ВЕСТНИК РОССИЙСКОЙ АКАДЕМИИ НАУК, 2024, том 94, № 7, с. 623–628

С КАФЕДРЫ ПРЕЗИДИУМА РАН

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ГУМАНИТАРНОЙ СФЕРЕ. УГРОЗЫ И ВОЗМОЖНОСТИ

© 2024 г. А.И. Аветисян[a],*

[a]*Институт системного программирования им. В.П. Иванникова РАН, Москва, Россия*

**E-mail: arut@ispras.ru*

Поступила в редакцию 03.07.2024 г.
После доработки 04.07.2024 г.
Принята к публикации 06.07.2024 г.

В статье рассматриваются проблемы, возникающие в связи с активным внедрением технологий искусственного интеллекта (ИИ) в гуманитарной сфере и в медицине, причины этих проблем, а также меры, которые предпринимаются в мире для их решения. Обсуждается создание методов и инструментов разработки безопасных технологий ИИ в рамках Исследовательского центра доверенного искусственного интеллекта Института системного программирования им. В.П. Иванникова РАН. Автор приводит результаты междисциплинарных проектов, реализуемых центром, а также предлагает ряд мер для активизации развития технологий ИИ, предназначенных для гуманитарной сферы.
Статья подготовлена на основе доклада, заслушанного на заседании президиума РАН 12 марта 2024 г.

Ключевые слова: доверенный искусственный интеллект, регулирование, репозиторий доверенного программного обеспечения, слабый искусственный интеллект, нейросети, междисциплинарность, машинное обучение, цифровая лингвистика, цифровая медицина.

DOI: 10.31857/S0869587324070028, **EDN:** FMXFED

В современном мире искусственный интеллект (ИИ) внедряется во многих областях человеческой деятельности, потому что он существенно улучшает качество услуг, повышает производительность труда и, как следствие, положительно влияет на экономические показатели многих отраслей. Однако наряду с открывающимися возможностями широкое применение ИИ чревато и серьёзными угрозами. Чтобы разобраться, каковы они и как им противостоять, необходимо определить, что такое искусственный интеллект.

АВЕТИСЯН Арутюн Ишханович – академик РАН, заместитель президента РАН, директор ИСП РАН.

Как отмечено в обновлённой “Национальной стратегии развития искусственного интеллекта на период до 2030 года” [1], ИИ – это “комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их”. Этот комплекс включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение, процессы и сервисы обработки данных и поиска решений.

Современный ИИ основан на алгоритмах машинного обучения, а его активное развитие в наше время вызвано, в частности, ростом мощности вычислительных ресурсов (рис. 1) [2]. Кроме того, важную роль играет резкий рост объёма структурированных и неструктурированных данных, наблюдаемый в мире с 2010-х годов.

Наличие суперкомпьютерных мощностей и большого массива данных позволяет реализовывать проекты класса “мегасайенс” [3], актуальные для многих научных областей, в том числе гуманитарной

направленности. Например, в психологии можно анализировать в автоматическом режиме миллионы аккаунтов пользователей социальных сетей, вести исследования, касающиеся детектирования эмоций с помощью видео-, аудио- и текстовых материалов.

Необходимо отметить, что современный ИИ представляет собой технологию без субъектности. Это так называемый “слабый искусственный интеллект” (weak AI) [4], который базируется на методах машинного обучения. Слабый ИИ уязвим для предвзятостей и ошибок. Он извлекает информацию из ограниченного набора данных и решает только те задачи, на которые он запрограммирован. Если данные искажены, слабый ИИ может выдавать необъективный (неэтичный, дискриминационный) результат.

Сильный ИИ (strong AI) гипотетически будет способен делать интеллектуальные выводы, решать задачи на уровне умственных возможностей человека, использовать стратегии, планировать действия, функционировать в условиях неопределённости. Он сможет общаться на естественном языке и мыслить абстрактно. Но в настоящее время технологий сильного ИИ не существует, так как отсутствуют методы, на которых они могли бы базироваться. Этот факт признан специалистами во всём мире. Показателен такой пример. В 2022 г. старший инженер-программист холдинговой компании Alphabet, управляющей компанией Google Inc. и её дочерними структурами, Б. Лемуан самоуверенно заявил, что языковой чатбот LaMDA обладает собственным разумом, после чего был уволен за некомпетентность.

Несмотря на очевидные недостатки слабого ИИ, он находит всё более широкое применение в различных отраслях. Если говорить о гуманитарной сфере, можно привести в пример такое направление, как цифровая лингвистика, где ИИ помогает решать задачи создания систем машинного перевода, сервисов автоматической транскрибации (перевода устной речи в письменную), чат-ботов и голосовых помощников в смартфонах. По некоторым данным, число коммерческих программ машинного перевода в мире с 2017 по 2022 г. выросло почти в 5 раз [5]. Растёт рынок технологий распознавания голоса, необходимых для успешного функционирования и развития умных голосовых устройств. Так, в голосовой помощник Amazon Alexa уже добавлен генеративный ИИ.

Ещё одна в определённой степени близкая к гуманитарной сфере наука, в которой технологии искусственного интеллекта активно внедряются, это медицина. Они помогают решать задачи сбора, анализа и хранения больших объёмов медицинских сведений: изображений (рентгенография, компьютерная томография, электрокардиография, гистология), данных об эпидемиологических трендах, генетических исследованиях. ИИ внедряется в процессы испытания лекарств, а также разработку технологий так называемого Emotion AI (распознавание эмоций для анализа ментального здоровья). Современный ключевой тренд – предиктивная персонализированная медицина, направленная на сохранение здоровья и предупреждение заболеваний. С этим трендом ассоциировано распространение носимых электронных девайсов (фитнес-браслеты, часы с кардиографом и т.д.), с ним же связана концепция цифровой экосистемы “домашний госпиталь”, разработанная в США с целью снижения нагрузки на медицинские учреждения.

Фактически благодаря развитию и широкому внедрению ИИ в гуманитарной сфере появляются новые отрасли экономики. Однако у этого процесса есть и обратная сторона. Она проявляется в угрозах, связанных с несовершенством слабого ИИ. В числе этих угроз отметим следующие:

• нарушение конфиденциальности и приватности данных (ИИ-системам часто требуются большие

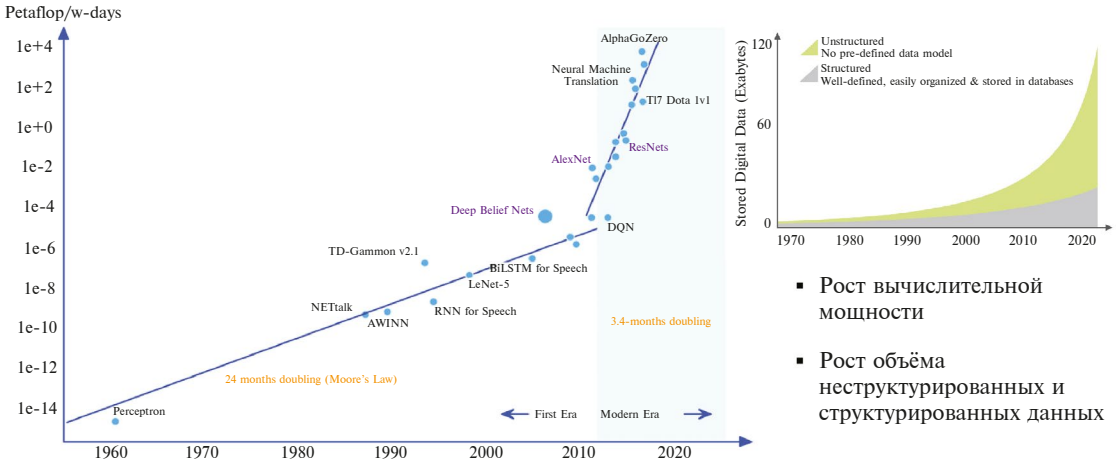

Рис. 1. Вычислительные ресурсы и большие данные – двигатели развития систем искусственного интеллекта

Источники: [2], https://dev.to/elastic/introduction-to-artificial-intelligence-and-data-analytics-47b5

объёмы информации для обучения и функционирования, в гуманитарной сфере это могут быть очень чувствительные данные, такие как медицинские записи);

• формирование и усиление неравенства (ИИ может обострить социально-экономическое неравенство, если его применение будет доступно только части населения или если результаты его работы продемонстрируют систематическую предвзятость);

• зависимость от технологий и потеря человеческого контакта (в сферах, где важно сохранение человеческого участия и эмпатии, таких как социальная или психологическая помощь, чрезмерная автоматизация может привести к снижению качества услуг);

• манипуляция сознанием и пропаганда (ИИ может использоваться в целях манипулирования общественным мнением, распространения дезинформации и усиления пропагандистского воздействия на население);

• проблемы ответственности (определение персональной ответственности за ошибки или вред, причинённый действиями ИИ, может оказаться очень сложным, особенно когда речь идёт о комплексных системах с автономными функциями);

• этические дилеммы (в гуманитарной сфере ИИ может столкнуться с этическими дилеммами, такими как выбор между разными видами помощи или распределение ограниченных ресурсов).

Число примеров реализации таких угроз стремительно растёт. Например, в феврале 2024 г. в Гонконге началось расследование не встречавшегося ранее проявления мошенничества: преступники обманом вынудили сотрудника транснациональной корпорации перевести им 25.6 млн долл., организовав для него фейковую видеоконференцию с участием якобы директора компании и других коллег (все фантомные собеседники были сгенерированы с помощью ИИ). Обман он обнаружил только после обращения в головной офис корпорации [6].

Пример из юридической области. В 2023 г. гражданин США подал в суд на авиакомпанию Avianca. Он заявил, что получил травму из-за того, что металлическая сервировочная тележка ударила его по колену во время полёта. Адвокаты представили суду записку на десяти страницах, где перечислялось несколько соответствующих судебных решений (в США действует прецедентное право). Но судья не смог найти документы, подтверждающие эти решения, потому что записку составил ChatGPT (чат-бот с генеративным искусственным интеллектом). Проведённое позднее исследование учёных Стэнфордского университета показало, что в 69–88% случаев искусственный интеллект неверно отвечает на вопросы, связанные с судебной практикой. Например, ChatGPT версии 3.5, разработанный компанией OpenAI, допустил ошибки в 69% случаев. Исследователи пришли к выводу, что пока искусственный интеллект не способен проводить юридический анализ ситуации [7].

Ещё один пример, более общий. ИИ может активно использоваться в целях пропаганды. Например, он способен в кратчайшие сроки проанализировать существующие тренды, подготовить массив фейковых новостей и распространить их по социальным медиа. Всё это осуществлялось и до появления технологий ИИ, так как границы между естественным развитием событий и спланированной акцией уже давно размыты. Достаточно упомянуть движение “жёлтых жилетов” во Франции или выборы президента США в 2016 г. Однако с привлечением технологий ИИ проводить пропагандистские кампании, психологически воздействуя на большие социальные группы, стало значительно проще и дешевле.

Помимо гуманитарных угроз существуют и техногенные. Тема статьи не предполагает их подробного рассмотрения, поэтому стоит привести лишь один пример – ДТП с участием беспилотных автомобилей. В октябре 2023 г. в США обычный автомобиль с водителем за рулём сбил пешехода. Пешехода отбросило под колёса беспилотного автомобиля Cruise производства компании General Motors, который остановился, но затем возобновил движение. Пострадавший оказался зажат под колесом и получил серьёзные травмы, так как Cruise проехал ещё 6 м. Спустя месяц 950 машин Cruise были отозваны для устранения недостатков программного обеспечения. По заверениям компании General Motors, в дальнейшем беспилотные Cruise будут полностью останавливаться в аналогичной ситуации [8].

Внедрять технологии искусственного интеллекта, безусловно, следует, это даёт значимый экономический эффект, однако области его применения должны быть контролируемыми. Нужны регуляторные механизмы, чётко определяющие, как и в каких условиях допустимо использовать интеллектуальные системы, а также границы их безопасности. Разработать и регуляторику, и безопасные ИИ-системы можно только с привлечением отраслевых специалистов, а для этого необходимы междисциплинарные проекты.

Потребность в регуляторных документах в области ИИ признают во многих странах мира. Например, в 2024 г. был одобрен EU AI Act (“Закон Евросоюза об искусственном интеллекте”) [9], в положениях которого предлагается разделить все системы с ИИ на следующие категории.

1. Системы с минимальными рисками (например, игры или спам-фильтры). Регулирование не требуется.

2. Системы с ограниченными рисками (например, системы генерации контента – изображений, аудио или видео). В этом случае применяется регулирование. В частности, контент должен быть обязательно помечен как сгенерированный.

3. Системы с высокими рисками (системы управления критической инфраструктурой, беспилотные автомобили, медицинские устройства с ИИ и др.). Необходимо жёсткое регулирование.

4. Системы с неприемлемыми рисками (системы социального скоринга[1], распознавания лиц в режиме реального времени и др.). Такие технологии будут под запретом (за редким исключением).

Что касается систем, обозначенных в пунктах 2 и 3, то для создания регуляторных механизмов достаточно технологического подхода, то есть разработки соответствующих методик и инструментов. В случаях появления систем, обозначенных в пункте 4, необходим ещё и гуманитарный подход, так как в каждом отдельно взятом обществе – своё понимание недопустимого. Это в очередной раз доказывает, что без междисциплинарного подхода создание регуляторики ИИ невозможно.

Разработка аналогичных документов ведётся и в США. В 2022 г. был опубликован проект "Билля о правах" ИИ [10], предлагаемого компаниями, общественными организациями и экспертными группами. В нём формулируются пять принципов создания и использования искусственного интеллекта, в числе которых разработка безопасных и эффективных систем, отсутствие алгоритмической дискриминации, обеспечение конфиденциальности данных. В 2023 г. в США на государственном уровне был одобрен ещё один важный документ – Executive Order on Safe, Secure, and Trustworthy AI ("Указ о безопасном, защищённом и доверенном искусственном интеллекте") [11], который устанавливает новые стандарты в сфере безопасного развития ИИ и содержит поручения ведомствам и разработчикам. Например, разработчики ряда значимых систем обязаны делиться с правительством результатами тестов на безопасность продуктов; кроме того, сгенерированный ИИ контент должен маркироваться специальными цифровыми метками. Последнюю инициативу разделяют и ведущие компании в области ИИ (OpenAI, Meta.Platforms, Alphabet и др.), которые уже обязались реализовать систему цифровых водяных знаков для всех форм синтезированного контента. В том же году Агентство национальной безопасности США объявило о создании Центра безопасности искусственного интеллекта, а Национальный научный фонд – о создании семи исследовательских институтов ИИ (один из них – Institute for Trustworthy AI in Law & Society, Институт доверенного искусственного интеллекта в юридических и общественных науках).

Если же говорить об общемировых подходах, то 19 мая 2023 г. на саммите глав государств "Большой семёрки" в Хиросиме был принят специальный документ для содействия развитию передовых систем искусственного интеллекта на глобальном уровне. 30 октября 2023 г. те же лидеры поддержали "Международный кодекс поведения" и "Руководящие принципы для организаций, разрабатывающих передовые системы ИИ" [12]. Например, в кодексе заявлено, что таким организациям следует "присоединиться к процессам разработки, продвижения и принятия, где это необходимо, общих стандартов, инструментов, механизмов и лучших практик для обеспечения безопасности, надёжности и достоверности передовых систем ИИ". А разработчики должны "добиваться полной прозрачности – документировать используемые наборы данных, процессы и решения, принятые в ходе разработки системы".

Разработка регуляторных механизмов ИИ активно ведётся и в нашей стране. Россия здесь – в числе лидеров. В 2019 г. Указом Президента Российской Федерации № 490 была принята Национальная стратегия развития искусственного интеллекта на период до 2030 года (обновлена в 2024 г.). В 2021 г. десятки компаний подписали Кодекс этики в сфере искусственного интеллекта [13], разработанный при участии Минэкономразвития России, Аналитического центра при Правительстве РФ, а также около 500 экспертов академического и бизнес-сообщества. Кодекс подчёркивает приоритет прав человека, ответственность человека за действия ИИ, потребность в безопасности и защищённости данных, а также необходимость разработки безопасных технологий. Появляются первые ГОСТы, например ГОСТ 59921 "Системы ИИ в клинической медицине" (принят в 2022 г.; устанавливает требования к клиническому тестированию ИИ-систем на основе глубоких нейронных сетей) [14].

Таким образом, в мире предпринимаются попытки ограничить внедрение небезопасных технологий и установить стандарты разработки безопасных, как это было ранее сделано для обычного программного обеспечения. Однако на практике общепринятых подходов и инструментов разработки безопасных технологий искусственного интеллекта пока не существует.

Центры по разработке таких технологий только начинают появляться в разных странах мира, в том числе и в нашей стране. В 2021 г. при поддержке Минэкономразвития России в Институте системного программирования им. В.П. Иванникова РАН учреждён Исследовательский центр доверенного искусственного интеллекта, цель которого – создать платформу доверенного ИИ. Платформа должна объединять программные инструменты и методики для противодействия принципиально новым угрозам, возникающим на всех этапах жизненного цикла технологий ИИ. Здесь получены результаты по ряду актуальных направлений, и что очень важно, эти результаты уже внедряются индустриальными партнёрами центра. Например, дове-

[1] Социальный скоринг – вид оценки платёжеспособности заёмщика банка по его социальным характеристикам и прогнозирования поведения клиента с помощью анализа его присутствия в социальных сетях.

ренные версии фреймворков TensorFlow и PyTorch внедрены в “Kaspersky Machine Learning for Anomaly Detection” v. 3.0 (“Лаборатория Касперского”).

Ряд проектов Центра реализуют междисциплинарные команды, когда лучшие практики создания ИИ-решений сопрягаются с прикладными исследованиями в различных областях, в том числе в гуманитарной сфере. Такой подход соответствует принципам “Национальной стратегии развития искусственного интеллекта на период до 2030 года”, где отмечается, что “для развития фундаментальных и прикладных научных исследований в области ИИ необходима в том числе реализация междисциплинарных исследовательских проектов в различных отраслях экономики”. В числе междисциплинарных проектов упомянем следующие.

Анализ ЭКГ (проект реализуется в рамках Научного центра мирового уровня совместно с Первым Московским государственным медицинским университетом им. И.М. Сеченова и другими заинтересованными организациями). Разработан макет системы разметки 12-канальных ЭКГ (http://ecg1.ispras.ru). Стандартизированная разметка по списку патологий помогает достичь высокой степени согласия между экспертами. Идёт работа по развитию и внедрению нейросетевой модели классификации 12-канальных ЭКГ. Модель проходит регистрацию в качестве медицинского изделия (уже интегрирована в систему “Единый кардиолог” Республики Татарстан). О проекте будет проинформирована и зарубежная научная общественность. Журналом “Biomedical Signal Processing and Control” (Q1) принята к публикации статья “Deep neural networks generalization and fine-tuning for 12-lead ECG classification” (“Дообучение и обобщаемость глубоких нейронных сетей для классификации ЭКГ в 12 отведениях”) [15].

Анализ соцсетей (совместно с Институтом психологии РАН). С помощью передовых методов ИИ проведено исследование более 750 тыс. постов социальной сети “ВКонтакте” для понимания связи между поведением в соцсетях и психологическими особенностями пользователей. В журнале “Scientific Reports” (Q1) опубликована статья “Applying explainable artificial intelligence methods to models for diagnosing personal traits and cognitive abilities by social network data” (“Применение интерпретируемых методов искусственного интеллекта к моделям диагностики личностных качеств и когнитивных способностей по данным социальных сетей”) [16].

Распознавание эмоций (совместно с Институтом психологии РАН). С 2023 г. развиваются два направления: детектирование ментальных проблем (посттравматического стрессового расстройства) и детектирование стресса и усталости у человека. Совместно с сотрудником Института психологии РАН С.Г. Мелик-Карамян ведётся совместная работа по составлению наборов данных для исследований в области автоматического распознавания эмоций с помощью нейросетей. Одна из актуальных задач – объединить данные нескольких типов: ЭКГ и детектирование эмоций (видео, аудио, тексты). Задача пока решается отдельно на открытых датасетах.

* * *

Подводя итоги, хотелось бы ещё раз подчеркнуть, что долгосрочное устойчивое развитие технологий ИИ в гуманитарной сфере требует комплексного подхода с привлечением как специалистов по информационным технологиям, так и представителей гуманитарных областей знания. Для активизации исследований целесообразны следующие меры.

Сформировать под руководством РАН постоянно действующую рабочую группу по направлению “Искусственный интеллект в гуманитарной сфере”.

Инициировать создание репозитория доверенных решений ИИ в сфере гуманитарных наук. Он будет содержать наборы данных, предобученные модели и средства поддержки полного жизненного цикла технологий ИИ, в том числе обеспечивающие возможность совместной работы распределённых междисциплинарных групп исследователей и гарантирующие безбарьерный, равноправный доступ к аппаратным средствам.

Инициировать запуск целевых комплексных научно-технологических проектов, а также постоянной программы поддержки научных групп. В программу будут вовлекаться (в том числе по грантам) исследователи из дружественных стран для реализации собственных междисциплинарных проектов с использованием технологий ИИ в гуманитарной сфере.

Проработать вопрос подготовки и реализации образовательных междисциплинарных программ (бакалавриат, магистратура, аспирантура), направленных на развитие компетенций по использованию технологий ИИ в гуманитарной сфере.

Проработать создание регуляторных механизмов, закрепляющих возможность беспрепятственного сбора и использования открытых данных. Данные, доступные в сети Интернет, позволят создавать наборы для научных исследований, в том числе междисциплинарных.

ЛИТЕРАТУРА

1. Указ Президента Российской Федерации “О развитии искусственного интеллекта в Российской Федерации”. Национальная стратегия развития искусственного интеллекта на период до 2030 года. http://static.kremlin.ru/media/events/files/ru/AH4x6HgKWANwVtMOfPDhcbRpvd1HCCsv.pdf

   Decree of the President of Russia “On developing artificial intelligence in Russia”, National strategy of developing artificial intelligence for the period until

2030. http://static.kremlin.ru/media/events/files/ru/AH4x6HgKWANwVtMOfPDhcbRpvd1HCCsv.pdf

2. *Zhang Yu., Nauman U.* Deep Learning Trends Driven by Temes: A Philosophical Perspective. https://www.researchgate.net/figure/Two-distinct-eras-of-computation-usage-in-training-AI-systemscredit-OpenAI-48_fig3_346359064.
3. Шагнуть за горизонт: что такое установки мегасайенс. https://xn--80aapampemcchfmo7a3c9ehj.xn--p1ai/news/shagnut-za-gorizont-chto-takoe-ustanovki-megasayens/

   Step beyond the horizon: what are megascience installations? https://xn--80aapampemcchfmo7a3c9ehj.xn--p1ai/news/shagnut-za-gorizont-chto-takoe-ustanovki-megasayens/
4. *Searle J.R.* Is the Brain's Mind a Computer Program? // Scientific American. 1990, vol. 262 (1), pp. 26–31.
5. Number of commercial machine translation (MT) programs available globally from 2017 to 2022. https://www.statista.com/statistics/1378793/mt-translation-programs-number/
6. Finance worker pays out $25 million after video call with deepfake 'chief financial officer'. https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html
7. Hallucinating Law: Legal Mistakes with Large Language Models are Pervasive. https://hai.stanford.edu/news/hallucinating-law-legal-mistakes-large-language-models-are-pervasive
8. Cruise recalls all self-driving cars after grisly accident and California ban. https://www.theguardian.com/technology/2023/nov/08/cruise-recall-self-driving-cars-gm
9. EU AI Act. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf
10. Blueprint for an AI Bill of Rights. https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf
11. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/
12. Hiroshima Process International Code of Conduct for Advanced AI Systems. https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems
13. Кодекс этики в сфере искусственного интеллекта. Ethics https://ethics.a-ai.ru/assets/ethics_files/2023/05/12/Кодекс_этики_20_10_1.pdf

    Codex on Artificial Intelligence. https://ethics.a-ai.ru/assets/ethics_files/2023/05/12/Кодекс_этики_20_10_1.pdf
14. ГОСТ Р 59921.2-2021 "Системы искусственного интеллекта в клинической медицине". https://docs.cntd.ru/document/1200181991

    State standard R 59921.2-2021 "Artificial Intelligence Systems in Clinical Medicine. Part 2. Program and methodology of technical validation". https://docs.cntd.ru/document/1200181991
15. *Avetisyan A., Tigranyan Sh., Asatryan A. et al.* Deep Neural Networks Generalization and Fine-Tuning for 12-lead ECG Classification. https://arxiv.org/abs/2305.18592
16. *Panfilova A., Turdakov D.* Applying explainable artificial intelligence methods to models for diagnosing personal traits and cognitive abilities by social network data. https://www.nature.com/articles/s41598-024-56080-8

ARTIFICIAL INTELLIGENCE IN THE HUMANITARIAN FIELD. THREATS AND OPPORTUNITIES

A.I. Avetisyan[a],*

[a] *Ivannikov Institute for System Programming of the Russian Academy of Sciences, Moscow, Russia*
**E-mail: arut@ispras.ru*

In this paper we discuss the problems arising out of active introduction of artificial intelligence (AI) technologies to medicine and other humanities, as well as causes of these problems and steps that are taken worldwide to solve them. We focus on the methods and tools for developing trusted AI technologies, which are being created within the ISP RAS Trusted AI Research Center. We present the results of interdisciplinary projects executed in the Center and suggest a number of solutions to speed up the development of humanitarian AI technologies. The paper expands on the report given at the General Assembly of the RAS on March 12, 2024.

Keywords: trusted artificial intelligence, regulation, trusted software repository, weak artificial intelligence, neural networks, federative learning, machine learning, digital linguistics, digital medicine.