

НАУЧНЫЕ ПРОБЛЕМЫ ОБЕСПЕЧЕНИЯ ТЕХНОЛОГИЧЕСКОГО СУВЕРЕНИТЕТА В ОБЛАСТИ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

© 2024 г. В.Б. Бетелин^{а,*}

^аФедеральный научный центр “Научно-исследовательский институт системных исследований РАН”,
Москва, Россия

^{*}E-mail: betelin@niisi.msk.ru

Поступила в редакцию 02.05.2024 г.

После доработки 04.05.2024 г.

Принята к публикации 12.05.2024 г.

В статье обсуждаются научные проблемы обеспечения технологического суверенитета в области технологий искусственного интеллекта. Одна из ключевых — моделирование на компьютере с конечными ресурсами когнитивных функций человека, обеспечивающих возможность обрабатывать и анализировать наряду с конечномерными и бесконечномерные данные. Известное решение этой проблемы — метод регуляризации А.Н. Тихонова, вследствие высокой сложности и стоимости реализации которого для построения решений искусственного интеллекта (ИИ-решений) используются менее сложные и дорогостоящие эмпирически построенные зарубежные искусственные нейронные сети (ИНС), но без гарантий безошибочности результата. Следствие этого — довольно низкий уровень внедрения ИИ-решений в промышленности, федеральных органах исполнительной власти и практическое отсутствие данных об экономическом эффекте. Предлагается считать основным критерием необходимости разработки отечественных ИИ-решений их экономическую эффективность. Статья подготовлена на основе научного сообщения, заслушанного на заседании президиума РАН 12 марта 2024 г.

Ключевые слова: технологический суверенитет, технологии искусственного интеллекта, когнитивные функции человека, конечномерные и бесконечномерные данные, метод регуляризации, эмпирически построенные ИНС, экономическая эффективность ИИ-решений.

DOI: 10.31857/S0869587324070031, EDN: FMSVIQ

Искусственный интеллект — это компьютерное моделирование только конечномерной части интеллектуальной деятельности человека. Национальная стратегия развития искусственного интеллекта на период до 2030 г. (далее Стратегия) утверждена Указом Президента Российской Федерации № 490 от

10 октября 2019 г. Президентским Указом № 124 от 15 февраля 2024 г. в Стратегию внесён перечень изменений.

Согласно Стратегии, *искусственный интеллект — комплекс технологических решений, позволяющий имитировать когнитивные функции человека (поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их.*

Когнитивные функции человека обеспечивают способность воспринимать, обрабатывать, анализировать, сопоставлять как конечномерные, так и бесконечномерные данные и их множества. Например, бесконечномерное (счётное) множество бесконечных в обе стороны спиц на плоскости. На каждой из этих спиц бесконечномерное (счётное) множество вращающихся колёсиков, на каждом из



БЕТЕЛИН Владимир Борисович — академик РАН, научный руководитель ФНЦ НИИСИ РАН.

которых присутствуют все буквы алфавита от а до я. Утверждается, что бесконечномерное множество текстов, образующихся при вращении этих колёсиков, будет включать как все уже написанные человеком литературные произведения, так и те, которые будут когда-либо написаны человеком. Когнитивные функции человека обеспечивают понимание им как справедливости этого утверждения, так и того, что это утверждение неверно для любого сколь угодно большого, но конечномерного подмножества и спиц, и колёсиков с алфавитом.

Однако имитация этих функций человека комплексом технологических решений искусственного интеллекта (далее ИИ-решения) на компьютере с конечными аппаратными (ёмкость оперативной и внешней памяти, быстродействие и т.д.) и программными ресурсами не обеспечит возможности понимания ИИ-решением справедливости этого утверждения на основе представленного на компьютере только конечного подмножества конечномерных спиц и колёсиков с алфавитом на этих списках. Другими словами, для того чтобы с помощью технологий ИИ обеспечить решение задач с бесконечномерными данными, необходимо не только свести эти задачи к конечномерной постановке, но и доказать теорему существования такой конечномерной постановки и её эквивалентность исходной бесконечномерной задаче. То есть, по сути дела, проблема создания таких технологий ИИ сводится к проблеме решения некорректных задач методом регуляризации А.Н. Тихонова [1]. Вместе с тем ввиду высокой сложности и, как следствие, высокой стоимости такого способа решения на практике применяются менее сложные и дорогостоящие эмпирические методы сведения бесконечномерной задачи к конечномерной и построения на этой основе ИИ-решения исходной задачи, однако без каких-либо гарантий безошибочности результатов. Следствием понимания конкретных причин и непредсказуемости возникновения ошибочных результатов таких ИИ-решений является тезис об отсутствии понимания того, как искусственный интеллект достигает результата [2].

Вместе с тем для решения задач с конечномерными данными могут быть построены ИИ-решения с теоретическим обоснованием безошибочности результата. Действительно, например, на основе большого, но конечномерного множества шахматных партий, которые были сыграны реальными гроссмейстерами, на компьютере с конечными аппаратными и программными ресурсами можно обеспечить имитацию (моделирование) функций человека-гроссмейстера на основе ИИ-решения “цифровой гроссмейстер”, разработанного человеком. По сути дела, “цифровой гроссмейстер” — это аналог поисковой системы Google на конечномерном множестве конечномерных партий, сыгранных реальными гроссмейстерами, дополненный

авторами этого ИИ-решения метрикой близости партии к выигрышу. Каждая партия этого множества — упорядоченная последовательность положений 32 шахматных фигур на 64-клеточной шахматной доске. Математическое обоснование успеха ИИ-решения “цифровой гроссмейстер” — приложение теоремы А.Н. Тихонова о неподвижных точках отображения на упорядоченных, ограниченных множествах [2].

Аналогичным образом на основе сколь угодно большого, но конечномерного множества конечномерных письменных произведений реальных авторов на компьютере с ограниченными аппаратными и программными ресурсами можно обеспечить имитацию (моделирование) функции реальных авторов на основе ИИ-решения СНАТ GPT. Это ИИ-решение также служит аналогом поисковой системы Google на конечномерном упорядоченном множестве многих тысяч письменных произведений реальных авторов, дополненным авторами ИИ-решения метрикой близости фрагментов реальных произведений к заданному фрагменту. Математическое обоснование успеха ИИ-решения содержится в приложении к теореме А.Н. Тихонова.

Проблема создания авторами этих ИИ-решений метрик близости партии к выигрышу и фрагментов письменных произведений к заданному фрагменту включает в себя и проблему уменьшения перебора вариантов партий и фрагментов письменных произведений.

Аналоги поисковой системы Google созданы и используются для анализа и классификации изображений, в том числе из области медицины (компьютерная томография, рентген, маммография). Эти изображения, в отличие от шахматных партий и фрагментов письменных произведений, выступают объектами бесконечномерного пространства, и поэтому для их использования компьютером требуется создание конечномерных представлений тех же объектов и соответствующей метрики близости на конечномерном пространстве этих уже конечномерных объектов. По сути дела, аналог поисковой системы Google ведёт анализ и классификацию на основе метрики близости не исходных изображений, а их компьютерных конечномерных представлений, которые отличаются от исходных изображений объёмом эмпирически исключённых из их состава данных. Это означает, что результаты классификации исходных изображений и их конечномерных компьютерных представлений могут непредсказуемым заранее образом не совпадать. Причина ошибочности результатов классификации изображений заключается в отсутствии теоретического обоснования успеха аналога поисковой системы Google для этих объектов, а следствием такой неустойчивости поисковой системы Google служит тезис об отсутствии понимания того, как искусственный интеллект достигает результата [2].

Недостаточный уровень самостоятельности в области технологий искусственного интеллекта для промышленности и федеральных органов исполнительной власти. В числе основных принципов развития и использования технологий искусственного интеллекта, сформулированных в Стратегии, — обеспечение необходимого уровня самостоятельности России в этой области, в том числе посредством преимущественного использования отечественных технологий ИИ. В развитие Стратегии Министрством цифрового развития, связи и массовых коммуникаций РФ утверждён федеральный проект “Искусственный интеллект” (далее Проект). Срок реализации — 2024 год. Планируемый результат: предприятия и граждане используют продукты (услуги), основанные на преимущественно отечественных технологиях искусственного интеллекта, обеспечивающих качественно новый уровень эффективной деятельности.

Однако данные, приведённые в проекте Аналитического отчёта [3] (далее Отчёт) “Федерального центра прикладного развития искусственного интеллекта” (ФЦПРИИ) Минпромторга РФ и экспертно-аналитическом докладе “Искусственный интеллект в промышленности” [4] (далее Доклад), свидетельствуют о достаточно низком уровне использования технологий искусственного интеллекта как промышленными предприятиями, так и федеральными органами исполнительной власти. Например, в Докладе констатируется, что ИИ обладает серьёзным потенциалом применения в промышленности, однако большого распространения не получил, что чрезмерный оптимизм в отношении ИИ опасен, что не сбылись предсказания 2016 г. одного из ведущих мировых новостных порталов “Business Insider” о выпуске к 2020 г. 10 млн беспилотных автомобилей, также как и предсказания Илона Маска о появлении на дорогах в 2020 г. 1 млн беспилотных такси. Отсутствуют данные о результатах реализации проекта Минтранса России и ассоциации “Цифровой транспорт и логистика” по масштабному коммерческому использованию с 2024 г. беспилотных грузовиков на магистрали Москва—Санкт-Петербург и ходе осуществления проекта запуска автономных перевозок между Азией и Европой к 2030 г. [5].

Аналогичным образом в Отчёте на основе данных анкетирования промышленных предприятий констатируется довольно низкий уровень внедрения технологий ИИ в целом при наличии интереса к ним. Интерес промышленных предприятий связан с ожиданием технологий ИИ, которые будут обеспечивать, применительно к решению стоящих перед ними задач, результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их (п. 5 Стратегии развития ИИ). Однако к числу ключевых барьеров внедрения ИИ-решений в Отчёте отнесены завышенные ожи-

дания промышленных предприятий (чрезмерный оптимизм) в сравнении с достигнутыми ими результатами. Точнее, с результатами, которые предприятиям достичь не удалось. Например, ИИ-решения не справились с задачей контроля технологических операций; внедрение фотоконтроля для определения цветов проводов и кабелей не дало результатов даже при распознавании самых простых видов продукции. Формирование отчётов с помощью распознавания изображения документа на базе решения с использованием нейросети не позволило справиться с поставленной задачей.

К ключевым барьерам, сдерживающим внедрение ИИ-решений для обеспечения потребностей, приведённых в Отчёте на страницах 34—36, следует отнести и завышенные ожидания федеральных органов исполнительной власти по сравнению с достигнутыми ими результатами. Об этом можно судить по тому факту, что в Отчёте отсутствуют данные как о внедрённых успешных практиках применения ИИ в деятельности органов государственной власти и в субъектах Российской Федерации, так и о переходе системы государственной власти федерального и регионального уровня к модели управления на основе автоматизированного сбора и анализа данных (с. 34).

В Отчёте приводятся только фактические данные о затратах в 2019—2023 гг. на закупочную деятельность федеральных органов исполнительной власти, которая ограничивается исключительно приобретением оборудования и программного обеспечения (59% общих затрат), на проведение научно-исследовательских работ (29% общих затрат), цель которых состоит в описании предполагаемой архитектуры ИИ-решений, оценке их целесообразности формировании итоговых требований к ним, а также на разработку ИИ-решений (12% общих затрат). Общие затраты федеральных органов исполнительной власти на эту деятельность с 2019 г. составили 19.1 млрд руб., однако в Отчёте отсутствуют какие-либо данные о результатах научно-исследовательских работ и разработок ИИ-решений, таких, например, как “Интеллектуальный юридический помощник для конструирования текстов нормативных регламентов, ведомственных актов и поручений...” (с. 35) или ИИ-решения, которое “... должно самостоятельно понимать контекст запроса, анализировать и классифицировать обращения граждан и формировать ответы...” (с. 36).

Из вышеизложенного следует, что с 2019 по 2023 г. в результате применения технологий ИИ в промышленности и федеральных органах исполнительной власти не получены какие-либо результаты, сопоставимые с результатами интеллектуальной деятельности человека и тем более превосходящие их. Основная причина несоответствия декларируемого и реального кроется в отсутствии документов стратегического характера, нацеленных на достиже-

ние технологической независимости от иностранных технологий в области ИИ, что по сути дела означает недостаточность существующего уровня самостоятельности в этих технологиях.

Причина возникновения юридической ответственности при использовании технологий ИИ — отсутствие гарантий безошибочности результата. Согласно Отчёту, нормативно-правовое регулирование ИИ (с. 11–15) определяет Стратегия развития ИИ, предусматривающая создание системы регулирования возникающих в связи с этим отношений (с. 12), включая юридическую ответственность при использовании ИИ (с. 13). В Отчёте декларируется также необходимость проработки механизмов уголовной и административной ответственности в случае причинения вреда системами ИИ и определения лиц, которые будут нести ответственность за их действия (с. 13). Это, по сути дела, означает неустраняемое отсутствие гарантий штатного функционирования систем ИИ, для которых, согласно Стратегии развития ИИ, *“...отсутствует понимание того, как искусственный интеллект достигает результата”*. Использование таких систем ИИ для обеспечения как обороны и безопасности государства, так и здоровья, безопасности граждан представляет угрозу, которая не может быть парирована ни нормативными регламентами (с. 19), ни этическими нормами (с. 20).

Причиной постулирования в Стратегии развития ИИ непознаваемости его действий и связанных с этим угроз является использование зарубежных полуэмпирически построенных искусственных нейронных сетей (ИНС) для создания решений ИИ для промышленности и федеральных органов исполнительной власти. Для этих зарубежных ИНС отсутствуют доказательства их устойчивости и сходимости, что, собственно, и оказывается непосредственной причиной непознаваемости действий, созданных на их основе решений ИИ. Так, в Докладе отмечается, что системы на основе этих решений не способны адаптироваться к небольшим изменениям входных данных, они неустойчивы, что для таких систем характерно катастрофическое забывание (то есть при изучении нейронными сетями новой задачи старые задачи забываются), что нет способов объяснить результаты решения задач в ситуациях с высокой степенью риска.

Однако, несмотря на эти обстоятельства, технологии ИИ на основе зарубежных ИНС уже внедряются в лечебные учреждения. Так, по сообщениям прессы, «учёные Сбера создали систему моделирования и прогнозирования диагнозов “путём адаптации современной нейросетевой архитектуры Transformers” для работы с диагнозами пациентов» [6]. Сообщается также, что “в России всё шире применяются разработанные Сбером AI-серверы: компьютерное зрение для расшифровки медицинских изображений (КТ, рентген, маммография),

автоматическое заполнение врачебной документации (из голоса в текст), диагностика заболеваний и другие”.

По информации в прессе, датированной 9 февраля 2024 г., «диагнозы в поликлиниках теперь будут ставить не только врачи, но и искусственный интеллект... К октябрю во всех взрослых поликлиниках Москвы планируют ввести систему “Аида” — цифрового ассистента-диагноста на базе ИИ. Разрабатывали его специалисты Сбера и правительства Москвы. Сервис обучался на записях 30 млн визитов пациентов за два года... Разработчики уверены в надёжности системы. На стадии тестов сервис с точностью 87% определял 95 наиболее распространённых диагнозов» [7].

В связи с этим необходимо отметить, что результаты тестирования могут указать только на наличие ошибок, но не на их отсутствие, к тому же при 15 млн визитов пациентов в год 13% ошибок “Аиды” означают около 2 млн ошибочных диагнозов, что почти в 30 раз больше, чем 70 тыс. врачебных ошибок в год, по данным Минздрава. Поэтому для исключения возможных юридических последствий с непознаваемостью зарубежной архитектуры Transformers реализация планов, по сути дела, массового внедрения системы “Аида” в лечебных учреждениях должна быть обусловлена обеспечением математического обоснования её функционирования и замещением нейросетевой архитектуры Transformers отечественным аналогом.

Искусственный интеллект окажет существенное влияние на экономический рост в мире¹, в том числе и на экономический рост промышленности России, для чего необходимо ориентировать задачи п. 24 Стратегии прежде всего на существенное увеличение уровня внедрения технологий ИИ в промышленность путём достижения результатов, не достигнутых в 2019–2023 гг., решить проблемы с подтверждением экономического эффекта от внедрения ИИ-решений, а также отказаться от использования зарубежных ИНС, не обеспечивающих безошибочности результата. Для этого необходимо уточнить определение технологий ИИ следующим образом.

Технологии ИИ — это технологии компьютерного моделирования интеллектуальной деятельности человека при решении конкретных классов практических задач в различных областях человеческой деятельности. Обоснование разработки и применения таких технологий ИИ для решения данного класса задач определяется сравнительной оценкой стоимости их решения на основе технологий ИИ и существующих технологий. Для технологий ИИ должно быть обеспечено теоретическое обоснование адекватности решаемой задаче, реализуемости, понимаемости и устойчивости используемых методов и алгоритмов.

¹ О внесении изменений в Указ Президента РФ от 10 октября 2019 г. № 490 “О развитии ИИ в РФ” и в Национальную стратегию.

В соответствии с этим определением необходимо в рамках Стратегии разработать подпрограмму создания отечественных технологий ИИ для практических задач (промышленности, федеральных органов исполнительной власти, образования, медицины и других сфер), решение которых более эффективно (с точки зрения финансовых и временных затрат на получение результата) на основе технологий ИИ, чем на основе технологий существующих.

Для формирования такой подпрограммы необходимо предпринять следующие меры.

- Подвести итоги реализации Стратегии-2024. В частности, проанализировать основные результаты и проблемы применения технологий ИИ, основанных на искусственных нейронных сетях (ИНС) зарубежных компаний при решении практических задач российских промышленности, федеральных органов исполнительной власти, образования, медицины и т.д. Сформулировать перечень нерешённых практических задач и проблемных вопросов, которые необходимо решить в краткосрочной перспективе 3–5 лет.

- Сформировать перечень практических задач для достижения основной цели, а также сформулировать теоретическое обоснование существования необходимых для этого решений ИИ, адекватности их практическим задачам, реализуемости, устойчивости моделей и алгоритмов.

- Определить перечень отечественных технологий ИИ, необходимых для достижения основных целей подпрограммы, включая прогнозируемый экономический эффект от их внедрения. Основное требование к этим технологиям — реализация на основе отечественных искусственных нейронных сетей, для которых имеется доказательство сходимости и устойчивости.

- Сформировать технические задания и планы реализации отечественных решений ИИ, адекватных требованиям практических задач из перечня практических задач, а также технических заданий и планов реализации отечественных ИНС, адекватных требованиям к отечественным решениям ИИ.

- Подготовить итоговый отчёт о реализации подпрограммы, включающий основные результаты, в том числе полученный экономический эффект, а также нерешённые практические задачи и проблемные вопросы.

ЛИТЕРАТУРА

1. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М.: Наука, 1986.
Tikhonov A.N., Arsenin V.Ya. Methods of solving incorrect problems. M.: Nauka, 1986.

2. Бетелин В.Б. О проблеме доверия к технологиям искусственного интеллекта // Успехи кибернетики. 2021. № 2 (3). С. 6–7. DOI: 10.51790/2712-9942-2021-2-3-1.

Betelin V.B. About the problem of trust in artificial intelligence technologies // Successes of cybernetics. 2021. № 2(3). P. 6–7.

3. Федеральный центр прикладного развития искусственного интеллекта. Аналитический отчёт о потребностях в области искусственного интеллекта и опыте использования технологий искусственного интеллекта промышленными предприятиями и органами государственной власти. ФЦПР ИИ 2023. (В печати.)

Federal Center for the Applied Development of Artificial Intelligence. An analytical report on the needs in the field of artificial intelligence and the Federal experience of using artificial intelligence technologies by industrial enterprises and public authorities. Federal Target Program II 2023. (In print.)

4. Серия “Источники новых индустрий”. Вып. 3. Искусственный интеллект в промышленности. 2022. <https://spb.energy/treki/analiticheskie-issledovaniya/>
The series “Sources of new industries”. Is. 3. Artificial intelligence in industry. 2022. <https://spb.energy/treki/analiticheskie-issledovaniya/>

5. Бетелин В.Б. В России планируют приступить к тестированию беспилотных грузовиков // Аргументы недели. 2022. № 1 (796). 12–18 января. <https://argumenti.ru/science/2022/01/754576>

Betelin V.B. Russia plans to start testing unmanned trucks // Arguments of the Week. 2022. № 1(796). January 12–18. <https://argumenti.ru/science/2022/01/754576>

6. Учёные Сбера создали систему моделирования и прогнозирования диагнозов. 22.11.2023. <https://lenta.ru/news/2023/11/22/modelirovaniya/>

Sberbank scientists have created a system of modeling and forecasting diagnoses. 22.11.2023. <https://lenta.ru/news/2023/11/22/modelirovaniya/>

7. Диагнозы в поликлиниках теперь будут ставить не только врачи, но и Искусственный интеллект. Цифровой ассистент будет предлагать свои версии на основе медицинской истории. В будущем он научится распознавать даже редкие болезни. 09.02.2024. <https://achbd.media/a/aida-ai-helper>

Diagnoses in polyclinics will now be made not only by doctors, but also by Artificial intelligence. The digital assistant will offer its own versions based on medical history. In the future, he will learn to recognize even rare diseases. 09.02.2024. <https://achbd.media/a/aida-ai-helper>

SCIENTIFIC PROBLEMS OF ENSURING TECHNOLOGICAL SOVEREIGNTY IN THE FIELD OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES

V.B. Betelin^{a,*}

^a*Federal Scientific Center “Scientific Research Institute for Systems Research of the Russian Academy of Sciences”,
Moscow, Russia*

^{*}*E-mail: betelin@niisi.msk.ru*

The article discusses the scientific problems of ensuring technological sovereignty in the field of artificial intelligence technologies. One of the key scientific problems is computer modeling of human cognitive functions with finite resources, which provide the ability to process and analyze finite-dimensional and infinite-dimensional data along with finite-dimensional ones. A well-known solution to this problem is the A.N. Tikhonov regularization method, due to the high complexity and cost of implementation of which, less complex and expensive empirically constructed foreign INS are used to build AI solutions, but without guarantees of error-free results. The consequence of this is a rather low level of implementation of AI solutions in industry and in the field of information technology and a practical lack of data on the economic effect. It is proposed to consider the economic efficiency of domestic AI solutions as the main criterion for the need for their development.

Keywords: technological sovereignty, artificial intelligence technologies, human cognitive functions, finite-dimensional and infinite-dimensional data, regularization method, empirically constructed INS, economic efficiency of AI solutions.