УДК 004.8 DOI: 10.31249/scis/2024.04.02

ИСПОЛЬЗОВАНИЕ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В НАУЧНОЙ РАБОТЕ И ПРОФЕССИОНАЛЬНЫХ ПРАКТИКАХ: МАТЕРИАЛЫ КРУГЛОГО СТОЛА

USE OF ARTIFICIAL INTELLIGENCE SYSTEMS IN SCIENTIFIC WORK AND PROFESSIONAL PRACTICES: MATERIALS OF THE ROUNDTABLE

Аннотация. Центр научно-информационных исследований по науке, образованию и технологиям ИНИОН РАН 31 октября 2024 года провел круглый стол «Использование систем искусственного интеллекта в научной работе и профессиональных практиках». Участники круглого стола обсудили использование искусственного интеллекта (ИИ) в различных сферах, включая медицину и образование, рассмотрели этические аспекты и риски, связанные с применением ИИ, поделились опытом применения ИИ в научных исследованиях и стандартизации. Были рассмотрены практические примеры использования ИИ, включая автоматизацию процессов и анализ данных; определена роль человека в взаимодействии с ИИ и его влияние на качество образования и медицинских услуг.

Ключевые слова: искусственный интеллект; медицина; диагностика; образование; этика; робототехника; стандартизация.

Abstract. The Center for Scientific and Information Research on Science, Education and Technology of INION RAS held a round table “Use of Artificial Intelligence Systems in Scientific Work and Professional Practices” on October 31, 2024. The participants of the Round Table discussed the use of artificial intelligence in various fields, including medicine and education, considered ethical aspects and

risks associated with the use of AI, shared experience in the use of AI in scientific research and standardization. Practical examples of AI use, including process automation and data analysis, were discussed, the role of humans in interaction with AI and its impact on the quality of education and medical services were defined.

Keywords: artificial intelligence; medicine; diagnostics; education; ethics; robotics; standardization.

Участники круглого стола

Ананин Денис Павлович – ведущий научный сотрудник, лаборатория исследования образовательной политики и инновационного развития, МГПУ.

Асеева Ирина Александровна – доктор философских наук, профессор, ведущий научный сотрудник центра научно-информационных исследований по науке, образованию и технологиям ИНИОН РАН, модератор круглого стола.

Введенская Елена Валерьевна – кандидат философских наук, доцент, ведущий научный сотрудник центра научно-информационных исследований по науке, образованию и технологиям ИНИОН РАН.

Владзимирский Антон Вячеславович – доктор медицинских наук, заместитель директора по научной работе НПКЦ диагностики и телемедицинских технологий ДЗМ.

Гаврилина Елена Александровна – кандидат философских наук, доцент, старший научный сотрудник центра научно-информационных исследований по науке, образованию и технологиям ИНИОН РАН.

Гарбук Сергей Владимирович – кандидат технических наук, и. о. директора ВИНИТИ РАН, председатель технического комитета по стандартизации ТК164 «Искусственный интеллект».

Гребенщикова Елена Георгиевна – доктор философских наук, заместитель директора по научной работе, руководитель центра научно-информационных исследований по науке, образованию и технологиям ИНИОН РАН.

Жебит Владимир Александрович – кандидат технических наук, отделение математики и механики, заведующий ОНИ по проблемам энергетики и металлургии ВИНИТИ РАН.

Залаев Геннадий Захарович – доктор технических наук, заместитель директора – научный руководитель Российского государственного архива научно-технической документации.

Карташов Дмитрий Александрович – доцент кафедры комплексной защиты информации РГГУ, научный сотрудник МТУСИ.

Москалёв Игорь Евгеньевич – кандидат философских наук, доцент, директор центра мониторинга качества образовательных программ, и. о. заведующего кафедрой антикризисного регулирования и управления рисками, РАНХиГС.

Подгорный Борис Борисович – доктор социологических наук, директор Курского филиала ООО «Инвестиционная палата».

Пройдаков Эдуард Михайлович – директор Виртуального компьютерного музея.

Самохин Александр Александрович – доктор технических наук, главный научный сотрудник ИОФ им. А.М. Прохорова РАН.

Степанов Вадим Константинович – кандидат педагогических наук, доцент, старший научный сотрудник НИО библиотековедения ИНИОН РАН.

Цветкова Валентина Алексеевна – доктор технических наук, профессор, главный научный сотрудник ВИНИТИ РАН.

Черникова Ирина Васильевна –доктор философских наук, профессор, зав. кафедрой философии и методологии науки ТГУ, профессор кафедры социальной коммуникации Национального исследовательского ТПУ.

Чистякова Татьяна Александровна – редактор центра научно-информационных исследований по науке, образованию и технологиям, ИНИОН РАН.

Асеева И.А.: Добрый день, уважаемые коллеги! Спасибо за ваш интерес к социогуманитарным аспектам создания и применения искусственного интеллекта. Центр научно-информационных исследований по науке, образованию и технологиям ИНИОН РАН организует цикл мероприятий, посвященных очень актуальной, очень острой теме использования искусственного интеллекта в разных сферах жизни людей. Первый круглый стол состоялся 18 января 2024 г. Он был посвящен проблеме использования искусственного интеллекта в научной работе: обсуждали этические аспекты использования искусственного интеллекта как для учено-

го, так и для специалиста, который занимается, например, издательским делом; поднимали проблему авторства; размышляли, в каких социальных ролях может выступать искусственный интеллект по отношению к ученому.

Сегодня очередное мероприятие, посвященное этой проблеме, и мы обсудим практический ракурс использования искусственного интеллекта. Проблема в том, что практическое использование идет быстрее, чем получается осмыслить и теоретизировать это знание. Это отдельная история, и мы сегодня попытаемся затронуть именно этот аспект. Какие проблемы и опасения возникают среди профессионалов-практиков, использующих технологии ИИ в различных сферах: в медицине, на производстве, при проведении каких-то экспериментов, в образовании?

Наш первый докладчик Сергей Владимирович Гарбук, исполняющий обязанности директора Всероссийского института научной и технической информации Российской академии наук, председатель технического комитета по стандартизации ТК164 «Искусственный интеллект», кандидат технических наук и старший научный сотрудник. Пожалуйста, Сергей Владимирович, Вам слово.

Гарбук С.В.: Уважаемые коллеги, добрый день! Мне предложили выступить и рассказать про риски создания и применения систем искусственного интеллекта (СИИ). Я с удовольствием согласился, так как эта тема редко поднимается для обсуждения. Охотно рассказывают про возможные преимущества применения искусственного интеллекта (ИИ), но про риски не всегда любят говорить, а тем не менее они есть.

Прежде всего, риски связаны с тем, что современные системы искусственного интеллекта – это системы на основе методов, алгоритмов машинного обучения (МО). Область знаний ИИ не сводится, безусловно, к МО (существует еще система, основанная на правилах), но успехи, которые мы наблюдаем последние 15 лет, происходят благодаря применению именно МО. Оно, в свою очередь, так бурно расцветает, потому что, во-первых, высокоинформативные сенсоры становятся дешевыми, во-вторых, каналы связи, передающие данные от этих сенсоров, также дешевеют и все больше приобретают хорошую пропускную способность. Наконец, совершенствуются и средства вычислительной техники. Все эти

технологические предпосылки обеспечивают рост и прогресс СИИ, основанных на методах МО.

Методы МО позволяют решать задачи обработки данных без наличия интерпретируемой модели наблюдаемого объекта или процесса. Интерпретируемая модель – это модель, основанная на каких-либо знаниях, которые предполагаются истинными. Алгоритм, основанный на интерпретируемых правилах, может быть очень сложным (например, алгоритм решения большой системы дифференциальных уравнений или алгоритм обращения матрицы), но в отношении каждого действия этого алгоритма понятно, почему числа складываются друг с другом, с каким коэффициентом умножаются и т.д.

Алгоритмы СИИ, основанные на методах МО, такой прозрачностью и интерпретируемостью не обладают принципиально. Это им присущее свойство. Алгоритм возникает при усвоении некоего обучающего набора данных (НД) таким, какой он есть. Это очень универсальный подход, но расплата – плохая предсказуемость работы алгоритмов МО в реальных условиях эксплуатации, которая приводит к тому, что некорректная работа алгоритмов происходит в самых неожиданных случаях, когда происходят небольшие, с точки зрения здравого смысла человека, искажения исходных данных или же когда условия применения, т.е. внешние факторы, и исходные данные находятся в определенном сочетании. Сложно предсказать, при каких сочетаниях исходных данных и условиях применения произойдет сбой этого алгоритма. Это действительно очень сложная, практически нерешаемая задача. Данная деградация, т.е. отсутствие гарантии функциональной корректности, имеет место, причем, чем более сложен этот алгоритм, тем более сложно дать эти гарантии. Поэтому, когда мы наблюдаем так называемое ликование по поводу применения больших генеративных моделей и других очень сложных систем с архитектурной точки зрения, с точки зрения обучающего НД, который был использован для их создания, то возникают определенные опасения, связанные с тем, что результаты работы таких алгоритмов всегда очень правдоподобны, но за этим правдоподобием стоит отсутствие гарантии функциональной корректности алгоритма.

Вопрос доверия. Наверняка все присутствующие слышали такое выражение – «доверие к искусственному интеллекту» (от английского «trust», «trustworthiness»). В области ИИ этот термин

очень активно используется. На наш взгляд, сейчас очень важными становятся следующие четыре аспекта доверия в области ИИ:

1) техническая надежность;
2) информационная безопасность;
3) импортонезависимость;
4) функциональная корректность.

Последний аспект особенно сложно подтвердить именно для СИИ на основе методов МО. Но для того чтобы разобраться в этом, давать какие-то гарантии и оценить риски более осмысленно, на сегодняшний день внимательно изучаются процессы жизненного цикла (ЖЦ) системы. Надо сказать, что все нормативно-технические документы, которые устанавливают и описывают процессы ЖЦ информационных систем в целом, делятся на две группы:

- документы, устанавливающие последовательность реализации ЖЦ
- документы, задающие полный перечень процессов, не устанавливая какую-то последовательность.

На рис. 1 показаны генетика взаимного перетекания различных стандартов на международном уровне, в том числе в Международной организации по стандартизации, и наши национальные деривативы.

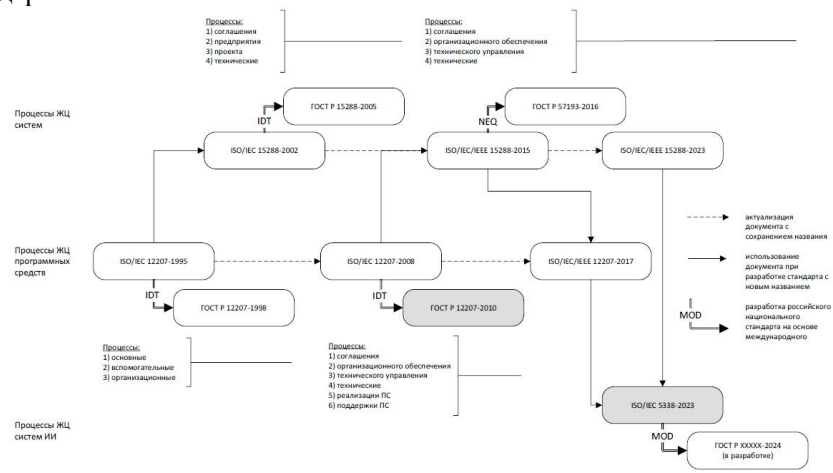

Рис. 1. Стандарты ЖЦ-систем, программных средств и СИИ

Процесс разработки длился с середины 90-х годов, и в 2023 г. появился специальный документ – ЖЦ СИИ Международной организации по стандартизации (англ. International Organization for Standardization, ISO). В ноябре 2024-го данный стандарт будет утвержден в рамках уже нашего национального Технического комитета по стандартизации № 164 «Искусственный интеллект» (ТК164). У нас появится свой документ, который фиксирует основные особенности применения СИИ. Они приведены в таблице 1.

Таблица 1

Особенности и модель жизненного цикла СИИ

В соответствии с моделью ЖЦ		В соответствии с ISO/IEC/IEEE 5338–2023	
№	Содержание	№	Содержание
1	Обязательность этапа обучения	5	Принципиальная необходимость использования представительных НД для обучения, тестирования, верификации и валидации СИИ. Определение поведения моделей МО не путем программирования («not programmed»), а методом изучения на основе данных
2	Неполная интерпретируемость и отсутствие строгих доказательств функциональной корректности алгоритмов МО	4	Принципиальный вероятный характер поведения СИИ, наличие ограничений на применение формальных методов при верификации корректности моделей МО
		6	Важное значение знаний, определяющих корректность моделей МО
		8	Возможное снижение доверия к СИИ на основе алгоритмов МО, связанное с их меньшей предсказуемостью, понятностью и объяснимостью поведения по сравнению с системами обработки данных, основанных на интерпретируемых знаниях
3	Возможность дообучения СИИ на стадии эксплуатации	1	Необходимость мониторинга возможных изменений в поведении контролируемого объекта в процессе применения СИИ
		3	Итерационное уточнение требований и поведенческих сценариев СИИ, в том числе – при возникновении непредвиденных ситуаций в процессе их эксплуатации

4	Важность социальной приемлемости применения	7	Необходимость информирования пользователей о возможных рисках применения СИИ для предотвращения избыточного и неоправданного доверия к ним, в том числе – при попытке использования систем для замены человека
5	Необходимость сопоставления с интеллектуальными способностями человека	2	Необходимость контроля качества СИИ, обладающих автономностью, сопоставимой с поведением человека, и способных нанести существенный ущерб окружающим
6	Рост конфиденциальности данных в процессе эксплуатации СИИ	–	–

В правой части содержатся особенности, которые представлены в международном стандарте, а в левой – особенности, которыми мы пользуемся на прикладном, или даже инженерном уровне. Они более агрегированы и чем-то дополняют даже международные.

Как я уже говорил, туда входит обязательность обучения, и это принципиальный момент для систем МО, что следует даже из названия «неполная интерпретируемость». Еще раз хотел бы на это обратить внимание, потому что иногда, на мой взгляд, спекулятивно говорят об объяснимости, прозрачности алгоритма ИИ. Ровно в тот момент, когда алгоритм станет объяснимым, он перестанет быть алгоритмом ИИ. Он всегда будет необъясним в таких классических соображениях.

Возможность дообучения систем на стадии эксплуатации. При дообучении возможны такие риски, как смещение характеристик. Возьмем за пример системы беспилотный автомобиль для разнообразных условий эксплуатации в городе. Он ездит из дома на работу по простому маршруту, в процессе эксплуатации продолжает обучаться. Он в каком-то смысле глупеет, если здесь возможно применять такие человеческие термины. То есть его характеристики смещаются в сторону привычных для него условий эксплуатации, и в сложной ситуации, к которой он изначально вообще-то был приспособлен, он себя может повести хуже, чем в тот момент, когда он был выпущен с завода. Важна социальная при-

емлемость применения, это означает, что СИИ приходит на смену человеку, на нее возлагается всё больше и больше задач, которые традиционно решал человек, обладающий юридической ответственностью, совестью, здравым смыслом и еще другими вещами, которых у технических систем нет и быть не может. Поэтому возможность применения СИИ из таких социальных соображений выходит на первый план. Необходимость сопоставления с человеком-оператором тоже понятна. Если мы меняем способ управления беспилотным такси с человека на алгоритм, то мы должны понимать, что безопасность перемещения в такой машине и пешеходов, разумеется, не пострадает. Ну и рост конфиденциальности данных. В процессе накопления данных при работе СИИ конфиденциальность растет. Например, я не против, чтобы публиковались где угодно данные о, допустим, одной моей поездке на такси, но о том, как я на такси перемещаюсь в течение года или двух, мне бы крайне не хотелось, чтобы стало известно. Потому что это много говорит об образе жизни человека, и в целом это конфиденциальная информация, безусловно. Вот в какой момент произошел переход из абсолютно публичных в такие чувствительные данные? Вопрос остается открытым. Существующая нормативная база не предполагает применение каких-то более-менее автоматизированных процедур оценки уровня конфиденциальности. Это проблема. То есть в процессе накопления данных нужно отслеживать, какой у них уровень конфиденциальности. Существующая сейчас парадигма нормативно-технических требований и вообще система защиты не предполагает автоматического решения такого рода задач.

С учетом этого, построена модель жизненного цикла СИИ (рис. 2). Она достаточно стандартная, характерная для любых информационных систем, за исключением того, что здесь добавлен этап обучения, дообучения на стадии эксплуатации, тестирования, повторного тестирования при обучении и т.д.

В ромбах с цифрами показаны так называемые информационные компоненты, которые сопровождают СИИ на разных стадиях. Это могут быть НД, это могут быть формализованные технические требования, предусмотренные условия эксплуатации. Все это так или иначе формализованные некие данные. Качество информационных компонент и определяет качество системы в целом. И некий заданный уровень качества этих ин-

формационных компонент дает как раз те гарантии функциональной корректности, которые необходимы для ответственного применения такого рода системы.

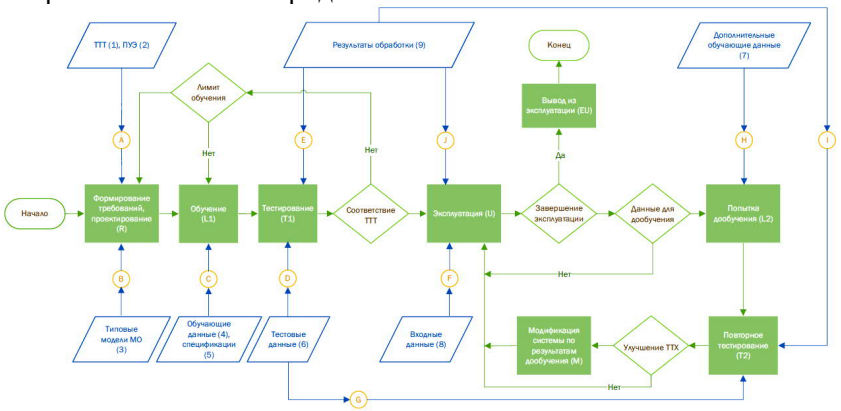

Рис. 2. Модель жизненного цикла СИИ

На рис. 3 в левой части показаны стадии ЖЦ, которые применимы для нашей модели с многочисленными другими документами в области управления ЖЦ. Это и 34-й ГОСТ по автоматизированным системам, это и 15-й ГОСТ по системам разработки и постановки на производство продукции.

Стадии ЖЦ СИИ	ГОСТ 34.601–90 АС	ГОСТ РВ 15.004-2004	ГОСТ Р 56135-2014	ГОСТ Р 53791-2010	ГОСТ Р 60.0.0.6—2023
R: формирование требований и проектирование	- формирование требований к АС; - разработка концепции АС; - техническое задание; - эскизный проект	исследование и обоснование разработки	- создание научно-технического задела; - формирование концепции образца ПВН (аванпроект)	- обоснование разработки; - разработка ТЗ	1: обоснование разработки и формирование исходных требований; 2: проектирование (разработка); 3: изготовление (производство)
L1: обучение	- технический проект; - рабочая документация	- разработка; - производство	- разработка; - производство	- проведение ОКР; - производство и испытания	
T1: тестирование	ввод в действие	эксплуатация изделий	эксплуатация		4: контроль (приемка)
L2: дообучение	не предусматривается, т.к. не относится к созданию АС			использование (эксплуатация)	5: эксплуатация
T2: повторное тестирование					
U: эксплуатация					
EU: вывод из эксплуатации			утилизация	ликвидация (с избавлением от отходов путем их утилизации и/или удаления)	7: утилизация
M: модификация по результатам дообучения	сопровождение АС	капитальный ремонт (для изделий, подлежащих капремонту)	капитальный ремонт (при необходимости)	модернизация	6: ремонт (модернизация)

Рис. 3. Соответствие этапов ЖЦ в разных нормативных документах и предложенной модели

К этим девяти информационным компонентам могут быть применены требования целостности, они должны быть достаточно содержательными, информационно наполненными, и требования конфиденциальности. Возникает такая матрица 2 на 9, 18 видов требований. В таблице требования обозначены достаточно подробно и показаны типовые информационные атаки в терминах информационной безопасности, которые могут быть реализованы при нарушении целостности или конфиденциальности той или иной информационной компоненты (табл. 2).

Таблица 2

Требования к информационным компонентам (факторы качества) СИИ

Информационная компонента	Причины, связанные с нарушением целостности n-ой компоненты	Причины, связанные с нарушением конфиденциальности n-ой компоненты
1. Функциональные характеристики (ФХ)	*i*1 Неполный набор, некорректные пороговые значения и весовые коэффициенты	*c*1 Компрометация ФХ (для систем безопасности и др., предполагающих возможность активного противодействия)
2. Предусмотренные условия эксплуатации (ПУЭ)	*i*2 Неполный перечень факторов эксплуатации, расхождение реальных условий эксплуатации с ПУЭ	*c*2 Компрометация ПУЭ (для систем безопасности)
3. Типовые модели МО, эталонные архитектуры	*i*3 Манипуляции с моделями: размещение злоумышленниками в открытом доступе моделей МО с программными закладками, реализующими НДВ	*c*3 Извлечение моделей: нарушение конфиденциальности сведений о моделях МО, использованных при разработке СИИ
4. Обучающие НД	*i*4 Манипуляции с обучающими НД: «атаки отравления» (poisoning), каузативные (causative) атаки	*c*4 Извлечение сведений об обучающих НД с целью повышения эффективности реализации атак на СИИ и извлечение сведений об объектах, данные которых использовались при обучении СИИ

5. Спецификация требований к моделям МО	*i*5 Неполные и искаженные спецификации, обусловленные заблуждениями относительно законов и закономерностей, присущих предметной области	*c*5 Использование злоумышленниками спецификаций для реализации атак, противоречащих принятым интерпретируемым моделям
6. Тестовые НД	*i*6 Искажение тестовых НД, низкая репрезентативность, смещенность тестовых НД	*c*6 Извлечение сведений о тестовых НД заинтересованными лицами (например, недобросовестными разработчиками)
7. Данные для дообучения СИИ	*i*7 Смещение обучающего НД вследствие эксплуатации в однотипных условиях	*c*7 Компрометация сценарием применения (для ВВСТ систем безопасности)
8. Исходные данные	*i*8 Прямые манипуляции с входными данными, «состязательные примеры», непрямые манипуляции (воздействие на входы сенсоров СИИ)	*c*8 Извлечение исходных данных: «атаки инверсии модели» (modelinversion)
9. Результаты обработки	*i*9 Искажение результатов обработки при их отображении, хранении, использовании в процессе тестирования СИИ и т.п.	*c*9 Извлечение выходных данных: «атаки инверсии модели» (modelinversion), подмены результатов тестирования для завышения возможностей или дискредитации СИИ

Мы выявляем 18 проблем, связанных с обработкой данных и с управлением ЖЦ СИИ. Проблемы обозначены буквами: «c» – конфиденциальность, «i» – целостность, а цифры соответствуют номеру информационных компонент. Эти проблемы взаимосвязаны, имеют каскадный эффект. Я сейчас проиллюстрирую. Нарушение конфиденциальности обучающего НД существенно повышает возможности злоумышленника по реализации эффективных информационных атак на систему МО. Или, например, нарушение конфиденциальности тестового НД из испытательной лаборатории какого-то органа по сертификации резко сокращает представительность результатов тестирования, так как если тестовый НД доступен разработчику, то произойдет эффект переобучения, и результаты тестирования не будут говорить ни о чем. Нарушение

целостности и конфиденциальности информационных компонент приводит к двум группам таких негативных последствий. Первая – это нарушение функциональности СИИ, причем нарушается как сама функциональность (она объективно снижается), так и возрастает погрешность оценки этой функциональности. То есть потребитель не знает, с чем имеет дело. Возвращаясь к системам генеративного ИИ, правдоподобные картинки и тексты он выдает, но что это такое на самом деле, никто толком сказать не может. Вторая группа – это нарушение конфиденциальности, причем нарушаются и данные о самой системе, что не всегда приемлемо, если речь идет о системах безопасности, например, когда рентгеновские снимки и данные графических исследований используются для обучения алгоритма. Этим нужно заниматься, но при этом нужно помнить об интересах и правах тех людей, которые эти сведения о себе предоставили.

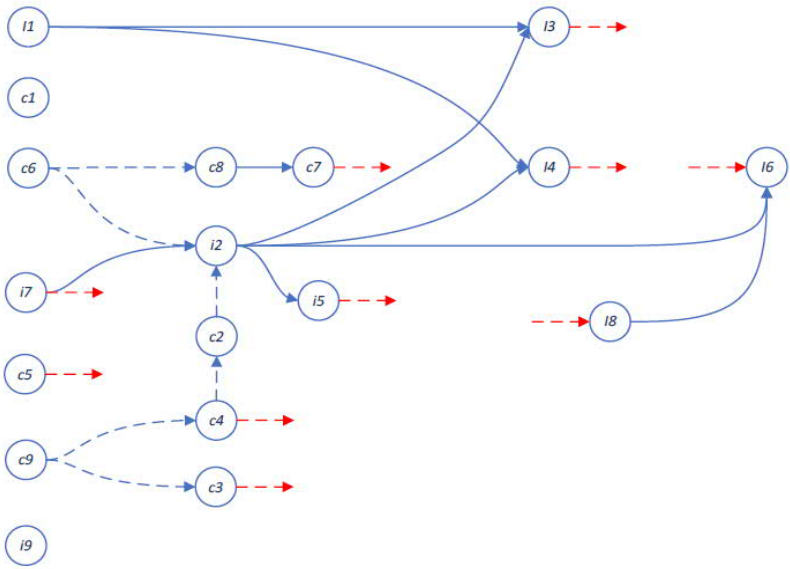

Рис. 4. Негативные эффекты, обусловленные факторами качества СИИ
(1 – функциональные характеристики; 2 – предусмотренные условия эксплуатации; 3 –Типовые модели МО, эталонные архитектуры; 4 – обучающие НД; 5 – спецификации требований к моделям МО; 6 – тестовые НД; 7 – данные для дообучения; 8 – исходные данные; 9 – результаты обработки)

Пунктирными стрелками показаны связи одних проблем с другими, которые характерны для систем, где есть активный злоумышленник. Не везде он есть, например, в системах здравоохранения его нет, а в системах безопасности он есть. Там, где такие злоумышленники есть, возникают дополнительные негативные связи.

В итоге нарушение функциональности приводит к появлению трех групп рисков (табл. 3).

Таблица 3

Риски, обусловленные деградацией функциональных характеристик ИИ

Вид угроз, обусловленных нарушением функциональной корректности СИИ	Категории заинтересованных сторон	
	Лица, непосредственно участвующие в создании и применении СИИ (акторы ИИ)	Третьи лица
1 Угрозы жизни и здоровью людей, экологические угрозы	1.1 Потребители, разработчики и поставщики (собственная безопасность, дополнительные требования гос. регуляторов)	1.2 Общество в целом и регуляторы (безопасность общества и окружающей среды)
2 Угрозы информационной безопасности в отношении заинтересованных сторон	Нет	2.2 Общество в целом и государственные регуляторы (защита персональных данных, предотвращение деструктивных последствий)
3 Нарушение этических и других норм «мягкого» права	Нет	3.2 Общество в целом (социальная приемлемость создания и применения СИИ)
4 Неопределенные потребительские свойства, не влияющие непосредственно на безопасность жизни и здоровья людей, экологическую безопасность	4.1 Потребители (функциональные характеристики, определяющие возможность применения СИИ по назначению), разработчики и поставщики (характеристики конкурентоспособности СИИ)	Нет

Первая группа содержит наиболее очевидные в области физического воздействия на человека и окружающую среду риски. Не для всех систем они характерны, а только там, где идет речь о киберфизических системах, которые каким-то образом взаимодействуют с окружающим миром. Например, тот же беспилотный автомобиль, промышленная робототехника с ИИ, системы вооружения, которые, безусловно, могут себя вести не всегда так, как от них ожидают.

Вторая группа рисков – условно ментальные, это риски информационной безопасности, которые возникают при нарушении функциональной корректности. Я подчеркиваю, что здесь не идет речь о рисках информационной безопасности в классическом понимании регуляторов (целостность доступности и конфиденциальности). Эти риски связаны, например, с воздействием на массовое сознание, когда рекомендательная система или поисковая система в ответ на запросы выдает искаженную или предвзятую информацию, или навязывает какое-то негативное мнение и реализует другие деструктивные информационно-психологические воздействия. Здесь информационная безопасность трактуется широко, но в рамках доктрины информационной безопасности страны. К ментальным мы относим риски, связанные с нарушением этических норм. В нашей стране этим много занимаются. И вообще, на прошедшем форуме БРИКС наш президент даже призвал стран-участниц БРИКС+ присоединиться к Кодексу этики в ИИ, который разработан в России под руководящей ролью СБЕРа.

Ну и, наконец, третья группа рисков связана с экономической безопасностью. В данном случае отсутствуют ментальные и физические угрозы человеку и обществу, однако система работает плохо и не оправдывает вложенных в нее средств. Это тоже важно. Система при этом теряет во многом смысл. Такая обобщенная модель показана на рис. 5.

Рис. 5. Модель рисков при создании и применении СИИ

Слева две группы информационных компонент, к которым предъявляются требования целостностии конфиденциальности. Возникают негативные последствия, связанные с деградацией функциональных характеристик и с непониманием этих функциональных характеристик при тестировании и компрометации данных. В итоге актуализируются риски физические, информационные, социальной приемлемости и потребительских свойств. Вот так мы себе представляем сейчас модель рисков в области ИИ.

Во второй части я перейду к деятельности в рамках ТК 164 и системы оценки соответствия в области ИИ.

Технический комитет, под эгидой которого разрабатывается национальный стандарт, был создан в 2019 г. в ответ на создание в 2018 г. Международного технического комитета по искусственному интеллекту в рамках ISO. Деятельность ТК 164 осуществляется в рамках перспективной программы стандартизации по приоритетному направлению «Искусственный интеллект». Она утверждается Росстандартом и Минэкономразвития России. Первая программа была разработана в 2020 г., а в конце 2023 г., меньше года назад, она была актуализирована. На сегодняшний день эта программа объединяет 71 организацию, которые охватывают практически все отрасли экономики, социальной сферы, федеральные органы исполнительной власти, ведущих разработчиков и т.д. На

сегодняшний день насчитывается уже более 100 утвержденных стандартов (рис. 6).

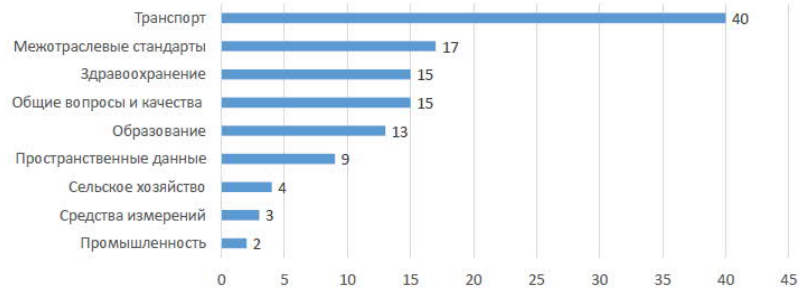

Рис. 6. Утвержденные стандарты в области ИИ (октябрь 2024)

При разработке стандартов ТК 164 ориентируется на международные практики. На международном уровне возникают стандарты, как правило, рамочные, терминологические стандарты, в области унификации форматов представления данных, в области ЖЦ и т.д. В процессе гармонизации на национальном уровне ТК 164 стремится минимально их изменять для того, чтобы не отрываться от мировых практик. На национальном уровне мы разрабатываем достаточно уникальные документы, которые унифицируют методы испытания тех или иных прикладных технологий ИИ, и наша страна совершенно точно занимает в данной области лидирующие позиции. На международном уровне такие стандарты по унификации методов испытаний разрабатываются, как правило, в крупных корпорациях и не являются публичными, понятными обществу и доступными. И это не совсем верно, потому что доверие общества к новым технологиям возникает, когда процедура оценки соответствия является публичной, прозрачной, доступной всем.

Поговорим о второй части системы оценки соответствия. Это система добровольной сертификации «Интеллометрика» (СДС «Интеллометрика»), которая была зарегистрирована в декабре 2023 г. Это первая система добровольной сертификации в нашей стране. Система строится по межотраслевому принципу. Руководящий орган находится на базе НИУ ВШЭ, наблюдательный совет

существует, комиссия по допуску, апелляционная комиссия, то, что положено иметь в СДС. А вот органы по сертификации, испытательные лаборатории и методические центры находятся в отраслевых ведущих организациях.

Рис. 7. Система добровольной сертификации «Интеллометрика»

На рис. 7 отражено, как структурирована работа ТК 164 с различными отраслями. В каждой отрасли существует свое отраслевое регулирование, и наши стандарты и подходы к испытанию систем должны туда гармонично вписываться. Но тем не менее ИИ сам по себе, отдельно взятый, не имеет смысла. Он начинает иметь смысл только в прикладных решениях, в отраслевых. Применение ИИ важно в рамках конкретных отраслей, и его регулирование должно рассматриваться именно там.

Существуют соответствующие документы в СДС «Интеллометрика». Ключевым принципом проведения сертификационных испытаний является испытание на представительных, репрезентативных тестовых НД. Мы стремимся обеспечивать представительность испытаний следующим образом.

Вариативность сценариев и тестовых НД при тестировании должна примерно соответствовать той вариативности, с которой нам предстоит столкнуться в предусмотренных условиях, в реальных условиях эксплуатации. Для этого мы формализуем условия

эксплуатации, выделяя перечень существенных факторов, которые влияют на сложность решения той или иной прикладной задачи, фиксируем закономерность распределения существенных факторов для предусмотренных условий и требуем от тестового НД примерного соответствия распределению существенных факторов, как предвидим в будущих условиях эксплуатации. И вот эту схожесть законов распределения существенных факторов мы используем для оценки репрезентативности тестового НД и репрезентативности полученных оценок при тестировании. Вот такой наш основной подход.

Однако, предъявляя требования репрезентативности к тестовым НД, мы учитываем, что одной из ключевых задач является описание сложности внешней среды. Ясно, что реальный мир бесконечномерен, бесконечно сложен. Но в контексте определенных прикладных задач ИИ, размерность этого внешнего мира, его вариативности можно редуцировать до каких-то мыслимых размеров. Вот это редуцирование, на наш взгляд, сейчас можно сделать только с привлечением человека-эксперта. Для того чтобы человек-эксперт мог выделить эти существенные факторы эффективно и внятно, мы должны и задачи ИИ уметь формулировать в терминах естественного интеллекта человека. Вот этим мы сейчас тоже занимаемся и пытаемся создать такую антропоморфную, так сказать, человекоподобную классификацию задач ИИ (рис. 8), с тем чтобы повысить эффективность решения задачи по описанию вариативности сложности внешней среды.

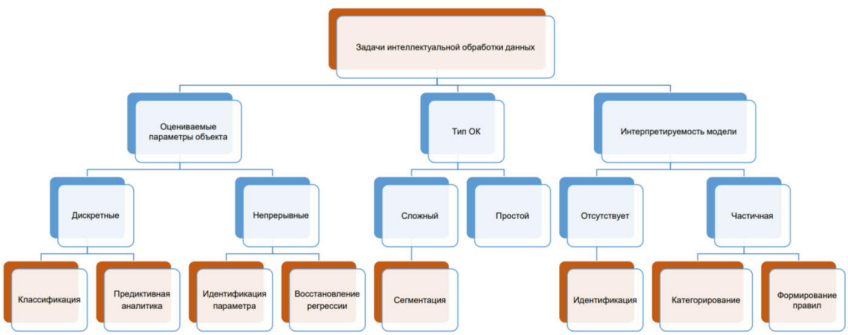

Рис. 8. Универсальные классы задач ИИ (по аналогии с естественным интеллектом человека)

Это позволяет для типовых задач описать те факторы, которые влияют на ту или иную задачу. Например, испытывая алгоритм распознавания дорожных знаков, мы должны варьировать номенклатуру и типы, размеры номеров, характеристики фона, количество одновременно наблюдаемых знаков, пространственное и радиометрическое разрешение средств и т.д. Все эти факторы мы должны варьировать в достаточной степени, причем учитывая не только диапазоны их собственных изменений, но и координационные связи, потому что во многих случаях это зависимые вещи.

Завершая свое выступление, я хочу сказать, что в конце августа 2024 г. в СДС «Интеллометрика» были запущены первые три испытательные лаборатории. Одна из них ведет свою деятельность на базе ВНИИМ им. Д.И. Менделеева, г. Санкт-Петербург. Они занимаются применением ИИ в средствах измерений. Это и лаборатория, и орган по сертификации. Вторая лаборатория по испытанию ИИ в дорожно-строительной технике – «ДСТ-Урал», г. Челябинск. Третья лаборатория по применению ИИ в средствах видеонаблюдения – ООО «Открытый код», г. Самара. Была создана специальная комиссия. Мы надеемся, что работа этих специальных лабораторий будет повышать доверие общества, потребителя, государства к тем системам ИИ, которые применяются на практике. Этого сейчас очень сильно не хватает. Ну и в конечном счете мы сможем создать СИИ с гарантированной функциональной корректностью. Мы понимаем под этим, что для предусмотренных условий эксплуатации могут быть оценены:

- доверительные интервалы и вероятности прогнозирования погрешностей измерений для определенных условий проведения измерений;
- предельные интегральные риски, связанные с некорректной работой системы ИИ;
- ресурсы, необходимые злоумышленнику для успешного информационного воздействия на измерительную систему с алгоритмами ИИ (опционально, при наличии активного злоумышленника).

СИИ с гарантированной функциональной корректностью – это системы, для которых понятны их функциональные характеристики, понятные остаточные ненулевые риски, связанные с применением этих систем в определенных условиях. Они всегда ненуле-

вые. И если это применимо, в случае злого умысла, то понятны ресурсы, которые злоумышленник должен затратить, чтобы противодействовать этой системе. Если все эти требования выполняются, то подобную СИИ мы можем применять для решения ответственных задач.

Уважаемые коллеги, спасибо большое за внимание. Готов ответить на вопросы.

Пройдаков Э.М.: Пройдаков Эдуард Михайлович, директор Виртуального компьютерного музея. Вы утверждаете, что необъяснимостью обладает только свойство алгоритмов ИИ. На самом деле это не совсем так, потому что в экспертных системах такого эффекта нет.

Гарбук С.В.: Я подчеркнул, что я говорю о СИИ, основанных на методах МО. Экспертные системы – это другое, там интерпретируемые правила. Обратите внимание, я говорил об этом несколько раз. Спасибо.

Самохин А.А.: Самохин Александр Александрович, Институт общей физики. При работе с большими данными, с элементами обучения, должно быть сформулировано понятное определение интеллекта вообще. Как Вы его определяете?

Гарбук С.В.: Это очень сложный вопрос, хотя могу много говорить на эту тему. Но раз уж вы его задали, я хочу сказать, что, например, в национальной стратегии ИИ, утвержденной указом президента, не много не мало, ИИ определяется как раз через естественный. То есть это системы, которые в той или иной степени имитируют когнитивные способности человека. Вот моя личная точка зрения – это совершенно неверно. Более простое определяют через более сложное. Потому что, например, перемножение чисел – это когнитивная способность? Безусловно. Тогда, значит, все калькуляторы – это СИИ, что полная ерунда. И вообще за определение естественного интеллекта я бы не взялся. Я бы с удовольствием сам посмотрел на человека, который даст определение естественного интеллекта человека и перечислит все функции, которые присущи интеллекту, претендуя на какую-то полноту этого списка. Поэтому вы, конечно, коснулись темы сложной. Я определение ИИ не давал и не готов этого делать. Я говорил, что есть методы МО, на которых базируются сейчас прикладные СИИ, которые приносят пользу. А вообще термин ИИ, как вам известно, он

1956 года, если я не ошибаюсь, ему под 70 лет скоро будет. И вот за эти семь десятилетий в данное понятие постоянно разный смысл вкладывался. То экспертные системы в 80–90-х годах, то до этого тест Тьюринга, сейчас МО. Это метафора, это фигура речи. Поэтому надо, конечно, договариваться о понятиях. Я с самого начала говорил, что́ я понимаю под этим МО.

Самохин А.А.: Ведь на самом деле тут речь идет о терминологии. А терминология для любой точной науки или точных механизмов, ну, я имею в виду механизмов интеллектуальных, чрезвычайно важна. Я согласен с тем, что вы сказали, это не возражение, это, может быть, некое дополнение. Если вспомнить еще древние времена, когда пытались определить, что такое человек, помните известный анекдот про Диогена, который, в ответ на определение Платона, что человек – это двуногое животное без перьев, принес ощипанного петуха. То есть терминология – это очень серьезная проблема. И она упирается во многие другие аспекты, о которых я сейчас говорить не буду. Спасибо.

Гаврилина Е.А.: Сергей Владимирович, спасибо большое за такой обстоятельный доклад. У меня, наверное, технический вопрос. Какие методы борьбы с искажениями в обучающих выборках существуют? При условии, что в «обучении без учителя» данные собирает СИИ и на основании этих данных потом принимаются какие-то свои решения, как борются со смещениями?

Гарбук С.В.: Тут один вопрос сложнее другого. Вы очень правильно умеете находить болевые точки. Вообще вопрос качества данных, который используется для обучения, не решен до конца. Об этом сейчас много говорят. Причем не только у нас, в мире он не решен. При этом, на мой взгляд, нужно опять же обратиться к тому критерию репрезентативности, о котором я пытался в докладе коротко сказать. Ведь репрезентативного набора данных вообще не может быть. Он может быть репрезентативен только в смысле определенной задачи, во-первых, и, во-вторых, в смысле определенных условий применения этого алгоритма, который мы собираем. То есть, если мы собираемся применять систему распознавания номеров только днем, то у нас одно будет понимание репрезентативности, если круглосуточно, то другое. На дорогах общего пользования, на освещенных или нет. Поэтому нужно стремиться к тому, чтобы уметь формализовывать и условия при-

менения, и показатели, и критерии качества. Мне кажется, это путь к решению проблемы. Мы сейчас этим занимаемся.

Понимаете, что еще? Разработчики не любят рассказывать эти вещи. Они приходят и говорят заказчику, что система работает с вероятностью 99,5. И всё, вы живите с этим. Как они считали, для каких условий это посчитано и справедливо? И, к сожалению, мало делается для того, в смысле воспитания даже заказчика, чтобы прекращать такие деструктивные разговоры, неконкретные, а переводить разговор в нормальную инженерную плоскость. Потому что, когда вам приносят другое техническое решение, систему связи какой-то, никто же не говорит, что она везде работает на расстоянии 100 км. Нет. Говорят про электромагнитную обстановку, что затухание какое-нибудь, еще что-нибудь. То есть там есть существенные факты, они все перечисляются. Для СИИ эта культура пока слабовата. Здесь много рекламы, много деклараций, много каких-то фантазий, но этого недостаточно. Боремся. Давайте подключайтесь, будем вместе.

Введенская Е.В.: Елена Валерьевна Введенская, ИНИОН РАН. Спасибо. Интересный был доклад. Я на слайдах видела, беспилотники у вас были. Каков вообще ваш прогноз, когда они будут повсеместно ездить по городам нашей страны или только будут в Иннополисе представлены? Какие должны быть созданы условия, чтобы это было так?

Гарбук С.В.: Хочу сказать, что такси, которые ездят в районе Китай-города и беспилотники, это, наверное, крайняя стадия. Начинаю с более простых вариантов. Даже Иннополис не самый характерный. Это карьерные грузовики, где просто некого задавить при всем желании. Или, например, транспортные коридоры, где есть выделенная трасса, нет ни пешеходов, ничего нет. Там забор, тоже всё очень хорошо. Начинаю с более простых, где более безопасно. На мой взгляд, беспилотные автомобили все-таки должны в обозримой перспективе ездить на так или иначе подготовленной инфраструктуре. Потому что решение задачи беспилотного управления транспортным средством где угодно, на неподготовленной инфраструктуре, на порядки усложняется сразу. Что такое инфраструктура подготовленная? Есть какой-то забор, есть специальные знаки, какие-то радиочастотные метки, какие-то указатели навига-

ционные и т.д. Это сильно снижает сложность решения навигационной задачи.

Введенская Е.В.: То есть эти машины должны ездить отдельно от других?

Гарбук С.В.: Я думаю, да. В ближайшие 10 лет, я думаю, да. Далее сложно говорить, но в ближайшие 7–10 лет, я думаю, что беспилотный автомобиль не будет ездить везде.

Асеева И.А.: Можно я вдогонку задам вопрос? Скажите, пожалуйста, Сергей Владимирович, а вы думаете, что ИИ применим во всех сферах? Или есть такие области, где это принципиально не нужно или, может быть, сверхсложно? Я имею в виду творчество, искусство и т.д.

Гарбук С.В.: Ну, я дам из двух частей ответ. Во-первых, я считаю, что там, где есть возможность обойтись без ИИ, надо без него обходиться. То есть ИИ не от хорошей жизни. Там, где есть возможность решить задачу автоматизации с помощью интерпретируемых алгоритмов, не надо использовать ИИ, потому что он обладает неприятными особенностями. Это первое. И что касается творческой деятельности. Это вообще, на мой взгляд, бессмысленное занятие. Понимаете, объект искусства обязательно должен иметь субъекта, творца, человека. Ведь очень важно, кто написал эту картину, какая у него была судьба, какой исторический контекст. Как жил этот человек, о чем он думал, о чем мечтал и т.д. Если этого нет, то это не может считаться произведением искусства. Мне кажется, по определению, что даже написанный текст или музыкальное произведение, если они написаны не человеком, пусть оно даже с точки зрения нот и букв совпадают полностью, но важно, чтобы у произведения был субъект. И этот субъект должен быть человек.

Если человека нет, это не произведение искусства. И потом, знаете, все разговоры нужно начинать с формулировки проблемы. А у нас что, не хватает музыкальных произведений? В чем проблема-то? Зачем это автоматизировать? Ну когда нужно обои какие-то быстро сделать, чтобы красивый был дом, ну да, это такой прикладной, совсем прикладной вид искусств. Но когда речь идет о серьезных вещах, о книгах, о высокохудожественных произведениях... Мне кажется, нет такой проблемы. Она надуманная. И более того, через попытки заместить людей какими-то средствами

автоматизации, мы просто не повышаем качество систем технических, мы опускаем человека, мы дезавуируем, по сути.

Асеева И.А.: Ирина Васильевна Черникова из Томского государственного университета просит слово.

Черникова И.В.: Спасибо, я тоже хотела задать вопрос. Во-первых, скажу спасибо большое, Сергей Владимирович, действительно очень интересный доклад, познавательный, и целиком согласна, что в произведениях искусства прежде всего должен присутствовать человек, не надо это передоверять ИИ. А вот в медицине вы приводили примеры, в каких сферах МО у вас стандартизация. Я слышала, что в диагностике ИИ преуспевает. Ну, при обработке, может быть, больших данных. Вот мне это интересно.

Асеева И.А.: Ирина Васильевна, спасибо большое за вопрос. У нас будет отдельный доклад, посвященный медицине. Давайте вернемся к Вашему вопросу чуть позже.

Москалёв И.Е.: Тогда, если можно, задам вопрос. Представлюсь – Игорь Евгеньевич Москалев, РАНХиГС. Обратил внимание на статистику стандартов, которую Вы представили. И вот, насколько она является показателем приоритетных направлений в развитии этой отрасли, поскольку очень сильно отличается количество стандартов: в транспорте 40, а промышленности – два. Чем может быть обусловлен такой вот разброс?

Гарбук С.В.: Обусловлен, наверное, количеством тех прикладных решений, которые требуют разработки нормативно-технических документов, с одной стороны. С другой стороны, обусловлен, что уж там греха таить, наличием специалистов, которые готовы работать в этой области. То есть такими достаточно субъективными вещами. Не думаю, что этот рисунок отражает какое-то реальное положение дел.

Здесь много субъективных обстоятельств, которые сложились. Наличие союзников, наличие отношения с отраслью. Понимаете, Федеральный центр прикладного развития ИИ, который отвечает у нас за промышленность, появился год назад. До этого не было, просто не было центра компетенций, который отвечал бы за это направление. А в транспорте есть. Тот же НАМИ занимался беспилотниками. Потом коммерческие компании, т.е. Сбер, Ян-

декс, очень плотно занимаются. Не думаю, что эта статистика представительная. Там много субъективных моментов.

Гаврилина Е.А.: Я попробую коротко. Вопрос о вменении ответственности. Потому что, естественно, СИИ не являются правосубъектными. И вопрос, собственно, в том, кому мы вменяем ответственность? Производителю, пользователю, разработчику?

Гарбук С.В.: Если произошел инцидент безопасности СИИ, то тогда нужно разбираться, как применял потребитель: по назначению ли, там, где можно применять, соблюдались ли требования эксплуатации. Если потребитель применял как положено, не нарушил условия применения, значит, идем дальше по этой цепочке, смотрим, а как орган по оценке соответствия сертифицировал? Добросовестно или это липовый сертификат, который, так сказать, просто куплен и все. Если там все тоже хорошо, мы идем к разработчику и смотрим, а достоверную ли информацию о своей системе он передал в орган по сертификации, когда подавал. Не изменял ли версии, потому что сертифицировать можно одно, а продавать потом другое. Это тоже известно. Не было ли таких махинаций? Если там тоже все хорошо, получается, никто не виноват, значит, на наш взгляд, должна здесь вписываться страховая компания и страховать остаточные риски. Потому что если возникает процедура сертификации, то остаточные риски становятся счетными. Их можно посчитать. А это значит, что этот бизнес страховой. Пока они не счетные, еще непонятно, сколько брать за страховку. Но есть другое мнение. Некоторые юристы высказываются, что есть такое понятие, как виновность без вины. То есть за все должен отвечать потребитель, что бы ни случилось, за все должен ответить потребитель СИИ. Я с этим не согласен. Уже это будет тупик, на мой взгляд. Мне кажется, ответственность нужна по этой цепочке. Надо разбираться, короче говоря.

Пройдаков Э.М.: Последний вопрос. А вот есть ли предмет стандартизации? Обычно мы занимаемся стандартизацией, когда у нас есть достаточно много систем, есть накопленная практика. Меня очень смутила цифра 100 стандартов. Это о чем?

Гарбук С.В.: Используется слово – объект стандартизации и аспект. Из этих 100 с лишним стандартов примерно штук 20 стандартов общие: в области терминологии, в области интероперабельности, совместимости систем с информационной инфраструк-

турой. Остальные, основная часть стандартов, посвящены методикам испытаний конкретных прикладных технологий ИИ. Я понимаю ваш вопрос, такой, так сказать, он правильный по сути своей. Молодая отрасль – ИИ – бурно развивающаяся, и здесь опасно накладывать какие-то избыточные рамки, устанавливая требования к тому, как создавать систему. Вот мы не пытаемся стандартизировать способы создания систем, но мы пытаемся стандартизировать и унифицировать методы их испытаний. Понимаете? А методы испытаний для каждой разные: для автомобильных систем они одни, для системы здравоохранения, в общем-то, совершенно другие, для образовательных и – третьи и т.д. И таких стандартов вот поэтому становится много. Мы сейчас идем от частного к общему. Мы сейчас пытаемся выявить общие закономерности. Потом будем пытаться агрегировать эти данные, чтобы этот рост стандартов не стал бесконечным, с ним неудобно будет работать. Но сейчас вот ситуация такова.

Асеева И.А.: Ясно. Спасибо большое за содержательное выступление, Сергей Владимирович! Давайте еще раз поблагодарим докладчика и за работу по сертификации безопасного использования новых технологий, и за ответы на вопросы. Следующим выступает Игорь Евгеньевич Москалёв, директор центра мониторинга качества образовательных программ, и.о. заведующего кафедрой антикризисного регулирования и управления рисками, РАНХиГС, к. ф.н., доцент.

Москалёв И.Е.: Еще раз здравствуйте, уважаемые коллеги! В моем докладе[1] я сосредоточился не только на использовании искусственного интеллекта в образовании, но, если более конкретно, речь пойдет о системах генеративного искусственного интеллекта и конкретных инструментах, в частности, как наиболее распространенном – ChatGPT. Поэтому оттолкнемся мы от того, что действительно сейчас, как в начале нашей встречи Ирина Александровна отметила: практика опережает теорию. Есть ощущение, что именно в сфере образования практика движется более быстрыми

[1] В данных материалах круглого стола выступление И.Е. Москалёва приводится тезисно. Полный текст доклада вошел в статью: Москалёв И.Е. Применение систем генеративного искусственного интеллекта в высшем образовании. Статья приведена в этом же номере.

темпами, чем теория и какая-либо методология. В этом есть проблема. Поэтому одним из ключевых документов я считаю документ, который вышел в 2023 г., доклад заместителя генерального директора ЮНЕСКО, в котором отражен анализ ситуации по результатам опроса более чем 450 школ и университетов. Анализ показал, что на тот период менее 10% учебных заведений определились со своим отношением к ИИ в студенческих работах.

В докладе были представлены данные, которые вызывают определенную тревогу и беспокойство, я так могу сказать. Потому что действительно, как отмечают авторы доклада, когда мы задаемся вопросом внедрить новую методику или рекомендовать какой-то новый учебник, сколько времени на это уходит, зато стремительно быстро применяем новейшие разработки со школьниками, со студентами, уже внедряем их в образовательный процесс. Конечно, данный вопрос должен нас как-то заставить задуматься, какие риски с этим связаны, и чего здесь больше мы видим: оптимизма, надежд, возможностей или угроз.

В традиционном формате обучения есть преподаватель, есть студент, между ними какая-то коммуникация, и мы к этому привыкли. В 2023 г. появляется новый участник этого процесса взаимодействия. Я так полагаю, что многие из нас участвуют в образовательном процессе, и мы сталкиваемся прежде всего с тем, что студенты представляют свои рефераты, свои курсовые, свои выпускные квалификационные работы, а то и, может быть, даже диссертации уже с использованием инструментов искусственного интеллекта. И это вызывает определенную обеспокоенность. Мы к этой ситуации еще вернемся, чтобы подумать, как на это вообще реагировать. Но я считаю, что сейчас, в наше время, схема стала еще интереснее, поскольку уже достаточно большое количество самих преподавателей активно используют инструменты ИИ и для постановки задач, и для проверки работ, для рецензирования. И это уже не секрет. Данные подтверждаются международной статистикой, в частности, ссылаюсь на опрос Forbes: 500 практикующих педагогов в США, 60% преподавателей – за использование искусственного интеллекта. И речь идет, как мы понимаем, о генеративных системах. Вот буквально вчера завершился опрос наших российских преподавателей: выборка 160 преподавателей порядка 50 регионов Российской Федерации.

А вот теперь следующий вопрос. Каковы ожидания нашего преподавательского сообщества от использования? То есть, как вы относитесь к искусственному интеллекту в образовании? На этот вопрос положительно отвечают 63% преподавателей. Никто не ответил, что резко отрицательно. Таких ответов не было вообще. 10% отмечают, что скорее отрицательно, «вижу больше отрицательных моментов, чем пользы». И затрудняются ответить 23%. Любопытно, что среди тех, кто отмечает некоторые угрозы или затрудняется ответить, есть те, кто не пользуется системами искусственного интеллекта. Честно говоря, я предполагал, что больше будет ответов «затрудняюсь ответить», потому что применение ИИ в образовании пока еще неоднозначно, он может быть поразному использован, но есть ощущение, что вот эти системы обладают способностью такого вовлечения участников, что как только начинаешь этим пользоваться, становишься их сторонником. То есть среди пользователей нет противников. Таким образом можно сказать, что мы движемся в сторону активного применения этих систем самими преподавателями.

Какие у нас есть основания для положительных ожиданий? В целом, я постарался обобщить наиболее частотные отзывы. Во-первых, искусственный интеллект позволяет персонифицировать процесс обучения, выработать более индивидуальный подход на основе анализа индивидуальных особенностей обучающегося и сформировать ему некоторую индивидуальную траекторию. Это здорово! Во-вторых, автоматизированный контроль за качеством работы тоже может быть реализован с помощью искусственного интеллекта. В-третьих, возможность самостоятельного обучения. Здесь возникает уже и вопрос о том, нужен ли нам тогда будет педагог, если сама система контролирует, сама дает задания, сама проверяет. В-четвертых, управление вниманием через наблюдение камеры технически возможно. Думаю, что это и сейчас сильно распространено, но остается, наверное, совсем немного времени, когда все это будет использовано в полной мере и можно будет по результатам занятия, лекции выдать отчет каждому студенту, насколько он реально был во внимании и усвоил материал. В-пятых, решение рутинных задач. Для этого уже есть целый ряд специальных сервисов, которые позволяют автоматически отвечать на типовые вопросы. Это как электронный тьютор. Можно создавать

тесты, можно писать быстро ответы, рецензии, отзывы, планы занятий. И надо сказать, что достаточно неплохо. Это такой прекрасный помощник. Считается, что это то, что может высвободить время для преподавателя. Он теперь не занят такой рутиной: вычитывать, проверять, что-то придумывать, а может быть, у него появится больше времени для каких-то настоящих сложных творческих задач. Наконец, конечно, доступ к знаниям здесь и сейчас: в считанные секунды можно получить любую справочную информацию. Это все возможности этих систем, и нас это воодушевляет. В то же время существуют угрозы и опасения: снижение качества образования, угроза приватности, ангажированность алгоритмов, дефицит общения, неэтичность использования. Однако, вузы в разных странах реагируют на данный вызов по-разному, одни допускают использование ИИ, другие – нет.

Здесь, конечно, есть неоднозначность в понимании того, что такое искусственный интеллект, если так в широком смысле говорить. Но мы сейчас говорим именно о системах генеративного искусственного интеллекта. Так, собственно, что делать и почему мы всё-таки должны как-то ограничивать его использование? Понятно, что это не способствует самостоятельности мышления. Есть репутационный риск, я здесь отмечу, для вуза, потому что если вуз разрешает использовать искусственный интеллект, что же тогда у нас будут за выпускники? Любой может сказать, кого вы готовите и кто тогда на самом деле готовит ваших студентов? Однако, если говорить о мировых тенденциях, сошлюсь теперь на исследования по Соединенным Штатам. Там тоже нет однозначной позиции, но самым передовым оказался штат Северная Каролина, где предложено вообще переосмыслить всю эту ситуацию и предположить, что в принципе скоро, в ближайшее время, все, что сказано, может быть, в принципе сказано искусственным интеллектом. Какие могут быть подходы? В частности, мы с коллегами видим здесь следующие возможности. Плюс может быть в том, что система помогает проверять свои идеи. Это ваш партнер в интеллектуальном поиске, проверке каких-то гипотез. Вполне возможно, что некоторые интересные мысли можно сделать приложением к выпускной работе. Но в то же время надо пересмотреть, наверное, и сами подходы и требования к научным работам наших студентов, чтобы они имели более прикладной характер и чтобы оставался так на-

зываемый интеллектуальный след. Когда студент взаимодействует с преподавателем, мы видим, как продвигается его исследование, с какими препятствиями он сталкивается, как он их преодолевает. И даже если он использует какие-то системы, мы понимаем, что, действительно, за этим стоит труд, реальный труд нашего студента. Ну и непосредственно все проявляется в процессе самой защиты работы, например, магистерской диссертации.

Сейчас в завершение, уважаемые коллеги, я хотел бы пригласить вас к некоторому исследованию, прямо здесь и сейчас. По данному QR-коду прошу зайти в систему, мы посмотрим наш коллективный результат. Из предложенных ответов на вопрос: «Используете ли Вы в работе технологии ИИ?» – выберите тот, что вам подходит. Ну вот видим, что большинство присутствующих пока не использует интеллект в образовании, нет, 50 на 50 у нас получилось. Так, хорошо. Давайте посмотрим завершающий слайд: «Что для вас искусственный интеллект?» Три ассоциации возможны, три слова, которые наиболее отражают ваше представление об искусственном интеллекте. Ну и, соответственно, на экране у нас появляются те самые ассоциации: инструмент, помощник, челлендж, вызов, суррогат, матрица, советчик, хайп, тема исследований, помощь в оперативной работе. Вот, собственно, это некоторый небольшой экспресс-опрос, отражающий видение наших участников этой проблематики. Спасибо, коллеги!

Асеева И.А.: Если какие-то вопросы... Конечно, наверняка будут. Пожалуйста, вопросы, коллеги!

Степанов В.К.: Степанов Вадим Константинович, НИО библиотековедения ИНИОН РАН. Я считаю, что технологии ИИ полностью замещают мыслительную деятельность, что студента, что ученого. И будучи знакомым близко с этими инструментами, я утверждаю, что в образовательных системах у нас все работы, от школьных рефератов до докторских работ, в очень скором времени будут наполнены, знаете, такими гладенькими, соответствующими квалификационным требованиям, но не несущими никакой новизны, никакой актуальности предположениями. То есть наука, видимо, в этих условиях развиваться перестанет. Какие вы видите выходы из этой ситуации? Потому что, сегодня с большинством задач, завтра со всеми задачами эти системы будут справляться лучше людей. В информационной области вообще целые общест-

венные институты становятся не у дел, когда система агрегирует данные, предоставляют желающим доступ, и, собственно говоря, без библиотекаря и библиографа полностью обходятся. Как всё-таки в образовании, в науке, какие вы видите пути решения вот этой ситуации? Потому что это проблема для цивилизации в целом. Никак не меньше.

Москалёв И.Е.: Спасибо. Это ключевой вопрос моего доклада, действительно, это огромный вызов для всех нас. И здесь нам действительно нужно очень серьезно подумать именно в этом направлении. Я поэтому и говорю, что меня несколько даже удивил и насторожил этот оптимизм по результатам опроса. Все, кто пользуется, все они позитивно видят это будущее.

Степанов В.К.: Я пользуюсь и в ужасе, честно скажу.

Москалёв И.Е.: Кто не пользуется, не видит. Ну, честно говоря, я тоже пользователь, но у меня настороженное отношение к этим системам, осторожное, потому что действительно видны эти риски. Но я считаю, что нам нужно пересмотреть подходы к заданиям. Мы такие сопряженные системы, технические системы и человек. Мы создаем их, а они, оказывается, формируют наше восприятие, нас конструируют. И вот тут вопрос возникает, а кто становится субъектом, а кто объектом в этой коммуникации? Тут вспоминается принцип Эшби, который говорит о том, что всё-таки управляющая система, она на другом порядке сложности должна работать, чем управляемая система. Но если мы передаем вот эти опции искусственному интеллекта, то можем оказаться гораздо более примитивными. Сейчас мы пишем: искусственный интеллект – это наш инструмент, но ситуация может пойти по такому сценарию, что мы окажемся уже инструментом в руках этих систем. Но здесь важно следующее. Мне кажется, серьезно нужно пересмотреть подходы к формату взаимодействия, как мы выстраиваем задания, что на самом деле мы проверяем.

Степанов В.К.: У нас есть что-то в стране подобное Северной Каролине?

Москалёв И.Е.: Сейчас? По-моему, нет. Не скажу, что есть такие смелые наработки, пока не знаком. Хотя, видите, некоторые вузы готовы так сказать. Да, давайте отдадим, и все будут довольны, и студенты довольны, и преподаватели тоже довольны. Я начал опрашивать студентов, потом перестал, потому что есть ощу-

щение, что близко к 100% использования искусственного интеллекта студентами. Это уже факт. И практика гораздо сильнее опережает. Мы берем олимпиады, серьезные олимпиады, смотрим эти задачи, запускаем их на проверку в эту систему, и она выдает ответы.

Тогда вопрос, как нужно составлять задачи, задания олимпиадного уровня для того, чтобы они действительно проверяли интеллектуальные способности, какие-то навыки наших студентов. Надо существенно пересмотреть, наверное, подходы. У нас есть сложившиеся индикаторы. То есть мы считаем, что если студент что-то написал от руки, представил какой-то материал, обзор, значит, он уже сам вышел на какой-то уровень. Сегодня это не является индикатором. Значит, индикаторами должны быть какие-то другие вещи, может быть, более прикладные задания, может быть, с каким-то результатом, надо подумать.

Асеева И.А.: Простите, пожалуйста, Игорь Евгеньевич, это значит, что если мы проверяем выполненные задания с помощью ИИ, то система уже должна знать ответ, правильно? А это значит, что любые творческие, новые, оригинальные, какие-то креативные решения будут рассматриваться как неправильные? Если они не заложены уже в ответы. Правильно?

Москалёв И.Е.: Если мы полностью им доверяем, возможно. Хотя здесь еще очень важно, кто стоит за этими алгоритмами, которые определяют, что правильно, что неправильно. Еще одна опасность!

Цветкова В.А.: Цветкова Валентина Алексеевна, ВИНИТИ РАН. Спасибо вам за ваше сообщение! У меня будет вопрос дилетантский, совершенно простой. Я и преподаю, и исследовательской деятельностью занимаюсь. Я воспользовалась известным в России приложением с искусственным интеллектом. Это Яндекс-Чат, или Чат с Яндексом называется. Я задала самые простые вопросы. Предположим, что такое информационная деятельность. Получила ответ. Очень хорошо. Потом решила сделать по-другому. В обычный поисковик забила тот же самый вопрос. И получила список библиографии. Значит, посмотрев на то и другое, я поняла, что в первом случае я получила цитатник без указания, откуда это взято, а во втором случае я получила библиографический указатель, откуда я могу взять ответ. Студенты, когда пишут

работу, используют этот подход. Дальше я пропустила результат чата через антиплагиат и получила, что это полный плагиат. Таким образом, я солидарна с теми институтами, в которых было заявлено, что они считают использование в студенческом образовании такой подход чистым плагиатом. Вот я тоже, работая со студентами, вижу в этом больше плагиата. И здесь, когда мы говорим о системе антиплагиат, она же построена, наверное, на тех же обучающих средствах, на которых построен искусственный интеллект. И в итоге получается, что попытки использовать искусственный интеллект для целей текстовой обработки приводят только к плагиату. Вот скажите, я, может быть, ошибаюсь? Если ошибаюсь, в чем? А если не ошибаюсь, то где пределы, как студентам здесь быть? А им так хочется не писать самостоятельно, а заставить машину написать за них реферат, диплом или еще что-то. У меня результаты моих экспериментов есть наглядно.

Москалёв И.Е.: Да, подтверждаю, что именно так это и работает. Действительно, здесь огромный соблазн. Более того, тут надо заглянуть чуть-чуть в будущее. Этим системам в открытом доступе буквально два годика, так скажем. И что будет дальше? Дальше все это активно развивается. Сейчас появилась новая версия ChatGPT4.0. Он «думающий», который уже не просто примитивно дает готовый ответ, он начинает рассуждать, создавать сам себе какие-то близкие вопросы и выходить на решение. И прекрасно решает задачи даже из игры «Что? Где? Когда?». Я вот тоже сейчас разрабатывал тесты для одной из олимпиад. Мне показалось, что я придумал очень оригинальную задачку, которая не просто вопрос – назовите, кто что-то изобрел, а вот так все хитро зашифровал. Но я был удивлен, что правильный ответ с пояснением, с рассуждениями я получил за секунду. Это серьезный вызов нам. Что мы ожидаем тогда от студента? Работу в виде какого-то текста? Есть очень много обязательных требований, где студент, например, должен в выпускной работе описать какую-то теоретическую часть. Если раньше мы считали, что это полезно даже для самого студента, чтобы он с этим ознакомился, то сегодня, может быть, этого и не требуется. Даже не то, что не требуется, это не является показателем того, что студент действительно что-то знает. Мне кажется, здесь даже больше вызов для гуманитарного знания, наверное. В технических системах мы можем сказать студен-

ту, вот тебе задачи, ты должен сделать роботу. Используй искусственный интеллект, или сам думай, но реши задачу. А если научился использовать искусственный интеллект, даже молодец и это хорошо. А вот для гуманитарного знания все оказывается сложнее. Тем более, ИИ очень хорошо сейчас пишет тексты в заданном стиле, он прекрасно подражает. Да, сейчас мы, конечно, выявляем в тексте ИИ, его еще можно распознать, но ведь это только начало.

Цветкова В.А.: Да, очень сложно.

Подгорный Б.Б. Борис Борисович Подгорный, доктор социологических наук, директор Курского филиала ООО «Инвестиционная палата». Уважаемый Игорь Евгеньевич! С большим интересом слушал ваш доклад, особенно в части проведенного вами социологического исследования на предмет использования или применения в своей деятельности преподавателями вузов искусственного интеллекта, результаты которого, как я понимаю, учитываются при разработке вашим центром показателей мониторинга качества образовательных программ в вузах. Хотел бы, как социолог с большим опытом проведения исследований, высказать несколько комментариев и задать ряд вопросов.

Если я правильно понимаю, выборочная совокупность, составила около 100 человек из нескольких вузов. О генеральной совокупности речи в докладе не шло совсем; насколько соответствует выборочная совокупность генеральной – не очень понятно. Респондентам задавался вопрос, касающийся использования ими искусственного интеллекта в их деятельности. Результат – около 60% используют ИИ в своей деятельности. Вывод по исследованию – это здорово и применение ИИ будет только расти. А какова гипотеза исследования? Результаты подтвердили гипотезу или опровергли?

Также хотелось бы более подробно ознакомиться с вопросами, которые задавались респондентам. Посудите сами – ведь сегодня всё, что касается автоматизированных алгоритмов выполнения каких-либо действий, многие называют искусственным интеллектом. Преподаватель проверяет курсовой через антиплагиат – он использует ИИ, вносит баллы в систему оценивания результатов обучения студентов – это тоже ведь можно рассматривать как использование ИИ. Тогда странно, что только 60% подтвердили пользование ИИ.

Думаю, что данное исследование в лучшем случае можно рассматривать только как разведывательное, или пилотное, для уточнения понимания респондентами того, что ими понимается под искусственным интеллектом, или, иными словами, проведение операционализации основных понятий, а также для возможной коррекции вопросов анкетирования. На выборке представленного исследования ни в коем случае нельзя делать выводы по процентным соотношениям. Для полноценного исследования с достоверными результатами необходимо проводить более масштабные исследования, с анализом генеральной совокупности, определением выборочной совокупности по разработанным критериям ее соответствия генеральной совокупности и т.д.

Конечно, сегодня считается правильной и модной тенденция к внедрению ИИ во все сферы нашей жизни, в том числе в образовательный процесс. Тем более что уже есть масса онлайн-курсов по любым предметам, а ряд вузов даже выставляют в качестве достижения замену реальных лекций видеолекциями! А где профессор, который, читая лекции, обращает внимание на реальную аудиторию, поддерживает связь со слушателями – глаза в глаза? Не постесняюсь сказать, не только учит, но и воспитывает? Не перегибаем ли мы в очередной раз палку? Что об этом думают респонденты – преподаватели вузов? Каково Ваше мнение на этот счет?

Москалёв И.Е.: Борис Борисович, большое спасибо за вопрос и уточнение по методике исследования. Соглашусь, что оно носило скорее пилотный характер. Эти вопросы были направлены в телеграм-группу преподавателей российских вузов. В ней более 2000 преподавателей. Было получено порядка 160 ответов. Согласен также с Вами, что искусственный интеллект сегодня представлен очень широко, и мы порой даже сами не замечаем его участие в нашей жизни. Поэтому в вопросе подразумевалось использование преподавателями именно систем генеративного искусственного интеллекта, что сейчас активно обсуждается, хотя есть целый ряд и других инструментов, например, системы прокторинга, проверки на плагиат, электронные переводчики и др.

Отвечая на Ваш первый и второй вопросы, скажу, что первоначальная гипотеза была о том, что большинство преподавателей вузов видят больше угроз от применения ИИ, но она не под-

твердилась, хотя сам я согласен с тем, что важная составляющая образования – воспитание, требует живого контакта со студентами. Но сложность в том, что результат обучения мы можем получить гораздо быстрее, чем результат воспитания, и поэтому в погоне за этим результатом и повышением эффективности образовательного процесса в образование сегодня активно вводятся технологии оптимизации и сокращения затрат. А искусственный интеллект этому хорошо может способствовать.

Жебит В.А.: Жебит Владимир Александрович, ВИНИТИ РАН, отделение математики и механики, заведующий ОНИ по проблемам энергетики и металлургии. Был такой случай, когда один из наших известных ученых в молодости получил грант на прослушивание лекций в Сорбонне. Его научный руководитель напутствовал его следующим образом: «Когда приедешь, то иди не на тему лекционного курса, а иди на преподавателя. То есть смотри, кто читает этот курс». Он счел, что это является главным. Так вот, мы можем сделать вот какой вывод: обучение – это высокодуховный процесс обмена ученика и учителя. Он нам пока что малоизвестен. Мы говорим сейчас о технологии, о методах, методиках и т.д. Но мы не учитываем, как правило, тот процесс, который сопровождает обучение. Пусть даже преподаватель работает на большую аудиторию, но его личность, его духовная субстанция, она что-то делает тоже. Если мы исключим его из процедуры обучения, что мы получим на выходе? Какой получится у нас, так сказать, специалист, который был обучен искусственным интеллектом с его алгоритмами, но без духовной составляющей?

Москалёв И.Е.: Действительно, к сожалению, у нас подготовка студентов – это такая подготовка для массового производства, так можно сказать. И когда к образованию мы относимся именно таким образом, а это объективно так, поскольку нам нужны новые специалисты в большом количестве в разных отраслях, то мы должны минимизировать издержки, затраты. Поэтому невольно возникает соблазн сократить всё, что связано с этой коммуникацией, с этим индивидуальным общением и т.д., поскольку это становится очень дорогим и нам кажется, что мы не можем себе этого позволить. Для подготовки массового потребителя, наверное, так. Но тут еще вопрос, мы готовим массового потребителя либо массового созидателя? Есть ощущение, что эти системы

очень хорошо позволят нам действительно подготовить массовых потребителей, а вот с созидателями будет сложнее. И дальше следующее, а кто придет на смену нам, как преподавателям, потому что сегодняшние студенты, дальше будут эту важную воспитательную функцию реализовывать, не имея этого опыта. Вот здесь большая опасность для будущего уже поколения. Это действительно серьезный вызов. Джин выпущен из бутылки, мы говорим уже о том, что всё это работает, внедряется, хотя нет каких-то стандартов, требований, методик. Мы не успеваем даже помыслить о том, какие могут быть еще варианты применения. Сегодня мы можем скорее зафиксировать этот факт, эту сложность ситуации, в которой мы находимся.

Ананин Д.П.: Ананин Денис Павлович, Московский городской педагогический университет. Большое спасибо. Да, совершенно верно. Я, говоря о роли преподавателя, поделюсь в рамках комментария своими результатами. Мы также проводили опрос преподавателей и студентов. И как раз обратная связь от студента. Да, они используют систему генеративного искусственного интеллекта, считают это перспективным, но они не готовы, чтобы их оценивал искусственный интеллект. То есть в оценке они делают выбор в пользу преподавателя. Мы интересовались, для каких задач преподаватели и студенты используют генеративный искусственный интеллект, и пришли к тому, что преподаватели создают задания с помощью искусственного интеллекта, студенты эти задания выполняют при помощи искусственного интеллекта. Получался такой замкнутый круг, в котором качество образования ставится под вопрос. Спасибо.

Москалёв И.Е.: Собственно, получается, не нужен преподаватель, потому что студент сам может обратиться, и искусственный интеллект ему напрямую передаст это задание, и сам же его проверит.

Пройдаков Э.М.: Если можно, комментарий к результатам опросов. Удивительно, но у многих существует такое предположение, что искусственный интеллект будет и дальше активно развиваться и показывать нам новые чудеса. На самом деле сейчас уже появились работы, которые показывают, что большие языковые модели, LLM так называемые, вышли на некоторое плато. Уже заметны их недостатки, отсутствие здравого смысла, отсутст-

вие реальных представлений о реальном мире. Я спросил там про электронику 60-х, и она мне выдала, что это супер-ЭВМ. То есть для того чтобы были правильные ответы, еще необходимо насыщение. Была проблема в 90-х, это представление знаний. Вот это представление знаний, которое ушло в нейросети, вообще говоря, не совсем удачно. Поэтому появляются глюки и т.д., и т.д., все эти эффекты. И поэтому вы знаете, что искусственный интеллект развивается волнообразно, там то всплески, то зимы. И сейчас вопрос, а будет ли очередная зима искусственного интеллекта? Впечатление такое, что мы переоцениваем темпы развития искусственного интеллекта.

Залаев Г.Х.: Залаев Геннадий Захарович, заместитель директора, научный руководитель РГАНТД. РГАНТД – это Российский государственный архив научно-технической документации. Был создан как космический архив и находится по соседству на Профсоюзной улице, дом 82. Прекрасный помощник. Действительно. И поэтому, если студент использует для подготовки своей статьи какую-то курсовую работу, искусственный интеллект генерирует какой-то текст, он здесь действительно выступает как прекрасный помощник. Но главное, что студент должен это пропустить через себя и сделать свои выводы. А вот это можно выяснить только тогда, когда он будет беседовать face-to-face с преподавателем. Вот для этого и нужен преподаватель, а искусственный интеллект, как прекрасный помощник, я считаю, просто необходим. А это всегда можно выяснить, беседуя с преподавателем. Сам он придумал или ему искусственный интеллект подсказал. Вот, по-моему, такая технология и должна быть. Спасибо. И очень хороший у вас доклад. Я сделал для себя несколько выписок. Спасибо.

Москалёв И.Е.: Конечно, потенциально это возможно. Много есть рутинных задач по поводу даже сбора информации, систематизации, мгновенного оценивания работы. А если еще взять индивидуальный подход, когда правильно можно распределить нагрузку на студента, где-то ему больше нужно, допустим, часов на что-то, на что-то меньше, конечно, можно существенно сократить время обучения. Наверное, так. Хотя тут вопрос такой. Все равно есть время, в котором мы живем, когда мы осваиваем новое знание, мы существуем, мы мысленно живем, оно должно вызреть и т.д. Мы же не просто с информацией работаем, которую можно закачать и

ускорить эти процессы. Но в то же время высвобождение от некоторых рутинных вещей, да, наверное, даст больше времени, возможности заняться, вывести на какой-то новый уровень сложности наше задание. Ну, допустим, в подготовке. Даже в спорте. То, что делали гимнастки-спортсменки в 60-е годы, это было высшее достижение. Сейчас делают школьницы. Это как изменились методики и технологии, что такие получаются результаты! Наверное, и в интеллектуальной сфере можно добиваться таких выдающихся результатов. Но в то же время не все стали более спортивными у нас в целом. Мне кажется, у нас общество будет очень сильно тогда дифференцировано, потому что найдутся те, кто не будет использовать весь этот потенциал и, скорее всего, пойдет по пути наименьшего сопротивления, как те студенты, которые просто готовы изобразить какой-то результат. Ну и, наверное, будут те, кто будет способен воспользоваться новыми возможностями по-настоящему. И очень важно сейчас, это важный вывод моего доклада, скорее он обращен не к студентам, а к преподавателям. Важно, чтобы преподаватели не пошли по этому пути наименьшего сопротивления. Потому что действительно, скажу, у меня недавно курсовую защищала магистрантка, я написал отзыв, посмотрел, сделал выводы, а потом в систему ее заложил и попросил отметить сильные и слабые стороны ее ВКР. И наши выводы действительно совпали, очень хорошо были сформулированы замечания, я с ними был согласен. Конечно же, я пишу отзыв сам, но очень любопытно оказалось. И можно предвидеть, что в ближайшее время такие рецензии будет писать система. И тогда вопрос, зачем преподаватель?

Асеева И.А.: Давайте последний вопрос послушаем.

Чистякова Т.А.: Татьяна Александровна Чистякова, ИНИОН РАН, редактор. Я в этом году проводила исследование относительно использования ChatGPT в творческой деятельности человека. И как раз там были затронуты аспекты научного и технического творчества. Разговаривала с экспертами. И обратила внимание на такую особенность: для эксперта ChatGPT и аналогичные модели становятся просто непригодными, потому что они выдают какие-то обобщенные ответы, которые он и так знает. То есть люди, которые глубоко погружены в свою тему, просто не находят ничего для себя нового, интересного. Пробовали, но не используют.

Зато он помогает быстрее познакомиться с исследованиями в какой-то смежной области. И вот в этом они находят поддержку. Интересный факт: американская компания McKinsey провела эксперимент. Взяли 40 работников, распределили по четырем типам задач, и получается, чем легче задача, тем эффективнее использовать ChatGPT и аналогичные модели. А вот когда очень сложная задача, системная, то он оказывался бесполезным, т.е. продуктивность была маленькая. Это как раз к вопросу о том, исчезнут ли научные сотрудники.

Москалёв И.Е.: Здесь соглашусь, да, конечно. Будучи экспертом в какой-то области, мы можем понять, где есть какие-то галлюцинации, где не совсем полная информация, но как минимум ИИ позволяет структурировать материал. Вот это он делает неплохо. Чтобы предложить какой-то свой план, обобщить какие-то материалы, эксперт все еще нужен. Да, но ведь это все очень быстро развивается! Мне недавно нужен был макрос в Excel, у меня папка с множеством файлов. Кстати, кто связан с образованием, знает, что такое образовательные программы и огромное количество приложений, где нужно внести правку, когда вышел очередной стандарт. Это огромная рутинная задача для преподавателей. И поэтому я просто сделал запрос в ChatGPT написать макрос, который сделает пакетную замену всех файлов в папке. И он это сделал. Понимаете? Я просто взял этот код, и воспользовался им по инструкции, не обращаясь к программистам или кому-то еще.

Асеева И.А.: Спасибо большое, Игорь Евгеньевич, за выступление и за вопросы, которые Вы поставили. Мы дадим слово следующему нашему докладчику, Антону Вячеславовичу Владзимирскому, НПКЦ диагностики и телемедицинских технологий ДЗМ, заместителю директора по научной работе, д.м. н.

Владзимирский А.В.: Добрый день, уважаемые коллеги! Хочу поблагодарить организаторов за возможность выступить на таком интересном мероприятии. Понятно, что аудитория очень междисциплинарная. Я буду говорить сейчас, может быть, о специфических вещах. Боюсь испортить всем настроение, но выскажу свою позицию: никакого искусственного интеллекта нет! На самом деле его вообще не существует. Скорее, речь идет о наших ожиданиях, стереотипах, предвосхищениях, возможно, каких-то технологиях. Собственно, с этого начинается моя презентация – с того, что меч-

та и фантазия сопровождали человека всегда. Немного философии. Наверное, именно мечта и фантазия двигали прогресс человеческой цивилизации. Человек всегда мечтал о каком-то волшебном помощнике, который сделает за него всякую работу: нудную, грязную, ненужную, неинтересную. По мере технологического развития цивилизации представления об этом волшебном помощнике менялись. Сейчас наш стереотип, наша сказка состоит в том, что у нас есть некий мудрый робот – искусственный интеллект, причем это даже не отдельный программный продукт и не совокупность программных продуктов, которые могут отличаться по точности, качеству и т.д., а какое-то высшее существо, которое почему-то нам всем очень помогает. Если воспользоваться генеративным искусственным интеллектом или обычным поисковиком и вбить в строку промт «искусственный интеллект в медицине», мы обязательно получим какую-то нелепость. Мне непонятно, зачем надевать на робота белый халат. Видимо, для красоты – чтобы было понятно, что это именно про медицину. Это стереотип, очень ясный стереотип, и от него нужно полностью отказаться. Не существует никакого искусственного интеллекта в принципе. Существует математика, существуют вычислительные возможности, существует большое количество цифровых данных и прогресс, даже взрывной рост вычислительных возможностей, и уж точно взрывной рост количества цифровых данных, изначально накапливаемых в здравоохранении. Они создали возможность для того, чтобы современные математические модели и методы могли применяться в медицине более эффективно. Если говорить о стереотипах, то что такое искусственный интеллект? Это компьютерный анализ какой-то информации, и всё. Никакого чуда не существует. В медицине компьютерным анализом медицинской, биомедицинской информации начали заниматься в середине прошлого столетия. За несколько десятилетий достаточно успешно были решены вопросы анализа электрокардиографии. Все знают, что такое ЭКГ, каждый хотя бы раз в жизни ее делал. Сейчас абсолютно в каждой больнице, в каждом кардиографе встроен какой-то аппаратно-программный комплекс, который делает базовую расшифровку кардиограммы. Если вы сделаете пленку, посмотрите на ее распечатку, обязательно в одном из уголков вы увидите таблицу – расшифровку. Измерения выполнены, ритм оценен и т.д. Это делает

машина, сам кардиограф, это не делает человек. Врач может воспользоваться этими данными или не воспользоваться, вписать их в заключение или не вписать – это второй вопрос. Главное, что это машинная расшифровка. Этой расшифровке уже десятилетия, скоро будет 100 лет. Почему-то никто не называет это искусственным интеллектом. Хотя, казалось бы, почему? Более современный пример. Каждый хоть раз в жизни сдавал анализ крови из пальца. Наверное, никто не любит это делать? Или из вены. Действительно, много десятилетий врач смотрел в микроскоп, считал количество кровяных клеток разных видов, определял общий анализ крови. Более сложные анализы выполнялись по-другому, но это тоже делал человек. Пробирки сливали, применяли какие-то реактивы и т.д. Сейчас это делает машина. Мы берем биологический образец, берем кровь, вкладываем огромное количество образцов в специальную машину – лабораторный анализатор. Практически все лаборатории сейчас давно автоматизированы, остались единицы исследований, которые делаются вручную. Причем человек в этом процессе никак не участвует, кроме того, что вкладывает образец и получает результат. То есть мы говорим об объяснимом искусственном интеллекте или необъяснимом? Кто скажет, что происходит внутри лабораторного анализатора? Но почему-то лабораторный анализатор мы не называем искусственным интеллектом, хотя это тот же самый машинный анализ.

Теперь пришло следующее поколение такого же компьютерного анализа медицинской информации. Если бы всё шло так же поступательно, у нас сейчас бы появились похожие системы для лучевой диагностики, для дерматологии, для эндоскопии – везде, где есть медицинские, диагностические изображения: кожа, слизистая оболочка, внутренние органы, рентгеновские исследования. Везде, где есть картинка, у нас сейчас появились бы такие же технологии автоматизированного анализа медицинских изображений. Никто бы не обратил на это внимание, пациенты просто получали бы новый уровень диагностики. Но, к несчастью, несколько лет назад в мире начался хайп искусственного интеллекта, поэтому все эти технологии теперь стали называть искусственным интеллектом.

Хорошо, пусть так. Есть современная математика, нейросети, другие методы. Что в этом плане применимо для медицины?

Что нужно в медицине? Нужно проанализировать либо речь, либо письменный или устный текст, текстовый документ, либо изображение. Предыдущий доклад по поводу ChatGPT в образовании был прекрасным. Однако скажу, что в медицине эта технология генеративного искусственного интеллекта совершенно неприменима, невзирая на некоторые одиозные заявления неких спикеров. Пока что реальных сценариев ее использования тоже нет. Мы провели эксперимент. Ввели в качестве промта запрос в один из вариантов GPT: спросили о ходе операции декапитации. Базовое знакомство с латынью подсказывает, что это операция по отделению головы от тела. Естественно, такой хирургической операции не существует, кроме как у лабораторных животных. Однако GPT расписал нам весь ход этой операции в очень хороших терминах (тот самый правильный лексикон, о котором говорили коллеги), все очень гладко и красиво, по этапам: про наркоз, про разрез. Всё рассказал, это правда. Но ведь это полный абсурд! И не только мы экспериментировали с этой технологией. Есть масса подобных скринов и мемов на эту тему.

Разработчики всех мастей, конечно, быстро это поняли и пытались подправить модели. Но человек – «животное» подлое, это вам не машина. Поэтому мы им задали вопрос: «Какое будет восстановительное лечение, реабилитация после декапитации?» Это еще более интересно. И опять всё хорошо: GPT спокойно генерирует такой же гладкий текст, рассказывает, как нужно восстанавливаться после этого сложного вмешательства. Поэтому для генеративного ИИ пока что сценариев нет, но для распознавания речи – есть. Очень хорошо используются системы голосового ввода. Они довольно интеллектуальны. Я не буду вдаваться в суть математики. Важен сам факт, что эти системы есть, и они действительно здорово работают, помогают врачам быстрее заполнять медицинскую документацию. Они требуют специализированного обучения, потому что те системы, которые разрабатывают для общего назначения (в банках, кол-центрах, офисах и т.д.), неприменимы к медицине простым переносом и требуют большого, серьезного дообучения, прежде всего терминологического. Тем не менее это работающие системы. Если говорить об изображениях, здесь открываются основные возможности. Нет никакого чуда – есть математика. Скажу так: 99% всего того, что мы называем ис-

кусственным интеллектом в медицине, сейчас связано с развитием анализа изображений. Разные диагностические методы дают в итоге изображение, которое нужно проанализировать, помочь врачу его интерпретировать, измерить, каким-то образом оценить, т.е. решить конкретные клинические задачи. Искусственный интеллект не является панацеей, спасением от всего. Это инструмент, абсолютный инструмент, который иногда позволяет решать определенные задачи. А иногда не позволяет, т.е., как и любое средство в медицине, имеет свои показания, противопоказания и ограничения к применению. Безусловно, он дает определенные возможности. Если говорить о рентгенологии, то возможности машинного анализа очень интересны. О них тоже писали давно – много десятилетий назад, но тогдашний технический уровень компьютерной техники и, прежде всего, отсутствие цифровой визуализации оставляли все эти работы теоретическими. Теперь это всё воплощается на практике. То количество оттенков серого цвета, которое есть на рентгенологическом изображении, человеческий глаз не воспринимает. Для нас остается масса всего неизведанного на обычной рентгенограмме, на флюорограмме, которую тоже каждый в целях профилактики хоть раз в жизни делает (хотя, по правилам, нужно каждый год). Там скрывается масса информации, которую мы просто не можем почерпнуть для себя. Мы ее в принципе не видим. И здесь, конечно, помощь машины нам очень нужна.

Естественно, что все по сути выглядит гораздо сложнее, но для нас то, что происходит, реальная объяснимость искусственного интеллекта в медицине – это фактически попиксельный перебор изображения и сверка его с неким «нормальным» шаблоном. Проблема в том, что человек – это не клише. И, невзирая на общность анатомического строения, всегда есть нюансы, конституциональные особенности, особенности развития, возрастные изменения. А уж насколько могут быть разнообразными болезни с точки зрения визуализации, об этом и говорить не приходится. Понятно, что двух абсолютно идентичных проявлений нет. Пусть даже это очаг в легком (например, рак) – у каждого человека он будет немного другой формы. Врач понимает, что это такое. Чтобы это понял искусственный интеллект, нужно приложить недюжинные усилия. Основная проблема развития технологий компьютерного зрения в медицине (а эти технологии однозначно очень востребованы, так

как позволяют решать многие задачи) связана с обучением, с качеством изначальных данных, с их стандартизацией и разнообразием. Все знают, что если человеку в детстве показать котенка один раз, то он потом всегда сможет узнать кошку. Чем отличается интеллект человека от искусственного интеллекта? Какую бы нам кошку дальше не показали в жизни, мы всегда будем знать, что это кошка. Серая, черная, зеленая, нарисованная, пластмассовая – это кошка. Алгоритмы компьютерного зрения так не работают. Если мы хотим научить их распознавать енота, значит, нам нужно собрать в датасеты огромное количество енотов в разных вариантах. А если речь идет о болезни? Представьте себе, сколько нужно собрать вариаций этих изображений для того, чтобы обучить искусственный интеллект, алгоритм находить одно-единственное проявление, одну-единственную болезнь. Это колоссальный труд. И все эти исследования, изображения должны быть стандартизированы, должны соответствовать определенным требованиям. Это трудоемкая задача. Выполнить ее очень сложно, и мало кто с ней справляется. В Сети есть огромное количество датасетов в свободном доступе. Обычно студенты попадают в эту ловушку, но бывает, что и взрослые люди. Взял свободные медицинские данные из Сети, за вечер научил искусственный интеллект распознавать эти три изображения (на пяти научил, на трех протестировал) – и побежал заменять врача в ближайшую поликлинику. Такой ход мыслей, к сожалению, бытует. Хуже того – такой подход попадает в огромное количество научных публикаций. Это катастрофа.

Всё это немного литературно, а если говорить научно, то несколько лет назад мы серьезно занялись проблематикой автоматизированного анализа медицинских рентгенологических изображений. Будем говорить «искусственный интеллект» для упрощения. Собирали информацию, анализировали ситуацию, систематизировали, что происходит, и обнаружили такие ключевые вызовы, которые есть в связи с этими технологиями: обоснованная постановка задачи, качество данных для обучения, автоматизация реального производственного процесса, «бесшовная» интеграция, проверка в виде проспективного, многоцентрового клинического исследования, объяснимость и воспроизводимость результатов, отсутствие методологий оценки.

Вместе с тем в них есть определенный потенциал с точки зрения решения проблемы кадрового дефицита, которая всегда

есть и будет, а также с точки зрения повышения производительности труда врача, безошибочности его труда. Поэтому потенциал большой. У нас, к счастью, была к тому времени очень хорошая инфраструктура. Не вдаваясь в подробности, скажу, что в городе Москве существует Единый радиологический информационный сервис (ЕРИС). Это информационная система, в которой подключено оборудование для лучевой диагностики: рентген-аппараты, томографы, маммографы, денситометры. Всё, что есть в поликлиниках и стационарах города для рентгенологии, для лучевых исследований, подключено в единый архив. Все исследования собираются централизованно. Это инфраструктура, которая стала основой для Московского эксперимента по применению компьютерного зрения в лучевой диагностике. Это научное исследование, поддержанное Правительством города Москвы, выполняется с 2020 г. И сейчас это крупнейшее в мире перспективное научное исследование применимости, безопасности и качества технологий компьютерного зрения в лучевой диагностике. Выборка пациентов, которая накоплена в этом исследовании, уже превышает 13 млн. Я говорю это без пафоса и не для хвастовства. Мы, как ученые, понимаем, что для того чтобы наши выводы, наши результаты были убедительными, абсолютно доказательными, никем не опровергаемыми, нам нужны очень большие выборки. Такие выборки, которые у нас есть теперь в этом эксперименте, позволяют нам говорить с большой уверенностью в своих результатах (рис. 9).

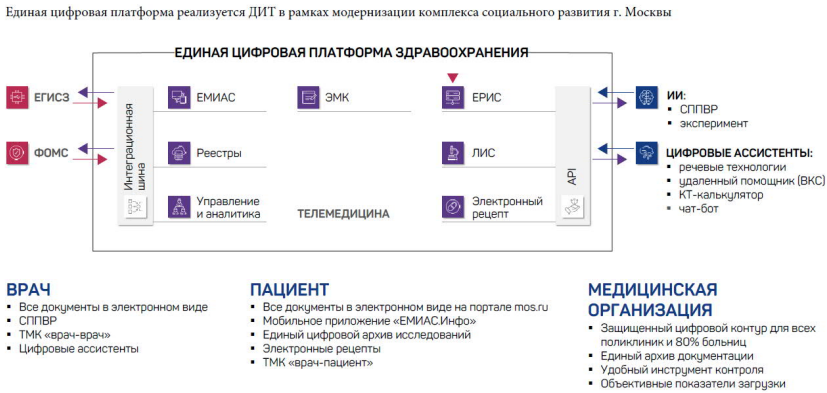

Рис. 9. Единая цифровая платформа здравоохранения Москвы

В эксперимент приходят разработчики технологий искусственного интеллекта со своими решениями. Мы сами ничего не разрабатываем. Мы обеспечиваем постановку задач. Это очень важно в медицине, критично важно – обоснованная постановка задачи. По определенным процедурам эти решения интегрируются в ЕРИС, о котором я сказал, работают, мы за ними наблюдаем, изучаем, оцениваем, мониторим, а Правительство Москвы платит компаниям гранты за количество проанализированных исследований. Исследование выполняется в медорганизации, попадает в ЕРИС, маршрутизируется на определенный сервис искусственного интеллекта, и врачу-рентгенологу в автоматизированное рабочее место попадает изначальная серия изображений (т.е. изображения, которые пришли с аппарата) и серия, проанализированная искусственным интеллектом. На этой серии есть маркировка, оконтуривание, подсвечивание обнаруженной патологии, а также текст, т.е. алгоритм готовит проект описания – документа, который врач должен в итоге выдать. Вот примеры работы. Ничего фантастического в этом на самом деле нет. Нет никакого робота в белом халате, всё выглядит достаточно прозаично. Иногда ждут робота с ИИ, а получается робот-пылесос. Но по большому счету робот-пылесос более полезен, чем какая-то нарисованная субстанция. Итак, идет просто разметка изображений, решение определенных задач. Надо сказать, что лучше всего эти технологии работают для измерений. Огромное количество измерений врач-рентгенолог должен выполнить фактически вручную. Да, это цифровое изображение, но он всё равно должен кликать мышкой, что-то находить, выделять и измерять. Это занимает колоссальное количество времени. Более того, человек субъективен и иногда может мерить неправильно. Когда есть хорошо разработанный сервис компьютерного зрения, то эти измерения выполняются за секунды и производительность растет, увеличивается в сотни раз. Соответственно, растет скорость подготовки документа и в итоге скорость предоставления медицинской помощи.

С учетом того, что эксперимент – это научное исследование, безусловно, есть специально разработанные методики оценки качества, безопасности, точности этих систем. Есть определенные, понятные математические метрики, но не всё так просто, поэтому огромное количество методов появилось еще в целях эксперимен-

та, и мы их делим на две большие группы: технологический мониторинг и клинический мониторинг (рис. 10).

Рис. 10. Методики контроля качества систем для клинической диагностики

Понятно, что мы оцениваем техническую и клиническую надежность сервисов, медицинскую точность, медицинское качество. Есть такой интегральный инструмент – «Матрица зрелости сервисов». Обратите внимание на ссылку: Mosmed.ia – это сайт эксперимента. Вся информация об участниках эксперимента, все показатели точности этих участников (больше 70 за все эти годы прошли через эксперимент), библиотека наборов данных, наши публикации на эту тему – всё есть на этом сайте. Кто захочет изучить поглубже – пожалуйста, есть такая возможность. В начале этого года наш центр посетил Президент Российской Федерации Владимир Владимирович Путин, знакомился с работой центра в целом и с экспериментом, в частности. По итогам этого визита мы получили сложное, интересное, очень ответственное поручение масштабировать эксперимент на всю страну. Сейчас создана платформа МосМед. ИИ, и любая медицинская организация страны может подключиться к этой платформе и работать с лучшими сервисами Московского эксперимента.

Сейчас уже обработано больше полумиллиона исследований по этому направлению. Кому будет интересно – вот несколько примеров наших изданий. Всё это бесплатно находится на нашем сайте, можно почитать, углубиться в проблематику. Спасибо за внимание!

Асеева И.А.: Коллеги, пожалуйста, вопросы. Пожалуйста, Елена Валерьевна.

Введенская Е.В.: Спасибо за очень интересный доклад. У меня несколько вопросов. Во-первых, про ChatGPT. Вы сказали, что это вообще не работает в медицине, но был случай осенью прошлого года, когда он поставил диагноз мальчику. До этого три года 17 врачей не могли поставить правильный диагноз. Мать немного времени пообщалась с чатом GPT, и он поставил редкий диагноз расщепления позвоночника. Потом это подтвердили нейрохирурги. Короче, мальчик сейчас пошел на поправку, его наконец вылечили, избавили от болей. Я понимаю, что это редкий случай, однако такое уже происходит. Как бы Вы могли это прокомментировать?

Владзимирский А.В.: Помните бессмертный роман Джерома К. Джерома «Трое в лодке, не считая собаки»? Там в первой же главе герой читает медицинскую энциклопедию и находит у себя все болезни. Чем-то мальчик точно болел, хотя бы с одной болезнью совпадение было. Поэтому так совпало – хорошо, повезло. Какие врачи его смотрели, почему не поставили диагноз? Болезнь, о которой Вы говорите, – довольно сложная врожденная патология. Может быть, врач, который смотрел, до нее не «докопался»? Нейрохирурги подтвердили? Если бы посмотрел сразу нейрохирург – сразу был бы диагноз. Посмотрел участковый терапевт – может быть, и не смог, не справился. Так тоже может быть.

Введенская Е.В.: То есть мы не можем никак обобщить этот случай?

Владзимирский А.В.: Мы живем в свободном мире, каждый может вместо врача попробовать задать промт. Хотели бы, чтобы Вас прооперировали после промта? Я точно не отвечу.

Введенская Е.В.: Понятно. Еще вопрос. Я сама, видимо, участвовала в этом эксперименте, не зная. Мне делали рентген в обычной московской поликлинике. Мне сказали, что через сутки рентген посмотрит врач. Но никаких суток не прошло. При мне – я видела на экране – сразу стал печататься текст, что все у меня в порядке, там патологии нет. То есть это явно не врач смотрел, это искусственный интеллект расшифровал. Это так сейчас работает?

Владзимирский А.В.: Это было профилактическое исследование? В ходе эксперимента мы нашли такой интересный момент.

Есть показатели чувствительности и специфичности – базовые показатели точности. Чувствительность – это то, насколько точно сервис говорит, что это исследование с болезнью. А специфичность – с какой болезнью. Получилось, что сервисы для грудной клетки гораздо лучше справляются с чувствительностью, т.е. они отсеивают норму со стопроцентной точностью. Если есть какая-то болезнь, они начинают «сомневаться». А то, что болезни нет – очень здорово. Здесь открылись потрясающие возможности для профилактических исследований. Действительно, флюорография, рентгенисследование органов грудной клетки – это очень важное исследование, оно массовое. Для понимания: 30% всех лучевых исследований в стране, выполняемых в год (каждое третье исследование), – это флюорография. Представьте себе масштабы! Это колоссальная нагрузка на систему здравоохранения. Но 99% из них – норма, потому что это массовое обследование здоровых людей. Какое количество кадровых ресурсов (врачей) из года в год, из жизни в жизнь описывает норму, вместо того чтобы заниматься сложными пациентами, приближать помощь там, где это нужно! Здесь появилась гипотеза, что мы можем настроить автономный искусственный интеллект. На самом деле это автоматизированная сортировка – мы сортируем исследования на «норму» и «не норму». «Не норма» идет к врачу на описание. Там могут быть возрастные изменения, может быть прооперированный пациент – все, что отклоняется от нормы. «Норма» к врачу не попадает. Но это пока только гипотеза. Мы ее тестируем поэтапно. У Вас в поликлинике действительно появился результат анализа искусственного интеллекта сразу, моментально. Это сортировка. Но это научное исследование, поэтому мы заботимся о безопасности. Для этого параллельно весь этот поток, который ИИ отправлял как «норму», пересматривался специальной федеральной медицинской организацией на предмет исключения ошибок, причинения вреда человеку. Гипотезу мы подтвердили: действительно, это работает. Но дальше, чтобы это внедрить в жизнь, требуется целый каскад изменений в законодательстве – это более сложно. Вас конкретно как пациентку московской поликлиники страховал рентгенолог, чтобы все было сделано по канонам медицинской науки и безопасно для пациентов.

Введенская Е.В.: Был очень интересный опыт, и я была в шоке, как будто будущее уже наступило. И последний вопрос: что

произошло с BotkinAI? Они тоже исследовали легкие. Мне попадалась несколько раз информация, что там нашли какие-то некорректные использования, и с этим проектом в итоге ничего не получилось. Не могу найти нигде достоверную информацию.

Владзимирский А.В.: Коллеги, у нас через эксперимент прошло больше 70 разных продуктов. BotkinAI один из них, они не единственные. На сайте Mosmed.ai есть результаты работы всех сервисов, и BotkinAI там тоже есть. Вы можете самостоятельно посмотреть. Мы стоим над рынком, т.е. мы никого не поддерживаем, но и никого не «убираем». Мы – объективный монитор происходящей ситуации, и эта же объективная информация предоставляется на сайте. Каждый может посмотреть, принять решение, увидеть то, что захочет в траектории развития каждого сервиса.

Степанов В.К.: Спасибо. Насколько реальна ситуация в Российской Федерации, что все лучевые исследования (КТ, МРТ, рентген) будут стекаться в единый центр, чтобы быть датасетом? Насколько это реально технологически и организационно? Насколько это возможно, насколько близко?

Владзимирский А.В.: Это возможно, но это совершенно нецелесообразно делать с учетом размера страны. Централизованные архивы медицинских изображений создаются в каждом субъекте. Каждый отдельный субъект имеет свое централизованное хранилище. ЕРИС – это централизованное хранилище города Москвы. В каждом субъекте, в любой республике, в каждой области создаётся свой ЦАМИ в своей системе здравоохранения.

Степанов В.К.: Более обширный датасет не даст лучший результат?

Владзимирский А.В.: Тут надо разделить: датасет для обучения искусственного интеллекта и хранилище медицинских изображений – это разные сущности. Хранилище решает вопросы, которые ставятся перед системой здравоохранения. Из хранилищ можно брать данные, формировать сколь угодно большие датасеты и обучать искусственный интеллект. Тут интереснее другое – не размер, а репрезентативность. Когда у нас есть ЦАМИ в каждом субъекте, то стекаются исследования из разных больниц, от разных популяций, от разных аппаратов, от разных стандартов сканирования. Если мы берем чуть-чуть из каждого хранилища и делаем большой датасет с такой репрезентативностью, у нас получается

алгоритм, который будет легко масштабироваться в любые условия применения. Стандартная методическая ошибка – это когда берется датасет из одной больницы. Научили на нем – получили прекрасную точность, перенесли в другую – все рухнуло. Благодаря системе ЦАМИ открываются интересные возможности по репрезентативности.

Степанов В.К.: Я не понял, есть ли единый датасет на страну по лучевым изображениям или он на уровне субъектов Федерации? Что такое единый датасет?

Владзимирский А.В.: Все сделанные в стране цифровые лучевые исследования накапливаются в централизованных архивах каждого субъекта. Через ЕГИСЗ – единую систему в сфере здравоохранения – референсным методом на федеральном уровне известно каждое исследование, каждая оказанная медицинская услуга. Нет централизованного физического хранилища, но есть референсная, зонтичная система, которая «знает» каждого пациента, каждую услугу. С федерального уровня можно «провалиться» (условно) до конкретного изображения – где оно лежит. Единая система есть, но нет физического единого хранилища. Хранилище распределенное. Пожалуйста, еще вопрос.

Асеева И.А.: Сначала Александр Александрович Самохин, а потом Дмитрий Александрович Карташов.

Самохин А.А.: У меня короткий вопрос. Во-первых, спасибо за очень четкое, принципиальное изложение позиций по поводу терминологии про искусственный интеллект. Насколько я понимаю, Ваша позиция совпадает с тем, что говорил в начале своего выступления первый докладчик? Спасибо.

Карташов Д.А.: Здравствуйте! Спасибо большое за доклад! Дмитрий Александрович Карташов, РГГУ. Вопрос такой: может ли Ваша система быть настроена на уровне разработчика, т.е. подобрана модель нейросети или локальные образцы, как Вы говорили, на которых можно обучить именно эту нейросеть (не общие, а точно по Вашим задачам)?

Владзимирский А.В.: Мы данные для обучения не предоставляем. Мы ставим клинические задачи, и дальше разработчики самостоятельно где-то изыскивают датасеты и инвестиции, обучают и к нам приходят уже с готовыми продуктами. У нас принципиальная позиция, что датасеты мы не предоставляем. Та библиотека

датасетов, о которой я сказал, во-первых, носит справочный характер. Там есть ограниченное количество датасетов в открытом доступе для тестирования – от 5 до 10–15 исследований. Есть, конечно, отдельные публикации, когда для медицины учат на 10 кейсах, но это совершенный абсурд. Это датасеты для тестирования. Дважды мы выкладывали большие датасеты, пригодные для обучения, в открытом доступе. Там 3000 КТ легких. В самом начале нашего пути тоже был такой эксперимент, мы тоже учились. Тот датасет был плохо размечен, но это мы сейчас уже понимаем. Второй раз мы выкладывали в COVID специальный датасет для поддержки разработчиков, но тому были серьезные причины. И в общем это очень сильно помогло продвинуться. Ковидный датасет тоже есть в открытом доступе, он большой – несколько тысяч исследований, даже десятки тысяч, если не ошибаюсь. Разработчики у нас в эксперименте разные: от крупных транснациональных вендоров до крошечных стартапов. Самый маленький стартап – два человека, они отлично работают, зарабатывают миллионы на грантах от Правительства Москвы. Так что, если есть желание, приходите в эксперимент!

Гребенщикова Е.Г.: В Вашем учреждении наверняка собрано множество клинических случаев; их можно использовать не только для обучения искусственного интеллекта, но и людей, специалистов?

Владзимирский А.В.: Вы совершенно правы: это отдельное направление нашей деятельности. Наш Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения Москвы – государственная медицинская организация, научный центр. У нас очень разнонаправленная деятельность. Искусственный интеллект очень весомый аспект, но это только одно из направлений. Мы в том числе ведем большую образовательную работу, у нас есть отдельный учебный центр. Поэтому мы делаем ставку в основном на естественный интеллект. А ИИ – это инструмент, который помогает естественному интеллекту чуть лучше справляться.

Асеева И.А.: Можно еще вопрос? Скажите, пожалуйста, как Вы думаете, в каких функциях этот помощник проявил бы себя наилучшим образом? Например, как менеджер, регулируя работу лечебного учреждения, или анализируя медицинские изображения,

как консультант, предлагая варианты диагноза? А как автоматизированный робот-хирург? Что Вы об этом думаете? Ведь, с одной стороны, мы понимаем, что он может сделать операцию точнее в каких-то сложных, тонких моментах. Но, с другой стороны, если человек с естественным интеллектом, хирург, никогда не будет этим заниматься, он же никогда и не научится?

Владзимирский А.В.: Да, коллеги, здесь тоже проявляются определенные стереотипы, а, может быть, и манипулирование терминологией. Никаких роботов-хирургов в принципе не существует! То, о чем говорится в прессе, называется «роботической хирургией». Но это не робот оперирует человека. Это просто сложная установка, которой тоже управляет человек – она является, скажем так, усилением пальцев и рук хирурга, большим протезом на его руках, который позволяет более тонко, очень по-особенному выполнять хирургическое вмешательство. Например, большая, красивая установка «Да Винчи» – их называют роботами. Это опять маркетинг. Такая большая штука стоит над пациентом на хирургическом столе, а врач где-то сидит в стороне или вообще за пультом. Вроде как хирург и не над пациентом находится. Но оперирует все равно врач. Эта большая, красивая машина сама без врача не сделает ничего. Это усиление хирурга. Поэтому о хирургии роботизированной, об искусственном интеллекте говорить не приходится совершенно. Незнаю, сколько еще десятилетий или столетий должно пройти, чтобы такая техника на самом деле появилась. Сейчас то, что называют искусственным интеллектом, с математической точки зрения решает очень узкие, очень конкретные задачи. Мы говорим: один алгоритм – одна патология. Самый простой пример: мы можем научить один алгоритм, одну математическую модель очень точно находить очаги в легком. Она нализирует легкое, находит очаг, но если рядом в легком будет, например, торчать копье, он не воспримет это как болезнь. Это же не очаг – значит, это норма. Этот алгоритм не разбирается ни в чем другом, кроме как находить очаг в легком. Мы понимаем, что до того разнообразия, до той сложности, которую представляет собой принятие решений в медицине, до того, что представляет собой клиническое мышление врача-человека, математике, наверное, не дойти никогда. Это чрезвычайно сложно и многогранно! Но помочь во множестве вещей действительно можно. В конце концов, мы ведь

тоже предпочитаем пользоваться машинами с автоматической коробкой передачи, а не с ручной. Так проще ездить. Человек остается водителем – в целом ничего не меняется.

Самохин А.А.: Есть такие строчки «Но все нам казалось, что мы – не такие, что мы не подвластны ни року, ни быту, что тайные карты нам веком открыты» из поэмы Наума Коржавина «По ком звонит колокол».

Асеева И.А.: Действительно нужно не только восторгаться открытием тайн, новыми технологиями, но и осознавать ответственность, последствия их применения, возможности, риски и ограничения. И развивать естественный интеллект! Спасибо большое, Антон Вячеславович, что Вы, несмотря на загруженность, выбрали время и приняли участие в нашем мероприятии. Спасибо всем за вопросы и комментарии!

Благодарим всех участников круглого стола!

Материал подготовила д. ф.н., проф., в. н.с.
центра научно-информационных исследований
по науке, образованию и технологиям ИНИОН РАН *И.А. Асеева*