvennaya literatura (in Russian).

Beghetto R.A. (2023) A New Horizon for Possibility Thinking: A Conceptual Case Study of Human×AI Collaboration. *Possibility Studies & Society*. Vol. 1. No. 3, pp. 324–341.

Budanov V.G. (2012) Post-Non-Classical Practices and Quantum-Synergetic Anthropology. In: Arshinov V.I. & Astafyev O.N. (Eds.) *Post-Non-Classical Practices: Experience of Conceptualization* (pp. 37–63). Saint Petersburg: Mir (in Russian).

Budanov V.G. (2016) The New Digital Technological Lifestyle: Prospects and Risks of Anthroposphere Transformations. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. No. 6, pp. 47–55 (in Russian).

Budanov V.G. (2022) Generalized Scientific Rationality: Origins, Structure, Prospects in the Digital Age. In: Budanov V.G. (Ed.) *Anthropomeasurability as a Challenge and Response to Modernity* (pp. 44–63). Kursk: Universitetskaya kniga (in Russian).

Budanov V.G. (2023a) Anthroposocial Challenges of Artificial Intelligence Expansion. *Proceedings of the Institute of Psychology of Russian Academy of Sciences*. Vol. 3, no. 2, pp. 23–31 (in Russian).

Budanov V.G. (2023b) Post-Critical Rationality as the Reverse Side of Digitalization. In: Trufanova E.O. (Ed.) *Constructivism, Realism, Human. On the 90th Anniversary of V.A. Lektorsky* (pp. 214–227). Moscow; Saint Petersburg: Tsentr gumanitarnykh initsiativ; Universitetskaya kniga (in Russian).

Budanov V.G. (2023c) Philosophy of Science in the Digital Age: On the Prospects of Post-Non-Classics. In: Lektorsky V.A. & Budanov V.G. (Eds.) *Third Step in Readings. Prospects of Philosophy of Science in the Modern Era: Proceedings of the International Conference* (pp. 41–52). Kursk: Universitetskaya kniga (in Russian).

Budanov V.G. (2023d) Metaphysics, Ontologies and Scenarios of the Great Anthropological Transition. In: V.G. Budanov & L.P. Kiyashchenko (Eds.) *Science and the Human Phenomenon in the Era of Civilizational Macroshift* (pp. 76–128). Moscow: Institut obshchegumanitarnykh issledovaniy (in Russian).

Danaher J. (2019) The Philosophical Case for Robot Friendship. *Journal of Posthuman Studies*. Vol. 3, no. 1, pp. 5–24.

Davydov A.P. (2021) Methodological “Middle-for” from the Perspective of V. Lektorsky’s Non-Classics, A. Akhiezer’s Mediation, and R. Grinberg / A. Rubinstein’s Principle of Complementarity. *Voprosy filosofii*. No. 4, pp. 191–202 (in Russian).

Davydov A.P. (2024) Mediation and Convergent Sociality. Toward a Theory of Social Dialogue. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. Vol. 67, no. 2, pp. 135–159 (in Russian).

Jain R., Garg N., & Khera S.N. (2023) Effective Human–AI Work Design for Collaborative Decision-Making. *Kybernetes*. Vol. 52, no. 11, pp. 5017–5040.

Latour B. (2014) *Reassembling the Social: An Introduction to Actor-Network Theory* (I. Polonskaya, Trans.). Moscow: HSE Publishing House (Russian translation).

Lektorsky V.A. (2021) On the Philosophical Issues of Artificial Intelligence and Cognitive Studies. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. Vol. 64, no. 1, pp. 7–12 (in Russian).

Lektorsky V.A. (2022) Artificial Intelligence in Human Study, Human in the World Created by Artificial Intelligence. In: V.A. Lektorsky (Ed.) *Human and Artificial Intelligence Systems* (pp. 10–29). Saint Petersburg: Yuridicheskii Tsentri (in Russian).

Lysenko V.G. (2024) Dialogue on Artificial Intelligence's Self-Awareness Between the Cognitive Science Expert and Large Language Model Claude 3 Opus: A Buddhist Scholar's Perspective. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. Vol. 67, no. 3, pp. 75–98 (in Russian).

Nyholm S. & Smids J. (2020) Can a Robot Be a Good Colleague? *Science and Engineering Ethics*. Vol. 26, no. 4, pp. 2169–2188.

Peeters M.M.M., van Diggelen J., Van Den Bosch K., Bronkhorst A., Neerincx M.A., Schraagen J.M., & Raaijmakers S. (2020) Hybrid Collective Intelligence in a Human–AI Society. *AI and Society*. Vol. 36, no. 1, pp. 217–238.

Perls F.S. (2010) *Gestalt Seminars*. Moscow: Institut obshchegumanitarnykh issledovaniy (Russian translation).

Polanyi M. (1985) *Personal Knowledge: Towards a Post-Critical Philosophy* (V.A. Lektorsky & V.A. Arshinov, Eds.; M.B. Gnedovsky, N.M. Smirnova, & B.A. Starostin, Trans.). Moscow: Progress (Russian translation).

Pruzhinin B.I. (1986) *Rationality and Historical Unity of Scientific Knowledge (Epistemological Aspect)*. Moscow: Nauka (in Russian).

Schütz A. (2003) *The Meaningful Structure of the Social World: Essays on Phenomenological Sociology* (A.Ya. Alkhasov & N.Ya. Mazlumyanova, Trans.). Moscow: Institute of the Foundation “Obshchestvennoe mnenie” (Russian translation).

Sokolova M.E. (2024) ChatGPT and Prompt Engineering: On the Prospects of Implementing Generative Neural Networks in Science. *Science Studies = Naukovedcheskie issledovaniya*. No. 1, pp. 92–109 (in Russian).

Stepin V.S. (2011) Historical Types of Rationality in Their Relation to the Problem of Complexity. In: Arshinov V.I. (Ed.) *Synergetic Paradigm: Synergetics of Innovative Complexity* (pp. 37–47). Moscow: Progress-Traditsiya (in Russian).