

ЦИФРОВАЯ ТРАНСФОРМАЦИЯ ГОСУДАРСТВЕННОГО УПРАВЛЕНИЯ

Ю.Ю. Петрунин

РАЗВИТИЕ КОНЦЕПЦИИ СОЦИАЛЬНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Исследования в области искусственного интеллекта (ИИ) выявили ряд концептуальных недостатков и рисков разработанных когнитивных моделей и алгоритмов ИИ при применении их в разнообразных сферах деятельности. Это, в частности, непрозрачность, непроверяемость логики рассуждений и выводов ИИ, недружественность по отношению к человеку, и, как следствие, недоверие человека к результатам работы ИИ. Недоверие между искусственным и естественным интеллектами ведет к снижению эффективности как ИИ, так и его взаимодействия с человеческим интеллектом. В качестве преодоления данного кризиса непонимания была предложена концепция социального искусственного интеллекта (СИИ), которая включает в себя не только когнитивные аспекты интеллекта, но и социальные (совместно с эмоциональными). В статье подводятся итоги тридцатилетнего развития концепции СИИ, выделяются современные проблемы моделирования и изучения СИИ.

Ключевые слова: искусственный интеллект, социальный искусственный интеллект, общий искусственный интеллект, социальные вычисления, интернет вещей, эпистемологически-антропный шовинизм.

Research in the field of artificial intelligence (AI) has revealed a number of conceptual shortcomings and risks of the developed cognitive models and AI algorithms when applied in various fields of activity. This, in particular, is the opacity, non-verifiability of the logic of reasoning and conclusions of AI, not friendliness towards a person, and, as a result, distrust of a person in the results of AI work. Distrust between artificial and natural intelligence leads to a decrease in the effectiveness of both AI and its interaction with human intelligence. As a way to overcome this crisis of misunderstanding, the concept of artificial social intelligence (ASI) was proposed, which includes not only cognitive aspects of intelligence, but also social (together with emotional ones). The article sums up the results of thirty years of development of the concept of ASI, highlights the current problems of modeling and studying ASI.

* Петрунин Юрий Юрьевич — доктор философских наук, профессор, факультет государственного управления, МГУ имени М.В. Ломоносова; e-mail: petrunin@sps.msu.ru

Key words: artificial intelligence (AI), artificial social intelligence (ASI), artificial general intelligence (AGI), social computing, internet of things (IoT), epistemological-anthropic chauvinism.

Словосочетание «социальный искусственный интеллект», или «искусственный социальный интеллект» (СИИ, ASI), активно используется в научной литературе и на практике. Можно выделить несколько взаимосвязанных значений данного термина, каждое из которых детерминируется теми или иными проблемами, связанными с пониманием, возможностями использования и взаимодействием человека с искусственным интеллектом (ИИ, AI).

Первое значение СИИ определяется как использование алгоритмов и технологий искусственного интеллекта (машинное обучение, искусственные нейронные сети, генетические алгоритмы, нечеткая логика, экспертные системы и др.) в социально-гуманитарных науках: экономике, социологии, истории, политических науках, менеджменте, государственном управлении. В большинстве случаев ИИ используется как альтернатива или дополнение к традиционным методам статистического анализа соответствующей предметной области, таким как кластерный, дискриминантный, регрессионный анализы, прогнозирование временных рядов и др. Преимуществом методов ИИ является их нечувствительность к ограничениям, налагаемым на эмпирические данные статистикой или эконометрикой, а также способность обнаруживать слабые, неочевидные, нелинейные связи между переменными. Таким образом, алгоритмы и технологии ИИ расширяют возможности предиктивной аналитики, глубину познания социальной реальности.

В некоторых работах применение методов ИИ не ограничивается только использованием аналитических (вычислительных) инструментов и технологий в исследовании общества, но включает в себя также заимствование методологии и парадигмы ИИ. При этом из концепции ИИ социально-гуманитарные науки активно заимствуют аналогии, метафоры, когнитивные модели и мировоззренческие подходы¹.

¹ *Andreyuk D.S., Petrunin Yu., Shuranova A., Ushakov V.L.* Information agenda as an analogue of attention in sociomorphic neuronal networks // *Procedia computer science*. 2022. № 213. P. 292–295; *Петрунин Ю.Ю.* Биоморфные модели в экономических исследованиях // *Биомедицинская радиоэлектроника*. 2017. № 10. С. 80–81. *Петрунин Ю.Ю.* Нейрокомпьютерная парадигма и общество. М.: Издательство Московского университета, 2012; *Burrell J., Fourcade M.* The Society of Algorithms // *Annual Review Sociology*. 2021. № 47. С. 213–37; *Петрунин Ю.Ю.* Нейрогосударственное управление // *Нейрокомпьютеры: разработка, применение*. 2016. № 6. С. 10–11.

Во **втором значении** СИИ понимается как решение прикладных задач в социальной сфере с помощью ИИ: в медицине², управлении персоналом³, общественном и личном транспорте⁴, социальном обеспечении⁵, разработке и поддержании функционирования социального рейтинга⁶, в борьбе с фейками и защите персональных данных⁷, общественной безопасности⁸ и др.

Однако применение достижений ИИ при решении общественно-значимых задач несет не только позитивные результаты, но и определенные риски, и негативные последствия. Рассмотрим примеры данной коллизии.

Первый пример связан с государственным управлением. Во многих странах мира учреждения социального обеспечения внедряют новые способы работы, связанные с ИИ и анализом больших данных (Big Data) для оптимизации процессов предоставления индивидуальных услуг, сокращения случаев мошенничества и ошибок, формирования стратегических решений на фундаменте эмпирических данных. Эти способы дают существенные преимущества для повышения эффективности решения социальных задач. Автоматизация процессов ведет к упрощению обработки

² Карпов О.Э., Андриков Д.А., Максименко В.А., Храмов А.Е. Прозрачный искусственный интеллект для медицины // Врач и информационные технологии. 2022. № 2. С. 4–11; Kundu S. AI in medicine must be explainable // Nature Medicine. 2021. № 27 (8): 1328.

³ Петрунин Ю.Ю., Пугачева А.Е. Имплементация технологий искусственного интеллекта для отбора персонала российских компаний // Безопасность цифровых технологий. 2021. № 2. С. 97–121; Петрунин Ю.Ю., Силюянова Ю.А. Статистические и нейросетевые методы в исследовании управленческих проблем в организации // Нейрокомпьютеры: разработка, применение. 2018. № 10. С. 39–47.

⁴ Девятый номер журнала из списка ВАК «Автомобильные дороги» за 2022 г. был целиком посвящен использованию искусственного интеллекта для решения транспортных проблем. <https://elibrary.ru/contents.asp?id=49450560&selid=49450572> (дата обращения 06.01.2023).

⁵ Venouachane H. Artificial intelligence in social security: opportunities and challenges // Журнал исследований социальной политики. 2022. № 20 (3). С. 407–418.

⁶ Фёдоров М., Цветков Ю. Искусственный интеллект и социальный рейтинг: начало эпохи цифрового концентрационного лагеря «в интересах человечества»? 25 июня 2021. <https://russiancouncil.ru/analytics-and-comments/analytics/iskusstvennyy-intellekt-i-sotsialnyy-reyting-nachalo-epokhi-tsfirovogokonsentratsionnogo-lagerya-v/> (дата обращения 06.01.2023).

⁷ Искусственный интеллект — наш помощник в решении социальных проблем // Курьер ЮНЕСКО. 2019. № 3.

⁸ Духихин В.В., Иванов А.С., Межуев И.Ю., Шоков А.В., Якупов И.Ю. Распознавание радиолокационных изображений, полученных в миллиметровом диапазоне, с использованием нейронных сетей // Вестн. Концерна ВКО «Алмаз — Антей». 2022. № 3. С. 48–58.

данных, централизации информации и уменьшения потребности во взаимодействии с человеком⁹.

С другой стороны, работа ИИ, как правило, связана с непрозрачностью предлагаемых решений и часто создает потенциальные этические и коммуникационные проблемы. Из-за этого приходится приостанавливать развертывание инструментов ИИ¹⁰. Эти этические проблемы могут приводить к результатам, которые являются необъяснимыми, несправедливыми, непрозрачными и/или предвзятыми по отношению к определенной группе пользователей данными услугами.

Фактически, прозрачность и объяснимость применения ИИ представляет собой важный вопрос, особенно в отношении решений, которые влияют на людей и/или сопряжены с социальными рисками. Известны случаи, когда внедрение интеллектуальных технологий в систему социального обеспечения приводило к росту количества жалоб со стороны населения на 87%¹¹!

Второй пример связан со здравоохранением, где имплементация технологий ИИ идет нарастающими темпами. Зарубежные и отечественные врачи отмечают отсутствие прозрачности в работе современных медицинских систем, использующих ИИ. «Даже когда традиционный «непрозрачный» (для человека! — Ю.П.) ИИ может выявить закономерность, указывающую на неизбежность заболевания, мы обычно не можем объяснить логику, лежащую в основе этого решения. Существуют технологические причины, по которым создание объяснимых СИИ является сложной задачей; «логика черного ящика» остается камнем преткновения»¹².

«Насколько можно доверять таким решениям и быть уверенным в том, что ИИ не ошибется в критической ситуации, когда на кону стоит здоровье и даже жизнь человека?»¹³ — вопрошают

⁹ ISSA Data-Driven Innovation in Social Security: Good Practices from Asia and the Pacific. Available at: <https://ww1.issa.int/analysis/data-driven-innovation-social-security-good-practices-asia-and-pacific> (accessed 06 Jan 2023).

¹⁰ Capgemini Research Institute. Why Addressing Ethical Questions in AI will Benefit Organizations. 2019. Available at: https://www.capgemini.com/gb-en/wp-content/uploads/sites/5/2022/05/Ethics-in-AI-Infographic_Web.pdf (accessed 06 Jan 2023).

¹¹ Nott G. Ombudsman: Centrelink OCI Lacking Usability and Transparency // Computerworld. 2017. Available at: <https://www.computerworld.com/article/3476375/ombudsman-centrelink-oci-lacking-usability-and-transparency.html> (accessed 06 Jan 2023).

¹² Карпов О.Э., Андриков Д.А., Максименко В.А., Храмов А.Е. Прозрачный искусственный интеллект для медицины // Врач и информационные технологии. 2022. № 2. С. 7.

¹³ Там же. С. 7.

врачи. Оптимальным выходом, считают они, может стать концепция «дополненного интеллекта», предложенная еще У.Р. Эшби¹⁴ и Фридманом¹⁵, или «ассистирующего ИИ» в медицине¹⁶.

Еще один пример взят из задачи отбора персонала в коммерческие компании с помощью ИИ. Работники отделов кадров отечественных компаний отмечают, что среди главных проблем использования технологий ИИ специалисты выделяют невозможность отследить работу алгоритмов, на основе которых даются рекомендации о трудоустройстве. В свою очередь указанная непрозрачность часто приводит к нежеланию кандидатов общаться с роботом¹⁷.

Об этих же проблемах пишут и юристы¹⁸, и социологи («люди еще не привыкли к этому виду интеллекта»¹⁹), и представители других общественных наук.

Описанные примеры подводят к выводу, что не(до)понимание между ИИ и человеком не только снижает эффективность ИИ, но и может привести к негативным результатам его использования. Очевидно, что решение проблемы требует теоретической и практической разработки механизмов согласованности/когерентности разных видов интеллекта.

Третье значение как раз определяет СИИ как концепцию, оптимизирующую взаимодействие естественного и искусственного интеллекта, как дополнение машинного/вычислительного интеллекта социальным измерением, делающим возможным понимание искусственного интеллекта (интеллектуального агента) естественным интеллектом. Последнее значение можно считать фундаментом для решения первых двух проблем и является важным элементом общей теории искусственного интеллекта (AGI).

¹⁴ Ashby W.R. An Introduction to Cybernetics. Chapman & Hall Ltd., L., UK, 1957.

¹⁵ Friedman C.P. A “fundamental theorem” of biomedical informatics // J. Am. Med. Inform. Assoc. 2009; 16: 169–170.

¹⁶ Карпов О.Э. и др. С. 7.

¹⁷ Петрунин Ю.Ю., Пугачёва А.Е. Имплементация технологий искусственно-го интеллекта для отбора персонала российских компаний // Безопасность цифровых технологий. 2021. № 2 (101). С. 97–121. DOI: 10.17212/2782-2230-2021-2-97-121. С.108-109

¹⁸ Харитонов Ю.С., Ян Т. Рекомендательные системы цифровых платформ в Китае: правовые подходы и практика обеспечения прозрачности алгоритмов // Закон. 2022. № 9. С. 40–49; Харитонов Ю.С., Савина В.С., Паньини Ф. Предвзятость алгоритмов искусственного интеллекта: вопросы этики и права // Вестн. Пермского ун-та // Юридические науки. 2021. № 3 (53). С. 488–515.

¹⁹ Цвык В.А., Цвык И.В. Социальные проблемы развития и применения искусственного интеллекта // Вестн. РУДН. Сер. Социология. 2022. Vol. 22. №. 1. С. 60.

Становление концепции СИИ в контексте ИИ

Искусственный интеллект (ИИ, AI) — междисциплинарное направление науки. В его разработке с момента возникновения участвовали кибернетика, информатика (computer science), биология, нейрофизиология, математика, логика, инженерно-технические науки, лингвистика, философия. В последнее время к ним присоединились юриспруденция, экономика, социология, политология, менеджмент, государственное управление и многие другие науки. Даже традиционные дисциплины, участвующие в становлении ИИ, сместили свои акценты: если раньше вклад биологии ограничивался изучением нейронных сетей, то теперь в неменьшей степени ее интересует поведение живых организмов (муравьиный алгоритм, природные алгоритмы), эволюционные процессы (эволюционное программирование, генетический алгоритм), в философии фокус исследования ИИ сместился с методологии на этику.

Создавая ИИ, нельзя забывать, что в мире уже есть как минимум один вид интеллекта — человеческий. Практика показала, что они работают (решают задачи) по-разному. Уже на старте ИИ геометрические доказательства интеллектуальной программы отличались от доказательств человека. То же мы наблюдаем в переводе с языка на язык, понимании речи, анализе данных при принятии решений, обучении, лечении больных, рекомендациях маркетинговых стратегий. Тем не менее, результаты во многих случаях являются сопоставимыми по эффективности, а часто — даже лучше, чем у Homo Sapiens. Парадокс, но машинное обучение в компьютерных играх открывает нам такие решения, которое не знали даже создатели программы²⁰.

Можно сколько угодно говорить о том, что ИИ не существует или даже не может существовать (хотя никто еще не доказал ни возможность, ни невозможность его существования)²¹, но фактом остается то, что технологии «ИИ» решают множество важных практических задач. Хотим ли мы сделать «ИИ» более полезным и эффективным? Тогда целесообразней не спорить о терминологии, а улучшать ИИ в направлении понятности, интерпретируемости, корректности, прозрачности и доверительности его выводов. Более того, если осознанно не заниматься поиском понимания разных

²⁰ Mnich V., Kavukcuoglu K., Silver D. et al. Human-Level Control Through Deep Reinforcement Learning // Nature 518. 2015. No. 7540. P. 529–33.

²¹ Когда-то античный философ Платон сомневался, что человек обладает интеллектом, считая, что интеллект присущ только богу. Платон. Законы // Соч. в четырех томах. Т. 4. М., 1995. С. 93.

видов интеллекта (человеческого и машинного/компьютерного), то не мы (люди) будем приспосабливать ИИ под себя, а он нас под свои стандарты (расчеловечивать?). Что, собственно, уже наблюдается в повседневной жизни. Интересной иллюстрацией этого является подзаголовок статьи про возможности полной автоматизации математических рассуждений: «Инструменты искусственного интеллекта формируют средства доказательства теорем следующего поколения, а вместе с ними и отношения между математикой и машинами»²².

Из данного факта следует, что ИИ не просто технический механизм или технологический процесс, **но автономный субъект социального действия**. Понимание этого делает необходимым создавать не только общую *рациональную (когнитивную)* концепцию ИИ, но и концепцию *социального* искусственного интеллекта. В качестве синонимов можно назвать «социально ответственного ИИ», «понятого ИИ», «доверительного ИИ» и т.п. «Социальный интеллект — это общий термин на стыке различных дисциплин, включая философию, социологию, экономику, юридические науки, психологию и т.д., а также информатику. В широком смысле социальный интеллект — это способность понимать других и действовать рационально и эмоционально в отношениях с другими. Это способность, которой обладают не только люди, но и искусственные агенты, моделируемые искусственным интеллектом»²³.

Термин Artificial Social Intelligence (ASI) появился почти 30 лет назад²⁴. В нашей стране схожая тема начала обсуждаться относительно недавно²⁵. Социальный интеллект в научном сообществе определяется как набор/система индивидуальных знаний об

²² Ornes S. How Close Are Computers to Automating Mathematical Reasoning? AI tools are shaping next-generation theorem provers, and with them the relationship between math and machine // Quanta Magazine. 2020.08.27. <https://www.quantamagazine.org/how-close-are-computers-to-automating-mathematical-reasoning-20200827/>

²³ Herzig A., Lorini E., Pearce D. Social Intelligence // AI & SOCIETY (2019) 34: 689. <https://doi.org/10.1007/s00146-017-0782-8>

²⁴ Bainbridge W.S., Brent E.E., Carley K.M., Heise D.R., Macy M.W., Markovsky B., Skvoretz J. Artificial social intelligence // Annual Review of Sociology. 1994. 20. P. 407–436. <https://doi.org/10.1146/annurev.so.20.080194.002203>

²⁵ Резаев А.В., Трегубова Н.Д. Готовы ли социологи к анализу «искусственной реальности»? Проблемы и перспективы исследований искусственного интеллекта в социальных науках // Мониторинг общественного мнения: экономические и социальные перемены. 2018. № 5. С. 91–108; Тавокин Е.П. Искусственность «искусственной социальности» // Социологические исследования. 2019. № 6. С. 115–122. DOI: 10.31857/S013216250005487-8; Алехина Т.А. «Искусственная социальность»: к вопросу о дискуссиях // Социальные и гуманитарные знания. 2021. Т. 7. № 1. С. 52–65.

обществе, социальных навыках и компетенциях. В настоящее время считается, что социальный интеллект тесно связан с когнитивным интеллектом и эмоциональным интеллектом.

Концепции социального и эмоционального интеллекта²⁶ как элементов человеческого интеллекта дополняют представления об интеллекте, заложенные в работах Н. Винера, А. Тьюринга и других пионеров ИИ. Эти элементы приобретают все большее значение по мере того, как человек в информационном обществе все чаще воспринимает машины как социальные сущности. Реализация искусственных агентов с сенсорными датчиками и модулями обработки сенсорной информации представляет собой практический пример технологии с некоторыми эмоциональными и социальными навыками.

Концепция СИИ сфокусирована на развитие способности интеллектуального агента сотрудничать с членами команды людей. Агентам, наделенным искусственным социальным интеллектом, потребуются различные возможности для сбора социальных данных, необходимых для взаимодействия с человеческими коллегами, умение моделировать, интерпретировать и предсказывать поведение и психические состояния людей для поддержки человека-агента. Искусственному социальному интеллекту потребуются возможности обработки, чтобы воспринимать, интерпретировать и генерировать комбинации социальных сигналов для работы в команде человек-агент. СИИ предлагает структуру, с помощью которой социально разумный агент может быть наделен способностью моделировать человеческих двойников и участвовать в эффективном взаимодействии человека и агента. Кроме того, моделирование искусственного разума может использоваться ASI для поддержки прозрачного общения с людьми, повышения доверия к агентам, чтобы они могли лучше предсказывать будущее поведение системы на основе своего понимания и поддержки доверия к искусственным интеллектуальным агентам²⁷.

Центральным вопросом исследований в области СИИ является вопрос о том, могут ли роботы быть социально разумными? В об-

²⁶ Среди относительно недавних обобщающих работ на эту тему можно выделить публикации *Шиллер А.В.* От теорий к моделям эмоций для искусственного интеллекта — основные методологические вопросы // *Ценности и смыслы*. 2018. Т. 4. № 56. С. 126–137; *Шиллер А.В.* Методологические основания моделирования эмоций в архитектуре искусственного интеллекта. Дисс. ... к.ф.н. М.: МГУ имени М.В. Ломоносова. 2019.

²⁷ *Williams J., Fiore S. M., Jentsch F.* Supporting Artificial Social Intelligence with Theory of Mind // *Frontiers in Artificial Intelligence*. 2022-02-28. DOI: 10.3389/frai.2022.750763

ласти искусственного интеллекта и робототехники социальный интеллект роботов появился сравнительно недавно. Однако стало ясно, что социальные и интерактивные способности необходимы в самых разных областях применения и контекстах, где роботы должны взаимодействовать и сотрудничать с другими роботами или людьми. Исследования взаимодействия человека и робота поднимают множество вопросов о природе интерактивности и «социального поведения» роботов и людей.

СИИ тесно связан с эмоциональным интеллектом и так называемыми аффективными вычислениями. Аффективные вычисления — это широкая область исследований, охватывающая системы, которые распознают, интерпретируют, обрабатывают или моделируют человеческие эмоции, настроения и чувства. Например, виртуальные голосовые помощники могут быть запрограммированы на шутивную беседу; это делает их более восприимчивыми к эмоциональной динамике человеческого взаимодействия, распознаванию испытываемых человеком эмоций, что облегчает взаимодействие человека с компьютером, а иногда и предупреждая о возможных рисках²⁸. В широко известных умных помощниках Siri или Алисе это неплохо реализовано: и предугадывание желаний собеседника, и изменение эмоциональной модальности разговора и многое другое.

Конечно, это может создавать у некоторых пользователей (людей) нереалистичное представление о том, насколько разумны существующие компьютерные агенты. Более того, эмоциональные вычисления в целом дают неоднозначные результаты. Мультимодальный анализ настроений, использующий ИИ для классификации эмоций, являются примером проектов такого типа²⁹.

Современное состояние исследований в области СИИ

В последние годы сформировалась новая вычислительная парадигма, известная как социальные вычисления. Она фокусируется на интеграции социального измерения с вычислительными системами. Социальные вычисления направлены на то, чтобы

²⁸ Барышев Д.А., Макаревич И.В., Зубанков А.С., Розалиев В.Л. Нейросетевой подход к определению эмоций человека по речи // Инженерный вестник Дона. 2022. № 5; Галичий Д.А., Афанасьев Г.И., Нестеров Ю.Г. Распознавание эмоций человека при помощи современных методов глубокого обучения // E-Scio. 2021. № 5 (56).

²⁹ How to Make Socially. Intelligent Robots Using AI? By Dr Nivash Jeevanandam. AI Concepts. Mar 15, 2022. India AI.

интеллектуальные системы могли общаться с пользователем и понимать его социальный контекст. Именно стремительный прогресс в применении искусственного интеллекта в социальных вычислениях привел к появлению практически реализуемого СИИ. У СИИ есть большой потенциал для решения проблемы адаптации социальных отношений в команде интеллектуальных агентов, поскольку он подходит к этой проблеме с точки зрения социальных вычислений, в отличие от обычного ИИ³⁰.

Еще одним явлением, повлиявшим на проблематику СИИ, стала другая современная технология — интернет вещей (Internet of Things, IoT). Огромный объем данных, генерируемых устройствами IoT, становятся богатым источником социальных свойств пользователя. Анализируя созданные пользователями следы социальных данных, записанные в различных ситуациях физического контакта и общения между пользователями (например, местоположение, журнал электронной почты, записи вызовов и текстовые сообщения), эмоциональные состояния и поведенческие модели, интеллектуальные устройства могут распознавать социальный контекст пользователей. Соответственно, они будут предлагать персонализированные услуги и индивидуальный контент в соответствии с их социальным контекстом. Обычные технологии анализа данных ИИ должны работать одновременно с СИИ, чтобы максимизировать знания системы о социальных аспектах пользователей. Применение обычного ИИ к отпечатку социальных данных пользователя с помощью систем IoT может помочь распознавать социальные функции и данные социального контекста³¹.

Роботы, виртуальные помощники и другие виды агентов, наделенных искусственным интеллектом, все чаще внедряются и будут дальше внедряться во многие отрасли, такие как государственное управление, здравоохранение, образование, транспорт, спорт, компьютерные игры и это лишь некоторые из многих областей, которые трансформируются благодаря достижениям в технологиях. ИИ также становится полезным во многих приложениях, от медицины до бизнеса.

С этой точки зрения «новый робот с ИИ» — это больше, чем «компьютер на колесах», как это воспринималось ранее в ИИ. Но-

³⁰ Williams J., Fiore S.M., Jentsch F. Supporting Artificial Social Intelligence with Theory of Mind // *Frontiers Artificial Intelligence*. 2022. 5:750763. <https://doi.org/10.3389/frai.2022.750763>

³¹ Sahraoui Dhelim, Huansheng Ning, Fadi Farha, Liming Chen, Luigi Atzori, Mahmoud Daneshmand. IoT-Enabled Social Relationships Meet Artificial Social Intelligence // *IEEE Internet of Things Journal*. 2021. № 1. arXiv: 2103.01776v2 [cs.CY] 25 May 2021.

вый робот с искусственным интеллектом олицетворен, учитывает ситуацию, отзывчив и восприимчив к обстановке³².

Технологические решения «старого ИИ», как правило, полагались на модели черного ящика, о чем уже говорилось ранее. Они были непрозрачны и плохо интерпретируемы. Это привело к развитию новых областей исследований, таких как «объяснимый искусственный интеллект» (Explainable AI, XAI). Его цель состоит в том, чтобы либо сделать основные процессы принятия решений более прозрачными, либо создать методы для преобразования, например, результатов машинного обучения во что-то понятное.

Отсутствие прозрачности в этих моделях способствует снижению доверия интеллектуальных агентов. Это очень важно, поскольку доверие признается специалистами важнейшей и весьма желательной характеристикой взаимодействия человека и интеллектуального агента. Понимание решений агентов способствует развитию доверия и имеет основополагающее значение для поддержки взаимодействия между пользователями и искусственными агентами. Обеспечение пользователей точным восприятием возможностей и намерений агентов, а также информирование их о контекстуально значимых ограничениях и других соответствующих знаниях, связанных с надежностью системы, будет становиться все более необходимым по мере развития автономных возможностей систем ИИ.

Если человек не доверяет агенту, он не будет делегировать ему задачи, независимо от того, действительно ли агент способен их решить, что приводит к неправильному/не эффективному использованию системы ИИ. Кроме того, чрезмерная зависимость от системы может привести к самоуспокоенности человека и невозможности обнаружить системные ошибки ИИ. Неправильная калибровка человеческого доверия к ИИ и, как следствие, необоснованная зависимость от агента, приводит к снижению производительности команд человек-агент. Таким образом, доверие является ключевым фактором взаимодействия человека и агента и влияет на то, будет ли человек полагаться на систему. Понимание людьми поведения и решений системы ИИ влияет на доверие человека к системе.

Повышение доверия, эффективное сотрудничество человека и агента требует, чтобы агент обладал социальным интеллектом, т.е. коммуникативными и социальными навыками, необходимыми агентам для поддержания эффективного обмена социальной информацией.

³² How to Make Socially. Intelligent Robots Using AI? By Dr Nivash Jeevanandam // AI Concepts. Mar 15, 2022. India AI.

Проблемы изучения и моделирования СИИ

1. Оценивание и измерение СИИ

Правильное оценивание и измерение социальных навыков является непростой задачей для исследователей. Хотя метрики и инструменты, с помощью которых можно количественно оценить социальный интеллект, были давно разработаны, не существует общепризнанного надежного инструмента для определения/измерения социального интеллекта. Например, методологии измерения СИИ часто требуют опроса нескольких человек (учителей, коллег, руководителей) для оценки одного человека, что делает каждую операцию оценивания/верификации сложной и трудоемкой. Кроме того, стандартные методы измерения (самоотчет и оценка поведения наблюдателями) часто не коррелируют друг с другом³³.

Существуют фундаментальные проблемы при попытке сравнения методологии изучения/измерения социального интеллекта с другими видами тестирования интеллекта. Например, использование методов самооценки для измерения социального интеллекта на индивидуальном уровне заметно отличается от традиционных методов оценки, применяемых при тестировании интеллекта. Однако это абсолютно необходимо для всесторонней оценки различных способов его использования в различных социальных взаимодействиях и контекстах. Поэтому методы определения компетентности в области социального интеллекта основываются на нескольких источниках, включая оценки сверстников, начальников и самооценки. Таким образом, множество различных компонентов социального интеллекта и разнообразие методов, используемых для его измерения, хотя и сложны, но помогают отличить его от других форм интеллекта.

2. Понимание намерений

Понимание намерений — сложная задача, поскольку даже у людей компетенция этого понимания сильно варьируется в зависимости от уровня развития их социального интеллекта. При взаимодействии человека с машиной могут возникать сбои, потому что когнитивный ИИ пренебрегает важностью социального интеллекта для оценки намерений. Для этого требуется эффективный обмен социальной информацией, что требует от агента способности понимать значение социальной информации, передаваемой

³³ *Silvera D., Martinussen M., Dahl T.I.* The Tromsø Social Intelligence Scale, a self-report measure of social intelligence // *Scand J. Psychol.* 2001. 42, 313–319. doi: 10.1111/1467-9450.00242

людьми, и способности передавать свою собственную интерпретируемую социальную информацию, устанавливая двунаправленность социальных коммуникаций. Задача усложняется, если учитывать многообразные контекстуальные факторы. Таким образом, успешное функционирование в сотрудничестве с людьми требует от агентов, наделенных СИИ, изучения социальных, культурных, моральных норм и неявных знаний, лежащих в основе понимания любой социальной ситуации. Кроме того, требуется, чтобы это предварительное знание могло быть использовано для вывода о том, что чувствуют и чего желают другие, и для прогнозирования их поведения.

Чтобы агент продемонстрировал социальный интеллект, он должен понимать, что намерения, чувства, психические состояния, личные особенности и другие качества человека могут быть воплощены через поведение. Таким образом, как и люди, агенты получают непрерывный поток сигналов во время взаимодействий, которые воспринимаются и анализируются. Это помогает разумно обосновать взаимодействие, чтобы намерения непосредственных действий человека могли быть распознаны другими агентами, и чтобы можно было делать прогнозы их ближайших и отдаленных действий. В более широком смысле эту информацию можно использовать для прогнозирования характера и качества социальных отношений, а также для определения общей атмосферы взаимодействия. Таким образом, вычислительный социальный интеллект требует понимания того, как сигналы воплощаются в действиях агентов.

3. Понимание социального и культурного контекста

СИИ должен быть в состоянии придерживаться контекстуально обусловленных социальных структур, таких как кодексы поведения, ритуальные формы общения, почетные знаки, традиции и культурные нормы. Однако необходимые знания и возможности СИИ могут регулироваться в зависимости от ситуативных контекстов. Культура, например, может операционализироваться как набор функций с различными параметрами, связанными с количественными социальными нормами. Например, количественно особенности культуры могут выражаться в интервалах межличностных расстояний, частотности словоупотребления и специфики лексики при общении в зависимости от речевого стиля или формата общения и т.д.³⁴.

³⁴ Hofstede G.J. GRASP agents: social first, intelligent later // AI Soc. 2019 34, 535–543. doi: 10.1007/s00146-017-0783-7

В этом направлении был разработан алгоритмический подход к социальным показателям, санкционированным культурой (Culture Sanctioned Social Metrics — CSSMs), в котором функции действия-воздействия изменяются в зависимости от контекста. В дальнейшем на этой основе была разработана вычислительная модель социальных норм³⁵. Это интегрирует культурные ценности и определяет их количественно с использованием ранее разработанной модели CSSM. Основываясь на ней, агент может максимизировать успешное внедрение и введение в действие норм во время социальных взаимодействий, одновременно учитывая, как это действие влияет на других агентов и его самого. Моделирование глубоких культурно обоснованных наборов убеждений/знаний особенно актуально для определения ситуации, в которой СИИ может способствовать объединению агентов в команду посредством активного взаимодействия. Принимая во внимание спектр задач, в которых преуспевают искусственные агенты и в которых участвуют люди, вполне вероятно, что СИИ может принести максимальную пользу в сценариях совместной работы, где потребность в социально обоснованном вмешательстве использует тот вид решения проблем, который требует больших вычислительных ресурсов, в котором ИИ уже чувствует себя как рыба в воде. Одним словом, СИИ должен быть полезным, жизнеспособным и применимым в широком диапазоне ситуаций. В отдельных случаях воплощение СИИ требует дополнительных конструктивных реализаций, таких как предоставление возможностей для выражения лица с помощью цифрового или физического дисплея или тембра голоса. Это увеличивает возможности СИИ, поскольку социальные компоненты могут существовать даже в виртуальной или резко упрощенной физической среде³⁶.

4. Заимствование «негативной» («злокачественной») социальности

Представление о создании когнитивного ИИ, очищенного от эмоций, социальных представлений и стереотипов не только не реалистично, но и не соответствует контексту возникновения ИИ. Истоки ИИ не ограничиваются аналитической философией, чистой математикой и Аланом Тьюрингом, но в меньшей степени вос-

³⁵ Bölöni L., Singh Bhatia T., Khan S.A., Streater J., Fiore, S.M. Towards a computational model of social norms // PLoS ONE. 2018. 13, e0195331. doi: 10.1371/journal.pone.0195331

³⁶ Hofstede G.J. GRASP agents: social first, intelligent later // AI Soc. 2019. 34, 535–543. doi: 10.1007/s00146-017-0783-7

ходят и к истории государственного управления³⁷. Обе системы обрабатывают большие объемы информации, используют иерархию заранее установленных, но адаптируемых правил. Система государственного управления в виде бюрократической машины предшествует системе машинной. История указывает на явную связь между организацией человеческих социальных структур и процессов (государства и корпораций) и цифровыми инструментами, предназначенными для их обслуживания. Замыслы Бэббиджа по созданию вычислительной техники в 1820-х гг. некоторые ученые рассматривают как «материализацию государственной деятельности»³⁸.

Аналогично российский чиновник С.Н. Корсаков в 1832 г. разрабатывал механизм поиска информации — прототип того, что сейчас называют нейрокомпьютером, — для решения административно-аналитических задач имперской бюрократии³⁹. Как пишет Дж. Пенн об одном из основателей исследований в области ИИ и Нобелевском лауреате по экономическим наукам Герберте Саймоне, «чтобы компьютер «думал» как человек, Саймон заставил его думать, как корпорация (государственная или частная)⁴⁰.

Имплицитно в строго аналитические модели ИИ с самого начала проникали социальные элементы, причем часто передающие ему не только позитивные, созидающие социальные импульсы, но и негативные, разрушительные.

И. Бахтина пишет: «Вольное или невольное закрепление сложившейся в обществе культуры притеснения тех или иных социальных групп или подрыв уважения к человеческому достоинству сегодня можно обнаружить в самых разнообразных системах алгоритмов, созданных по шаблонам, принятым в бизнес-среде. Поголовной этической экспертизы (и прививки) для всех создаваемых нейросетей пока не предусмотрено»⁴¹.

³⁷ Agar J. The government machine: a revolutionary history of the computer. The MIT Press Cambridge, Massachusetts L., England. 2003.

³⁸ Там же.

³⁹ Корсаков С.Н. Начертание нового способа исследования при помощи машин, сравнивающих идеи / Пер. с франц. под ред. А.С. Михайлова. М.: МИФИ, 2009. 44 с.

⁴⁰ Penn J. AI thinks like a corporation—and that's worrying // The Economist. Nov 26th 2018. https://www.economist.com/open-future/2018/11/26/ai-thinks-like-a-corporation-and-thats-worrying?__twitter_impression=true

⁴¹ Бахтина И. Кого копирует искусственный интеллект? // Колонка Ирины Бахтиной «Блеск и нищета корпоративных брендов». 15 декабря 2018 г. <https://rben.ru/corpequities-page/967-2018-11-12>

Бахтина приводит историю, произошедшую с ее коллегой, которая поделилась «недоумением по поводу того, как корпорация Apple в ходе опроса потребительского мнения потеряла к ней интерес, как только узнала ее возраст». «Очевидно, довольный собой, — продолжает автор, — и совсем не любознательный алгоритм оказался не в состоянии предположить, что “лица старше 50” способны приобретать гаджеты Apple — и не только для собственного пользования, а для всех членов семьи» и далее о том, «какое чувство брезгливости вызывают рекомендации брендов, основанные на весьма бесхитроном и насквозь пронизанном стереотипами анализе персональных данных AI»⁴².

Очевидно, что требуется не только контроль внедряемых в ИИ социальных элементов, но и сканирование/выявление уже присутствующих в интеллектуальных агентах (внесенных человеком-создателем) разрушительных, токсичных, зловердных вирусов антисоциальности: сексизме, эйджизме, расизме, эгоизме и т.д.

Заключение

Преодоление эпистемологически-антропного шовинизма является вызовом времени. Признание определенной социальной и правовой автономности ИИ не означает тождественности искусственного интеллектуального агента и человека, а лишь специфическую значимость и относительную самостоятельность ИИ. Если в правовой сфере существуют не только физические, но и юридические лица, имеющие весьма далекое отношение к «лицу» (человеку), то почему бы не быть и «искусственному лицу»?

Для решения сложнейшей задачи полноценного и эффективного развития ИИ и его использования в разных сферах общества необходимы междисциплинарные исследования, включающие в себя социально-гуманитарные науки. Этому учит и история исследований в области ИИ, и практика применения ИИ. Именно с помощью социально-гуманитарных наук интеллектуальные (машинные) агенты способны стать более полезными для общества, более гуманными, понятными («прозрачными»), готовыми к сотрудничеству с человеком. Общая концепция ИИ (AGI) возможна на основе не исключительно когнитивного интеллекта, но и искусственного социального и эмоционального интеллектов. В контексте сотрудничества между людьми отсутствие социального интеллекта мешает людям в ситуациях, связанных с обменом преимущественно социальным содержанием. ИИ будет эффективно

⁴² Там же.

участвовать в таких сценариях только в том случае, если он обладает социальным интеллектом. В более широком смысле, функциональный СИИ может представлять ценность в условиях, когда люди с низким социальным интеллектом или неадекватным социальным интеллектом (например, в межкультурных коммуникациях) могут испытывать трудности. Руководство таким исследованием требует от нас более полного понимания ключевых условий, например, контекста, условий задачи, межличностных отношений в команде. Изучение социального интеллекта, его компонентов, проявлений и влияния на взаимодействие человека с человеком и/или взаимодействие человека с ИИ-агентом/агентами позволяет заложить основу для исследований СИИ в более широком диапазоне.

Дополнительной потребностью в таких исследованиях является разработка метрик, позволяющих оценивать способности агентов в терминах психического состояния и то, как их реализация этих состояний влияет на команды агентов. Для этого необходимо использовать различные методы, в том числе оценку агентов с использованием социального интеллекта, полученных при изучении человеческого социального познания. Следует отметить, что измерение социального интеллекта у людей все еще активно изучается, поскольку его трудно операционализировать в зависимости от контекста. Тем не менее, есть хорошо зарекомендовавшие себя и последовательные методики оценки социальных когнитивных процессов на разных уровнях у людей. Агенты должны будут оцениваться по тем же критериям, что и их коллеги-люди, однако существующие оценки социального интеллекта должны быть адаптированы для агента СИИ, чтобы они могли получать, интерпретировать и генерировать данные социальных сигналов. Кроме того, агент также должен иметь возможность разрабатывать и поддерживать динамические ментальные модели отношений с другими сотрудничающими агентами в команде человек-агент, которые обновляются посредством обработки социальных сигналов.

С развитием этих более динамичных форм СИИ они могут в конечном итоге быть реализованы в условиях, потенциально слишком сложных для человека. Такие ситуации могут возникать из-за несоответствия ментальных моделей в команде (например, когда членам не хватает общих знаний) или когда формируются специальные команды (например, из-за отсутствия знакомства друг с другом). Здесь СИИ, способный диагностировать несоответствие ментальной модели и вмешиваться по мере необходимости, поможет команде быстрее достичь координации для достижения общей цели.

Список литературы:

Алехина Т.А. «Искусственная социальность»: к вопросу о дискуссиях // Социальные и гуманитарные знания. 2021. Т. 7. № 1. С. 52–65.

Барышев Д.А., Макаревич И.В., Зубанков А.С., Розалиев В.Л. Нейросетевой подход к определению эмоций человека по речи // Инженерный вестник Дона. 2022. № 5.

Бахтина И. Кого копирует искусственный интеллект? // Колонка Ирины Бахтиной «Блеск и нищета корпоративных брендов». 15 декабря 2018. <https://rben.ru/corpequities-page/967-2018-11-12>

Галичий Д.А., Афанасьев Г.И., Нестеров Ю.Г. Распознавание эмоций человека при помощи современных методов глубокого обучения // E-Scio. 2021. № 5 (56).

Дудихин В.В., Иванов А.С., Межуев И.Ю., Шоков А.В., Якупов И.Ю. Распознавание радиолокационных изображений, полученных в миллиметровом диапазоне, с использованием нейронных сетей // Вестн. Концерна ВКО «Алмаз — Антей». 2022. № 3. С. 48–58.

Искусственный интеллект — наш помощник в решении социальных проблем // Курьер ЮНЕСКО. 2019. № 3.

Карпов О.Э., Андриков Д.А., Максименко В.А., Храмов А.Е. Прозрачный искусственный интеллект для медицины // Врач и информационные технологии. 2022. № 2. С. 4–11.

Корсаков С.Н. Начертание нового способа исследования при помощи машин, сравнивающих идеи / Пер. с франц. под ред. А.С. Михайлова. М.: МИФИ, 2009. 44 с.

Петрунин Ю.Ю. Биоморфные модели в экономических исследованиях // Биомедицинская радиоэлектроника. 2017. № 10. С. 80–81.

Петрунин Ю.Ю. Нейрогосударственное управление // Нейрокомпьютеры: разработка, применение. 2016. № 6. С. 10–11.

Петрунин Ю.Ю. Нейрокомпьютерная парадигма и общество. М.: Издательство Московского университета. 2012. 304 с.

Петрунин Ю.Ю., Пугачёва А.Е. Имплементация технологий искусственного интеллекта для отбора персонала российских компаний // Безопасность цифровых технологий. 2021. № 2 (101). С. 97–121. DOI: 10.17212/2782-2230-2021-2-97-121

Петрунин Ю.Ю., Силюянова Ю.А. Статистические и нейросетевые методы в исследовании управленческих проблем в организации // Нейрокомпьютеры: разработка, применение. 2018. № 10. С. 39–47.

Платон. Законы. Соч. в четырех томах. Т. 4. М., 1995.

Резаев А.В., Трегубова Н.Д. Готовы ли социологи к анализу «искусственной реальности»? Проблемы и перспективы исследований искусственного интеллекта в социальных науках // Мониторинг общественного мнения: экономические и социальные перемены. 2018. № 5. С. 91–108.

Тавокин Е.П. Искусственность «искусственной социальности» // Социологические исследования. 2019. № 6. С. 115–122. DOI: 10.31857/S013216250005487-8

Фёдоров М., Цветков Ю. Искусственный интеллект и социальный рейтинг: начало эпохи цифрового концентрационного лагеря «в интересах человечества»? 25 июня 2021. <https://russiancouncil.ru/analytics-and-comments/analytics/iskusstvennyy-intellekt-i-sotsialnyy-reyting-nachalo-epokhi-tsvirovogokontsentratsionnogo-lagerya-v/>

Харитонова Ю.С., Савина В.С., Паньини Ф. Предвзятость алгоритмов искусственного интеллекта: вопросы этики и права // Вестн. Пермского ун-та // Юридические науки. 2021. № 3 (53). С. 488–515.

Харитонова Ю.С., Ян Т. Рекомендательные системы цифровых платформ в Китае: правовые подходы и практика обеспечения прозрачности алгоритмов // Закон. 2022. № 9. С. 40–49.

Цвык В.А., Цвык И.В. Социальные проблемы развития и применения искусственного интеллекта // Вестн. РУДН. Сер. Социология. 2022. Т. 22. № 1. С. 58–69.

Шиллер А.В. От теорий к моделям эмоций для искусственного интеллекта — основные методологические вопросы // Ценности и смыслы. 2018. Т. 4. № 56. С. 126–137.

Шиллер А.В. Методологические основания моделирования эмоций в архитектуре искусственного интеллекта. Дис. ... к.ф.н. М.: МГУ имени М.В. Ломоносова, 2019.

Andreyuk D.S., Petrulin Yu, Shuranova A., Ushakov V.L. Information agenda as an analogue of attention in sociomorphic neuronal networks // Procedia computer science. 2022. С. 292–295.

Agar J. The government machine: a revolutionary history of the computer. The MIT Press Cambridge, Massachusetts, London, England. 2003.

Ashby W.R. An Introduction to Cybernetics. Chapman & Hall Ltd., London, UK, 1957.

Bainbridge W.S., Brent E.E., Carley K.M., Heise D.R., Macy M.W., Markovsky B., Skvoretz J. Artificial social intelligence // Annual Review of Sociology. 1994. 20, 407–436.

Benouachane H. Artificial intelligence in social security: opportunities and challenges // Журнал исследований социальной политики. 2022, 20 (3): 407–418.

Bölöni L., Singh Bhatia T., Khan S.A., Streater J., Fiore S.M. Towards a computational model of social norms // PLoS ONE. 2018. 13, e0195331. DOI: 10.1371/journal.pone.0195331

Burrell J., Fourcade M. The Society of Algorithms // Annual Review Sociology. 2021. 47: 213–37.

Capgemini Research Institute. Why Addressing Ethical Questions in AI will Benefit Organizations. 2019. Available at: https://www.capgemini.com/gb-en/wp-content/uploads/sites/5/2022/05/Ethics-in-AI-Infographic_Web.pdf

Cuijpers R. Making robots socially intelligent <https://www.tue.nl/en/news-and-events/features/making-robots-socially-intelligent/>

Dhelim S., Ning H., Farha F., Chen L., Atzori L., Daneshmand M. IoT-Enabled Social Relationships Meet Artificial Social Intelligence // IEEE Internet of Things Journal. 2021. № 1.

- Friedman C.P.* A “fundamental theorem” of biomedical informatics // *J. Am. Med. Inform. Assoc.* 2009; 16: 169–170.
- Herzig A., Lorini E., Pearce D.* Social Intelligence // *AI & SOCIETY*. 2019. 34: 689 DOI: 10.1007/s00146-017-0782-8
- Hofstede G.J.* (2019). GRASP agents: social first, intelligent later. // *AI Soc.* 34, 535–543. DOI: 10.1007/s00146-017-0783-7
- How to Make Socially. Intelligent Robots Using AI? By Dr Nivash Jeevanandam. *AI Concepts*. Mar 15, 2022. India AI.
- ISSA (2022) Data-Driven Innovation in Social Security: Good Practices from Asia and the Pacific. Available at: <https://ww1.issa.int/analysis/data-driven-innovation-social-security-good-practices-asia-and-pacific>
- Kundu S.* AI in medicine must be explainable // *Nature Medicine*. 2021; 27 (8): 1328.
- Mnich V., Kavukcuoglu K., David Silver et al.* Human-Level Control Through Deep Reinforcement Learning // *Nature* 518, No 7540. 2015: 529–533.
- Nott G.* Ombudsman: Centrelink OCI Lacking Usability and Transparency // *Computerworld*. 2017. Available at: <https://www.computerworld.com/article/3476375/ombudsman-centrelink-oci-lacking-usability-and-transparency.html>
- Ornes S.* How Close Are Computers to Automating Mathematical Reasoning? AI tools are shaping next-generation theorem provers, and with them the relationship between math and machine // *Quanta Magazine*. 2020.08.27. <https://www.quantamagazine.org/how-close-are-computers-to-automating-mathematical-reasoning-20200827/>
- Penn J.* AI thinks like a corporation—and that’s worrying // *The Economist*. Nov 26th 2018. https://www.economist.com/open-future/2018/11/26/ai-thinks-like-a-corporation-and-thats-worrying?__twitter_impression=true
- Silvera D., Martinussen M., Dahl T.I.* The Tromsø Social Intelligence Scale, a self-report measure of social intelligence // *Scand. J. Psychol.* 2001. № 42. 313–319. DOI: 10.1111/1467-9450.00242
- Williams J, Fiore S.M., Jentsch F.* Supporting Artificial Social Intelligence with Theory of Mind // *Frontiers Artificial Intelligence*. 2022. 5:750-763. DOI: 10.3389/frai.2022.750763

Статья поступила в редакцию 15.01.2023.