

УДК 004.81;304.44

К ВОПРОСУ О ВКЛЮЧЕННОСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СИСТЕМУ СОЦИАЛЬНЫХ ОТНОШЕНИЙ

С.В. Шалагин¹, Г.Э. Шалагина²

¹Казанский национальный исследовательский технический университет
им. А.Н. Туполева – КАИ, г. Казань, Россия

²Казанский национальный исследовательский технологический университет,
г. Казань, Россия

Поступила в редакцию: 16.08.24

В окончательном варианте: 01.09.24

Аннотация. В работе ставится проблема антропометрической и безопасной включенности искусственного интеллекта (ИИ) в среду, созданную человеком в процессе общественного производства. С учетом развития информационных технологий компоненты ИИ (рецепторы, анализаторы и эффекторы) расширены до сенсоров, блоков принятия решений и исполнительных блоков. Указывается на социально-экономический характер аберрации когнитивного процесса при моделировании ИИ. Проблемы применения ИИ рассмотрены с позиций неокибernetики и эвристического подхода. В качестве возможного способа включения ИИ в систему социальных отношений смоделирована проекция авторского подхода на типологию социальных страт Платона. Сделаны выводы о возможных рисках и эффективности использования ИИ.

Ключевые слова: искусственный интеллект, информационное общество, социальные отношения, антропологическая экспертиза.

ON THE ISSUE OF ARTIFICIAL INTELLIGENCE INCLUSION IN THE SOCIAL RELATIONS SYSTEM

S.V. Shalagin¹, G.E. Shalagina²

¹Kazan National Research Technical University named after A.N. Tupolev,
Kazan, Russia

²Kazan National Research Technical University, Kazan, Russia

Original article submitted: 16.08.24

Revision submitted: 01.09.24

Abstract. The paper raises the problem of anthropometric and safe inclusion of AI in an environment created by man in process of social production. Taking into account the development of information technology, the components of AI behavior (receptors, analyzers and effectors) have been expanded to sensors, decision blocks and execution units. The socio-economic nature of the cognitive process aberration in AI modeling is indicated. The problems of AI application are considered from the standpoint of neo-cybernetics as well as heuristic approach. As a possible way to include AI in the system of social relations, the projection of the author's approach on Plato's typology of social strata is modeled. Conclusions about the possible risks and effectiveness of using AI are drawn.

Keywords: artificial intelligence, information society, social relations, anthropological expertise.

Информационные технологии сегодня являются базисом для развития искусственного интеллекта (ИИ) при реализации разнообразных форм человеческой деятельности. Согласно [1, с. 6] важным и представляющим интерес аспектом ИИ является его поведение как способ искусственной реализации разнообразных форм человеческой деятельности, а также как средство сделать антропомерными реакции искусственной среды. Под понятием «искусственная среда» будем понимать среду, созданную человеком в процессе исторического развития общественного производства. У поведения ИИ есть определенный порядок формирования [1, с. 25–28]: сначала информация поступает на рецепторы, обрабатывается, а далее сформированная реакция выдается на эффекторы. И главное – применение ИИ основано на принципиальной возможности адекватного компьютерного моделирования, адекватной «оцифровки» объектов из заданной области человеческой деятельности.

Классическая кибернетика в основном была ориентирована на управление, на выработку поведения для технических объектов, компьютерная модель которых формализуема по определению. Неокибернетика ориентирована на сложные киберфизические системы, действующие в социальном пространстве, на так называемые социокиберфизические системы [2, с. 83]. При формальном описании, моделировании отдельных аспектов сложных социальных систем наблюдаются когнитивный барьер и aberrации когнитивного процесса [3, с. 370–371]. Причины когнитивных проблем компьютерного моделирования зачастую носят не столько технический или парадигмально-научный, сколько социально-экономический характер [4, с. 157].

При обработке больших объемов данных со скоростью, недоступной людям ввиду объективных психофизиологических ограничений их нервной системы, ИИ, безусловно, будет полезен. Согласно [5, с. 13] ИИ является техническим, инструментальным продолжением человека. Всеобщая экспансия ИТ каузирует иллюзию их универсальности и способности разрешить любую проблему. Вместе с тем отмеченные возможности ИИ могут таить в себе различные угрозы. Они возникают из-за того, что в результате обработки колоссальных объемов информации со скоростью, недоступной людям, возрастает вероятность формирования поведения ИИ, которое будет как минимум неантропомерно, а в определенных случаях может содержать потенциальную угрозу для отдельных индивидов, их групп и всего человечества в целом. Чтобы не оказаться в фатальной зависимости от неантропомерного поведения ИИ, требуется выработать определенные подходы, связанные с обеспечением безопасного применения ИИ в человеческой деятельности.

ИИ – достаточно обширная тема для исследования. Среди различных направлений исследований в области ИИ мы рассматриваем технические и социальные аспекты эвристического (информационного) направления. Данное направление характеризуется исследованиями в области создания машинных способов решения интеллектуальных задач, для которых не имеет значения, как именно будут они (способы) устроены, насколько близки или далеки будут результаты к результатам, полученным людьми [6, с. 48].

Рассмотрим общую структуру технических систем ИИ (ИИ-систем), ориентированных на заданный круг задач в рамках эвристического (информационного) направления его развития. Согласно [1] для формирования поведения

примитивным ИИ-системам требуются три компонента: 1) рецепторы, 2) обработчики информации с фильтром данных, памятью и процессором, 3) эффекторы. С учетом развития информационных технологий, реализующих ИИ-системы, данные компоненты могут быть расширены до следующих блоков: 1) сенсоры; 2) блок(и) принятия решений (БПР); 3) исполнительные блоки (ИспБ).

Сенсор (или датчик) – это устройство, модуль, машина или подсистема, которые служат для обнаружения событий или изменений в определенной среде и для отправки данной информации в другие устройства, модули, машины или подсистемы. БПР – устройство, модуль, машина или подсистема, которая выдаёт один из множества вариантов решений при использовании сформированной базы данных (базы знаний), решающих правил и данных с сенсоров. ИспБ – устройство, модуль, машина или подсистема, которые передают воздействие с управляющего устройства на объект управления путем изменения поступающих на него энергии и/или материалов. Процесс принятия решения ИИ-системами выглядит следующим образом. Информация, требуемая для принятия решения, поступает через сенсоры. На сенсорах производится первичная фильтрация информации, которая затем поступает в БПР. Полученная информация сравнивается в БПР с некоторыми эталонными значениями, и по заданному правилу определяется возможное решение. Указанное решение поступает на ИспБ либо для передачи воздействия на управляемый объект, либо для блокировки такого воздействия.

Каковы же особенности включенности ИИ-систем в социальное пространство, в систему социальных отношений? ИИ позволяет существенно ускорить процессы принятия социально значимых решений. Причина ускорения заключается в следующем: всё большую роль в обществе станут играть акторы, направляемые или управляемые ИИ. Акторами могут быть как люди, ведомые «рекомендациями» алгоритмов, созданных на основе ИИ, так и сами алгоритмы ИИ, наделенные правами акторов. Человеческих индивидов, участвующих в социальных отношениях, будем называть естественными акторами, тогда как участников социальных отношений, которые находятся под контролем ИИ (роботов), будем называть акторами искусственными. Активное внедрение инструментов ИИ в социальную действительность, с одной стороны, и опыт осмысления социума в терминах акторно-сетевой теории Б. Латура [7], с другой стороны, предоставляет методологическую основу включения искусственных акторов в социальную структуру, в типологию социальных страт и классов. Искусственный интеллект сегодня поистине становится значимым социальным актором [8]. Например, в проекции нашего подхода на типологию социальных страт Платона рецепторные функции ИИ и эффекторы, влияющие на поведение людей, автономны друг от друга и располагаются в слое ремесленников (демоса); искусственные акторы, занятые обработкой информации с рецепторов, включены в страту стражей (воинов), реагирующих на заданные ситуации или состояния, а координацией этих функций на основе анализа информации, поступившей от эффекторов, занимаются «мыслящие» искусственные акторы, соседствующие с естественными акторами-мыслителями. Выбор платоновской модели обусловлен лаконичностью социальных составляющих, что характерно для обществ «закрытого»

типа. Современные типы обществ сложнее, но и там, основываясь на иных социально-философских моделях, можно проследить варианты включенности ИИ не только в искусственную среду, но и непосредственно в систему социальных отношений.

В случае генерирования социально значимых актов остро встает вопрос об антропомерности данной деятельности и таких социальных отношений. Указанная ситуация таит в себе как возможности для развития, так и всевозможные опасности – от потери ресурсов отдельных лиц до потери социально-экономических свобод индивидов и групп. Промежуточное положение занимают индивиды (естественные акторы), которые применяют ИИ в качестве системы поддержки принятия социальных решений. Следует понимать, что на данных индивидах лежит ответственность за то, какое влияние на социум оказывают решения, генерируемые под их руководством и выдаваемые в социум. Алгоритмы ИИ играют роль безответственных советников, которые, с одной стороны, позволяют сократить время выработки социально значимого решения, а с другой – сделать процессы социальных отношений неантропомерными со всеми вытекающими последствиями, опасными для естественных акторов. Ответственность за последствия решений всегда несет естественный актор, и фактически, и юридически.

Выходом из создавшейся ситуации может служить создание предохранительных механизмов внутри алгоритмов ИИ, основанных на автоматизированной оценке социальных решений, вырабатываемых алгоритмами, а также автоматизированной отбраковке заведомо неантропомерных социально значимых решений. В данной связи требуется формализовать понятие «антропомерность» для социально значимых решений, принимаемых в различных областях общественной жизни. Сами предохранительные механизмы, внедрённые в ИИ-систему, выполняют свою роль на трёх уровнях. На первом уровне подразумевается блокировка поступления на сенсоры информации, которая может каузировать выработку неантропомерных решений. На втором уровне БПР настроен таким образом, что заведомо не позволяет вырабатывать неантропомерные решения. На третьем уровне выработанные решения «тестируются» на антропомерность в ИспБ и блокируются в случае несоответствия критериям антропомерности. Предпосылки для решения данной задачи представлены в публикации [9]. Требуется адекватная антропологическая экспертиза и формализация процессов принятия решений алгоритмами ИИ, перманентное отслеживание аберрации когнитивного процесса на всех уровнях функционирования предохранительных механизмов [3; 4].

Таким образом, применение ИИ-систем в социальной жизни с учетом представленной методики позволит повысить эффективность и полезность принимаемых социально значимых решений, а наличие в ИИ-системах предохранительных механизмов даст возможность существенно минимизировать возможные негативные последствия, вызванные воздействием на общество неантропомерных решений, выработанных ИИ-системами.

В данной работе представлен подход авторов к вопросу о включенности искусственного интеллекта в систему социальных отношений, который позволяет развернуть дискуссию, а также приступить к отдельным (частным) и масштабным (общим) исследованиям по указанной теме.

Список литературы

1. *Гаазе-Рапопорт, М.Г.* От амебы до робота: модели поведения / М.Г. Гаазе-Рапопорт, Д.А. Поспелов. – Москва: Наука, гл. ред. физ.-мат. лит., 1987. – 288 с.
2. *Верзилин, Д.Н.* Неокибернетика: состояние исследований и перспективы развития / Д.Н. Верзилин, Б.В. Соколов, Р.М. Юсупов // Системный анализ в проектировании и управлении: сб. научных трудов XXIII Междунар. научно-практ. конф., Санкт-Петербург, 10–11 июня 2019 г. – Санкт-Петербург: ФГАОУ ВО «Санкт-Петербургский политехнический университет Петра Великого», 2019. – С. 81–98.
3. *Шалагин, С.В.* Когнитивные проблемы проектирования на основе компьютерных моделей: технический и социогуманитарный аспекты / С.В. Шалагин, Г.Э. Шалагина // Онтология проектирования. – 2016. – Т. 6. – № 3 (21). – С. 368–376.
4. *Шалагина, Г.Э.* Информационно-коммуникационные технологии как предмет социогуманитарных исследований / Г.Э. Шалагина, С.В. Шалагин // Вестник Московского государственного областного университета. Сер. Философские науки. – 2019. – № 2. – С. 154–163.
5. *Шалютин, С.М.* Искусственный интеллект: гносеологический аспект / С.М. Шалютин. – Москва: Мысль, 1985. – 199 с.
6. *Человек и системы искусственного интеллекта в военном деле / О.В. Масленников, Ф.К. Алиев, С.А. Беспалов, В.Е. Мишин // Военная мысль. – 2021. – № 6. – С. 46–56.*
7. *Латур, Б.* Пересборка социального: введение в акторно-сетевую теорию / Б. Латур. – Пер. с англ. И. Полонской; под ред. С. Гавриленко. – Москва: Изд. дом Высшей школы экономики, 2014. – 384 с.
8. *Шалагин, С.В.* Устойчивая организация в VUCA-мире: объективная и субъективная составляющие [Электронный ресурс] / С.В. Шалагин, Г.Э. Шалагина // VIII Садыковские чтения. XXI век: вариативность человеческого: матер. Междунар. научно-образоват. конф. (Казань, 19–20 ноября 2021 г.). – Казань: Изд-во Казанского ун-та, 2021. – С. 218–223 (дата обращения: 14.08.2023).
9. *Shalagin, S.* Concept Formalization in Designing: Roles of Natural and Artificial Actors / S. Shalagin, G. Shalagina // Lobachevskii J Math. – 2023. – № 44. – Pp. 757–763.

References

1. *Haase-Rapoport MG, Pospelov DA.* From amoeba to robot: models of behavior. Moscow: Science, Ch. ed. Physical-mat. lit., 1987. 288 p.
2. *Verzilin DN, Sokolov BV, Yusupov RM.* Neokibernetics: State of Research and Development Prospects. System Analysis in Design and Management: Sat. scientific works of the XXIII International. scientific practice. conf., St. Petersburg, June 10–11, 2019. St. Petersburg: St. Petersburg Polytechnic University of Peter the Great, 2019. Pp. 81–98.
3. *Shalagin SV, Shalagina GE.* Cognitive design problems based on computer models: technical and socio-humanitarian aspects. Ontology of design. 2016;3(21);368-376.
4. *Shalagina GE, Shalagin SV.* Information and communication technologies as a subject of socio-humanitarian research. Bulletin of Moscow State Regional University. Series: Philosophical Sciences. 2019;2;154-163.
5. *Shalyutin SM.* Artificial Intelligence: The Gnoseological Aspect. Moscow: Thought, 1985. 199 p.
6. *Maslennikov OV, Aliyev FK, Bespalov SA, Mishin VE.* Man and artificial intelligence systems in military affairs. Military thought. 2021;6;46-56.

7. *Latur B.* Re-assembly of social: an introduction to actor-network theory. Moscow: Ed. House of the Higher School of Economics, 2014. – 384 p.
8. *Shalagin SV, Shalagina GE.* Stable organization in the VUCA world: objective and subjective components. VIII Sadykov readings. XXI century: the variability of the human. Materials International. scientific and educational. conf. (Kazan, November 19–20, 2021) [Electronic resource]. Kazan: Publishing House of Kazan University, 2021. Pp. 218–223.https://kpfu.ru/portal/docs/F1326062238/_red._Sadykovskie_8.Shatunova._1_.pdf (accessed 14.08.2023).
9. *Shalagin S, Shalagina G.* Concept Formalization in Designing: Roles of Natural and Artificial Actors. *Lobachevskii J Math.* 2023;44:757-763.

Информация об авторах

ШАЛАГИН Сергей Викторович – доктор технических наук, доцент, профессор кафедры компьютерных систем ФГБОУ ВО «Казанский национальный исследовательский технический университет им. А.Н. Туполева – КАИ», г. Казань, Россия; eLibrary SPIN: 5718-4194.
E-mail: sshalagin@mail.ru

ШАЛАГИНА Гульнара Эдуардовна – кандидат философских наук, доцент, доцент кафедры философии и истории науки ФГБОУ ВО «Казанский национальный исследовательский технологический университет», г. Казань, Россия; eLibrary SPIN: 4187-2385.
E-mail: galanova@rambler.ru

Information about the authors

SHALAGIN Sergey V. – Doctor of Technical Sciences, Associate Professor of the Department of Computer Systems, Kazan National Research Technical University named after A.N. Tupolev – KAI, Kazan, Russia; eLibrary SPIN: 5718-4194. **E-mail:** sshalagin@mail.ru

SHALAGINA Gulnara E. – Candidate of Philosophical Sciences, Associate Professor of the Department of Philosophy and History of Science, Kazan National Research Technological University, Kazan, Russia; eLibrary SPIN: 4187-2385. **E-mail:** galanova@rambler.ru