ДОКЛАДЫ РОССИЙСКОЙ АКАДЕМИИ НАУК. МАТЕМАТИКА, ИНФОРМАТИКА, ПРОЦЕССЫ УПРАВЛЕНИЯ, 2023, том 514, № 2, с. 6–19

УДК 004.5, 004.8

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ОБЩЕСТВЕ

© 2023 г. Академик РАН А. Л. Семенов[1,2,3,*]

Поступило 30.10.2023 г.
После доработки 30.10.2023 г.
Принято к публикации 01.11.2023 г.

Статья представляет собой авторский обзор сингулярности, в которой развиваются события в области ИИ. Предлагается общий взгляд на роль информационно-технологических революций как расширяющих человеческую личность. Рассматривается современный этап расширения личности, охватывающий последнее десятилетие и особенно 2023 год. Рассматриваются наиболее важные и типичные общественно значимые документы, в которых выражается озабоченность результатами развития и применения ИИ, а также документы, в которых декларируются условия позитивного развития событий: этические принципы, приоритетные направления, требования и ограничения. Более детально рассматривается находящийся в стадии принятия Европейский закон об ИИ и реакция разных групп на него. Выделяются культурно-исторические факторы, которые могут противостоять негативному и катастрофическому развитию событий, которое становится возможным благодаря ИИ. Анализируются возможные механизмы сохранения подлинного знания в среде профессионалов и распространения его среди широких слоев населения.

Ключевые слова: искусственный интеллект, расширенная личность, озабоченность негативным влиянием ИИ, межгосударственное регулирование, Европейский закон об ИИ, Хиросимский процесс ИИ, Ковчег знаний, Энциклопедия

DOI: 10.31857/S2686954323350023, **EDN:** НКОНJC

1. ВВЕДЕНИЕ

Говоря об ускорении событий в мире, в области искусственного интеллекта (ИИ) – во многом ответственной за это ускорение, мы где-то внутри ощущаем разговоры об этом ускорении как метафорические, журналистские. Однако действительность каждый раз ставит наше подсознание на место. Оглядываясь за время, прошедшее от подготовки предыдущего “Путешествия в ИИ” (меньше года), мы чувствуем себя путешественниками в будущее: не могло все это произойти за один год!

Работая над настоящей статьей и просматривая новостные ленты, я вижу необходимость ежедневно вносить в текст изменения в связи с происходящими событиями.

Пару дней назад наш научно-методический совет по школьным математическим экзаменам принял решение, которого я добивался десятилетиями: разрешить школьникам на экзамене по математике в 9-м классе использовать непрограммируемый калькулятор. После рекомендации об использовании калькулятора на занятиях, изданной министерством образования СССР в 1982 году [1], и Постановления ЦК КПСС и Совмина СССР по цифровой трансформации школы 1985 года [2] это важнейший шаг в сокращении постоянно растущего отставания школы от жизни.

Вчера представители Большой семерки подписали Кодекс ИИ, свою подпись к нему добавила Президент Европейской комиссии. Сегодня в английском поместье, где Алан Тьюринг создавал свой компьютер для расшифровки немецкого военного шифра, открывается первый в истории Саммит по рискам ИИ.

Пытаясь удержать в своем быстро расширяющемся сознании происходящие изменения, мы пытаемся, с одной стороны, восстанавливать хронологию недавних событий, с другой – вписывать их в глобальную временну́ю (логарифмическую – экспоненциальную, а может быть – гиперболическую?) шкалу ускоряющихся изменений, начинающуюся с другой – аттосекундной (гиперболической, обращенной в другую сторону) космогони-

[1]*Российский государственный педагогический университет им. А. И. Герцена, Санкт-Петербург, Россия*
[2]*Институт образования НИУ Высшая школа экономики, Москва, Россия*
[3]*Московский государственный университет имени М.В. Ломоносова, Москва, Россия*
*****E-mail: alsemno@ya.ru*

ческой шкалы. Приближаясь к предсказываемому некоторыми лидерами ИИ концу человеческой истории, мы вглядываемся и в эту историю, и в текущие события.

2. ПОНЯТИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Обсуждение общей проблематики ИИ и конкретных применений технологий часто включает методологические элементы или даже вообще сводится к вопросам типа: это ИИ или не ИИ? Связано это, видимо, в том числе и с беспрецедентной скоростью развития событий в данной области.

Мы предпочитаем использовать технически самое простое определение: искусственный интеллект (ИИ) – это средства автоматизации интеллектуальной деятельности человека. По существу – это переформулировка первоначального, почти тавтологичного, определения МакКарти [3]. Часто ИИ – это элемент деятельности, которую ведет человек, например, редактируя текст. Но во многих случаях действия ИИ не встроены в деятельность человека непосредственно. ИИ оказывается все более и более автономным, как в случае автономных транспортных средств. Это порождает естественные опасения, которые рассматриваются дальше.

Теоретической же основой как ИИ, так и компьютерных технологий можно считать открытие математиками возможности формально описать математическую деятельность человека внутри самой математики: с помощью машин Тьюринга, рекурсивных функций, формул логики отношений или лямбда-исчисления и т.п. Это открытие было завершено в 1930-е гг.

3. РЕВОЛЮЦИЯ ИИ В РЯДУ ДРУГИХ ИНФОРМАЦИОННЫХ РЕВОЛЮЦИЙ В ИСТОРИИ ЧЕЛОВЕЧЕСТВА

Как говорил Л. С. Выготский [4, 5] и замечали многие другие, появление принципиально новых информационных технологий приводило к радикальному изменению того, как человек думает, общается, действует, учится, причем не только в условиях явного использования новой технологии. В качестве исходного момента в серии таких радикальных изменений – революций – естественно рассмотреть само появление человека разумного *Homo sapiens* несколько сотен тысяч лет назад. За этим последовало (указываем порядки в годах при обратном счете):

появление речи – 10^5;

появление письменности – 10^3;

появление искусственного интеллекта – 10^2.

Последняя революция также вызвала очередное революционное лишение человека его уникальности, центрального положения в мире, как и революции Коперника, Дарвина и Фрейда (см. [6]). Человек перестал быть единственным и исключительным носителем рассудка.

Вероятной следующей информационной революцией станет распространение прямого человеко-машинного взаимодействия, стирание информационной границы между мозгом и цифровой средой, когда, например, будет требоваться специальная дисциплина и усилия, чтобы сообразить, “где” ты помнишь тот или иной факт. Распространенным примером прямой передачи электронных импульсов в нервную систему, уже охватывающим миллионы людей, в том числе лауреатку Нобелевской премии мира Малалу Юсуфзай, является кохлеарный имплант [7]. Имплант осуществляет прямую передачу электрических образов звука на нервные окончания. Развиваются и другие технологии взаимодействия нервной системы с электронной средой. Предшествующим этапом такого будущего взаимодействия является сегодня смартфон или планшет, воспринимающий устный текст, подключенный к интернету, беспроводным наушникам и т.д. Для наших детей, пугающим нас образом, он становится основным окном в мир. Как известно, не меньшим шоком для цивилизации был переход знания из устной формы в письменную (см. [8], где Сократ предсказывает исчезновение знания в результате появления письменности).

С абстрактно функциональной точки зрения смартфон мало отличается от планшета или ноутбука. Однако его доступность для человека – “карманность” – принципиально меняет взаимодействие человека с цифровой средой. В этой сфере мы встречаемся с разнообразными неожиданностями, например, с тем, как на протяжении нескольких лет появился (а потом – исчез) новый вид взаимодействия: набор (прежде всего – детьми) текстов на клавиатуре мобильника двумя большими пальцами; средством коммуникации стали короткие видео или селфи.

Однако ситуация лишь в малой степени сводится к “аппаратной”.

Сегодня человек повсеместно и ежесекундно взаимодействует с цифровым информационным пространством, выход из него происходит не чаще, чем в предшествующие столетия вход в информационное пространство книг, писем, часов, карт и счетов. Ведущим каналом восприятия информации (органом чувства) является зрение, ввода – движения пальцев, доля звукового, устного канала растет. При этом идет интеграция устного и письменного текста, текста и изображения: “рассказ компьютера по картинке”, автоматическое аннотирование видео, синтез изображе-

ния по описанию. Одновременно наращивается и разнообразится текстовый диалог между человеком и окружающей цифровой средой, порождающей тексты по человеческим запросам и направляющей к нему разнообразные месседжи, влияющие на него (прямой диалог, например, с банковской системой или муниципальными службами все еще часто оказывается неизбежным мучением для человека).

Перечисленные выше глобальные революции сопровождали революции второго порядка (субреволюции), такие как появление часов, печатных книг (революция Гутенберга) или фотографии. Революция средств электронной передачи и хранения информации, начавшаяся еще до цифровой компьютерной революции, интегрировалась с другими цифровыми технологиями. Среди этих субреволюций: телевидение на фоне фотографии, кино и электрическая передача сообщений, интернет на фоне почтовой связи и электронной передачи сообщений. В области ИИ субреволюции соответствуют, грубо говоря, видам ИИ, созревание которых также занимает сокращающиеся интервалы времени:

– *рациональный ИИ,*

– *интуитивный ИИ,*

– *творческий ИИ.*

Сегодня, видимо, наступает пора *деятельного ИИ*. Человек все больше будет прямо или косвенно поручать ИИ действия. Тест Тьюринга 2023 года предлагает ИИ заработать 1 млн долларов, вложив 100 тыс. долларов [9].

4. РАСШИРЕННАЯ ЛИЧНОСТЬ

Как указал Л. С. Выготский, личность человека, его мышление, поведение зависят от имеющихся в его распоряжении информационных средств (источников, инструментов), информационной среды, в которой он существует. И для себя, и для других человек – это расширенная личность. В его распоряжении есть источник точного времени, и от него ожидают, что он придет вовремя, а если случится что-то непредвиденное, то он воспользуется сотовым телефоном. Он знает сотни номеров телефонов и держит это в своей памяти, только не во внутренней, и не в бумажной записной книжке, а во внешней – в облаке. От человека ожидают, что он позвонит, как обещал. Он сможет грамотно написать письмо и послать его по электронной почте и т.п.

Вымышленный диалог Эдисона с Эйнштейном [10] дефакто решается в пользу Эйнштейна. Квалификация человека состоит не в том, что он хранит ответ в своей черепной коробке и может его там найти, а в том, что он может правильно задать вопрос себе (в том числе – внешнему, а значит – всему человечеству) и отфильтровать среди ответов тот, который ему действительно нужен. Именно расширенная личность сегодня – субъект социальных отношений: моральных, экономических, юридических, образовательных. (Упомянутое разрешение калькулятора на экзамене в 9-м классе – шаг к признанию этого.)

Расширенная личность имеет подвижную границу с окружающей средой, а среда эта в первую очередь цифровая. Эта среда постоянно воздействует на нас информацией и постоянно собирает информацию о нас, в том числе – через видеокамеры, установленные везде в городах.

5. ВЫЗОВЫ, ОПАСНОСТИ. ПРЕДОСТЕРЕЖЕНИЯ

Существенное влияние как на формирование общественного мнения, так и на позицию правительств, корпораций, лиц, принимающих решения, имеют время от времени возникающие документы, к которым причастны лидеры в области ИИ, а также другие влиятельные ученые и политики.

Беспрецедентным является то, что эти документы не только являются попытками привлечь внимание к какой-то научно-технологической области в целях рекламы или конкурентной борьбы, но в первую очередь в этих документах выражается озабоченность проблемами, которые могут возникнуть в связи с использованием ИИ. Масштабы такого предостережения превосходят предостережения создателей ядерного оружия, которое при этом проектировалось именно как оружие. Самое существенное – инициаторами предостережений являются не “внешние” алармисты – “зеленые”, борцы за мир и т.п., а лидеры самой отрасли.

Ряд действий в направлении поддержки коллективного создания указанных документов предпринял международный (исходно – американский) Институт будущего жизни (Future of Life Institute – FLI [11]). В 2015 г. Институт участвовал в подготовке открытого письма “Научно-исследовательские приоритеты для создания надежного и полезного искусственного интеллекта”, которое подписали более 10 тыс. исследователей, среди которых видные фигуры мира ИИ. Письмо ссылается на статью трех авторов [12], где эти приоритеты описаны.

Одним из первых важных событий в современной озабоченности влиянием ИИ стала Асиломарская конференция, которую Институт будущего жизни провел в январе 2017 года (Асиломар – курорт и конференц-центр на побережье Калифорнии). Там была принята декларация “The Asilomar AI principles” [13], на которую ссылались авторы многих подобных документов в дальнейшем.

Одним из недавних самых известных текстов такого сорта стало обнародованное 22 марта 2023 г. тем же FLI “Открытое письмо: приостановить гигантские эксперименты в ИИ” [14]. Основной призыв, содержащийся в этом письме, – остановить по крайней мере на 6 месяцев “обучение систем, более мощных, чем GPT-4”. Письмо подписали глава SpaceX и создатель OpenAI Илон Маск, соучредитель Apple Стив Возняк, филантроп Эндрю Янг и еще десятки тысяч исследователей искусственного интеллекта, эксперты и другие общественно значимые фигуры.

FLI через 6 месяцев опросил ряд исходных авторов письма о том, что за эти шесть месяцев произошло [15]. Кажется, ни один не сказал, что было остановлено “обучение систем...”! Однако некоторые из этих авторов видели связь между появлением письма и активизацией работ по регламентации и ограничению ИИ на правительственном уровне.

Осенью 2023 г. FLI разработал рекомендации к Саммиту по ИИ в Великобритании [16], запланированному на 1–2 ноября 2023 г. Открытые письма, манифесты и меморандумы со стороны лидеров отрасли ИИ и международного сообщества продолжали возникать более активно в последние годы.

Российский Кодекс этики в сфере ИИ был принят 26 октября 2021 года [17]. Он разработан Альянсом в сфере ИИ, объединяющим ведущие технологические российские компании [18]. Почти одновременно рекомендации по этике ИИ были приняты в ноябре 2021 г. на 41-й сессии Генеральной конференции ЮНЕСКО [19].

28 июля 2023 г. состоялось обсуждение проблем ИИ в Совете безопасности ООН, на котором выступил Генеральный секретарь ООН.

Примечательно самое короткое обращение такого сорта. Оно содержит 22 английских слова и обнародовано Центром безопасности ИИ (Center for AI Safety), калифорнийской некоммерческой организацией, 30 мая 2023 г. Вот его текст [20]:

Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.

Вот возможный перевод на русский язык:

Снижение риска вымирания из-за искусственного интеллекта должно стать глобальным приоритетом, наряду с другими рисками масштаба целого общества, такими как пандемии и ядерная война.

Краткость текста, по объяснению авторов, должна отражать их консенсус и содействовать обсуждению высказанной мысли. Среди подписавших обращение: генеральный директор OpenAI Сэм Альтман, лауреаты премии Тьюринга Джеффри Хинтон и Йошуа Бенжио, главный научный сотрудник OpenAI Илья Суцкевер, генеральный директор DeepMind Демис Хассабис, а также эксперты и ученые из крупнейших университетов и технологических компаний.

21 июля 2023 г. семь ведущих компаний в области искусственного интеллекта – Amazon, Anthropic, Google, Inflection, Meta, Microsoft и OpenAI представили Президенту Байдену в Белом доме [21] свою совместную позицию о добровольном принятии обязательств по продвижению к безопасному и прозрачному развитию технологии искусственного интеллекта [22].

В сентябре 2023 г. прошла закрытая конференция AI Insight Forum, организованная лидером сенатского большинства Чаком Шумером [23]. В мероприятии приняли участие ведущие технологические лидеры (в том числе Илон Маск, Марк Цукерберг, Сатья Наделла), политики (65 сенаторов обеих партий), профсоюзные лидеры и защитники гражданских прав. Был достигнут широкий консенсус в отношении того, что искусственный интеллект должен начать регулироваться в ближайшее время, уже в 2024 г.

С содержательной точки зрения представляет интерес дискуссия, ведущаяся в DeepMind [24].

Из индивидуальных выступлений влиятельных современных мыслителей выделяется доклад “Искусственный интеллект и будущее человечества” Юваля Ноя Харари на Форуме фронтиров (Frontiers Forum) 29 апреля 2023 года в Монтре, Швейцария [25]. В своем выступлении Ю.Н. Харари подчеркивает опасность манипулирования массовым мировоззрением, возникновения новых религий, создание которых искусственным интеллектом может быть запущено людьми: “люди и компании, которые контролируют новые оракулы искусственного интеллекта, станут чрезвычайно могущественными”. Ключевым элементом здесь, как он полагает, является овладение ИИ человеческим языком, “взламывание” языка. Это дает возможность разрушить демократию, породить хаос. Возникает вопрос о цифровом колониализме [26].

Представляющая интересы писателей Авторская гильдия 18 июля 2023 г. опубликовала открытое письмо, которое затем подписали и тысячи индивидуальных литераторов. Письмо адресовано группе лидеров генеративного ИИ и ИТ в целом: Sam Altman, CEO, OpenAI; Sundar Pichai, CEO, Alphabet; Mark Zuckerberg, CEO, Meta; Emad Mostaque, CEO, Stability AI; Arvind Krishna, CEO, IBM; Satya Nadella, CEO, Microsoft [27]. Авторы требуют от этих лидеров обеспечить использование авторских работ генеративным ИИ только с разрешения авторов и со справедливой компенсацией авторам. Актуальность этого письма обусловлена появлением коммерческого потока текстов, не являющихся плагиатом в классиче-

ском смысле этого слова, но созданных с помощью генеративного ИИ “по мотивам” текстов, созданных людьми – писателями и журналистами.

В то же время обсуждение на высшем уровне в США привело к появлению письма сотен деятелей искусства [28], адресованного сенатору Шумеру (лидеру большинства в Сенате) и членам Конгресса США:

“Мы пишем это письмо сегодня как профессиональные художники, использующие инструменты генеративного искусственного интеллекта, которые помогают нам вкладывать душу в нашу работу <...>.

<...> искусственный интеллект делает искусство более доступным или позволяет осваивать совершенно новые художественные средства. Как и предыдущие инновации, эти инструменты снижают барьеры в создании произведений искусства <...>.

Сегодня мы выступаем в защиту будущего, наполненного более богатыми и доступными творческими инновациями для грядущих поколений художников. Художники вдыхают жизнь в искусственный интеллект, направляя его инновации на позитивную культурную эволюцию, одновременно расширяя основные человеческие аспекты, которых ему по своей сути недостает”.

6. ЕВРОПЕЙСКИЙ ЗАКОН ОБ ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ И СВЯЗАННЫЕ С НИМ ЗАКОНЫ

Большинство документов, относящихся к этике и политике в области ИИ, предостережениям и ограничениям, носит декларативный характер, в частности, не включают в себя непосредственные жесткие запреты и экономические санкции за их нарушение. В связи с этим представляет интерес рассмотрение процесса регулирования данной сферы, идущего в рамках Евросоюза.

6.1. Действующие и вводимые общеевропейские акты о цифровых данных, рынках и услугах

Общий регламент ЕС по защите данных (General Data Protection Regulation – GDPR) вступил в силу в 2018 году. Основные задачи, решаемые этим регламентом, – это защита персональных данных, фиксация прав граждан по отношению к организациям, которые эти данные собирают.

Летом 2022 года Европейский парламент принял Закон о цифровых услугах The Digital Services Act (DSA) и Закон о цифровых рынках The Digital Markets Act (DMA). Затем они были приняты Советом Европейского союза, подписаны президентами обоих институтов и опубликованы в Официальном журнале ЕС. Сроки ввода в действие отдельных норм законов различны, но с первых месяцев 2023 г. эти законы применяются в полном объеме. Вопросы, относящиеся к компетенции каждого из законов, связаны друг с другом.

Закон DMA нацелен на обеспечение конкуренции на цифровых рынках. DMA затрагивает только так называемых “гейткиперов” (“gatekeepers”), мы можем их назвать “смотрящими”. Это компании, представляющие крупные онлайн-платформы и очень крупные онлайн-поисковые системы – это платформы, которые в среднем ежемесячно посещают более 45 миллионов пользователей в ЕС. На данный момент выделено 19 платформ, при этом Amazon и немецкий ритейлер Zalando уже опротестовали свое включение в список этих онлайн-платформ.

Закон DSA в первую очередь заботится о прозрачности и защите прав потребителей, он требует от компаний оценки рисков, определения мер по их снижению и прохождения сторонних аудитов на предмет соответствия требованиям. В основном он относится к социальным сетям, сообществам пользователей и онлайн-сервисам с бизнес-моделью, основанной на рекламе. Закон требует раскрытия алгоритмов, относящихся к учету и классификации пользователей, рекомендательным алгоритмам и модерации контента. (Слово “алгоритмы” в публичном дискурсе начало употребляться в основном по отношению к алгоритмам манипулирования пользователем!) DSA распространяется на всех поставщиков цифровых услуг, при этом предъявляет дополнительные требования к тем, услуги которых используют более 10% населения ЕС. Платформам предписывается изыскать способы предотвращения или удаления публикаций, связанных с незаконными товарами, услугами или содержащих незаконный контент, а также обеспечить пользователям возможность сообщать о таком типе контента. Запрещается целевая реклама на основе интимных предпочтений, религии, этнической принадлежности или политических убеждений человека, а также ограничивается таргетинг рекламы на детей.

Европейская комиссия будет отвечать за обеспечение соблюдения DMA, в то время как национальные регулирующие органы будут отвечать за применение правил DSA. Оба закона грозят нарушителям существенными штрафами: 10 и 6% от общемирового оборота для DMA и DSA соответственно.

6.2. Предлагаемый Закон об ИИ

Для нас важно, что указанные законы DMA и DSA приняты в период подготовки европейского закона об искусственном интеллекте (Artificial Intelligence Act), точное название которого: Har-

monised Rules on Artificial Intelligence (EU Artificial Intelligence Act – EU AIA) [29]. Этот закон был предложен Европейской комиссией 21 апреля 2021 г. он будет введен в действие в полном объеме только с 2026 г. Но, несмотря на это, EU AI Act уже оказывает существенное влияние на события во всем мире ("Брюссельский эффект"). В данный момент рассмотрение EU AI Act находится на этапе "триалога", на котором три учреждения – Европейский парламент, Совет Европейского союза и Европейская комиссия – согласуют свои позиции по этому закону. Триалог начался 14 июня 2023 года, и ожидается, что окончательный текст будет подготовлен к концу 2023 года (жесткий дедлайн – февраль 2024 г.), в преддверии выборов в Европейский парламент в 2024 году.

Рассмотрим положения предлагаемого EU AI Act более подробно. Закон классифицирует системы, использующие ИИ, в соответствии с рисками, которые они несут:

1. **Системы низкого риска**. Они составляют большинство систем, используемых в настоящее время на рынке, например, спам-фильтры или видеоигры, применяющие ИИ. На эти системы законом не накладывается никаких дополнительных ограничений, они просто должны соответствовать действующему законодательству.

2. **Системы ограниченного риска**. От этих систем требуется прозрачность – пользователь должен быть проинформирован о том, что:

• он взаимодействует с системой искусственного интеллекта;

• система искусственного интеллекта будет использоваться для определения его характеристик или эмоций;

• контент, с которым они взаимодействуют, был сгенерирован с использованием искусственного интеллекта.

Примерами могут служить чат-боты и глубокие фейки.

3. **Системы высокого риска**. Это системы, которые могут оказать значительное влияние на жизнь пользователя. На них должны быть наложены существенные ограничения, включая обязательства по управлению рисками и управлению данными. (В тексте EU AI Act имеются подробные разъяснения, относящиеся к таким системам.)

4. **Системы неприемлемого риска**. Запрещены к продаже на рынке ЕС и включают в себя те, которые манипулируют людьми без их согласия или позволяют проводить социальный скоринг, а также системы биометрической идентификации как в режиме реального времени, так и удаленно, и отсроченно.

6.3. Системы высокого риска в Законе об ИИ

Согласно статье 6 закона EU AI Act, система имеет высокий риск, если она предназначена для использования в качестве компонента безопасности продукта. В этой ситуации она должна пройти оценку соответствия третьей стороной, связанную с рисками для здоровья и безопасности.

Кроме того, указаны 8 областей использования систем, влекущих их высокий риск:

1. Системы идентификации, используемые для биометрической идентификации и делающие выводы о личностных характеристиках, включая системы распознавания эмоций.

2. Системы для критически важной инфраструктуры и защиты окружающей среды.

3. Системы образования и профессиональной подготовки, используемые в процессе обучения (в частности, для оценивания).

4. Системы, влияющие на занятость, поиск и поддержку талантов и доступ к самозанятости.

5. Системы, влияющие на доступ и использование частных и государственных услуг и льгот, включая те, которые используются в страховании.

6. Системы, используемые в правоохранительных органах и от имени правоохранительных органов.

7. Системы управления миграцией, предоставлением убежища и пограничным контролем, включая системы, используемые от имени соответствующего государственного органа.

8. Системы, используемые при отправлении правосудия и демократических процессах, включая системы, используемые от имени судебной власти.

В обсуждениях 2023 года, естественно, центральную роль играли базовые модели (foundation models) и генеративный ИИ (generative AI). В частности, было предложено следующее определение базовых моделей:

"*Модели искусственного интеллекта, которые разрабатываются на основе алгоритмов, предназначенных для оптимизации общности и универсальности выходных данных. (Эти) модели часто обучаются на широком спектре источников данных и больших объемах данных для выполнения широкого спектра задач, включая некоторые, для которых они специально не разрабатывались и не обучались*".

Дополнительно разъясняется, что предварительно подготовленные модели, предназначенные для "более узкого, менее общего, более ограниченного набора приложений", не должны рассматриваться как базовые модели.

Закон EU AI Act также определяет генеративный ИИ:

"*Системы искусственного интеллекта, специально предназначенные для создания с различным уровнем автономии контента, такого как сложный текст, изображения, аудио или видео*".

Закон фиксирует обязательства для поставщиков базовых моделей соответствовать следующим требованиям, прежде чем выводить свой продукт на рынок ЕС, с помощью следующих механизмов:

1. Обеспечить идентификацию, выявление и снижение разумно ожидаемых рисков для здоровья, безопасности, основных прав, окружающей среды и демократии во время разработки модели посредством соответствующего проектирования, тестирования и анализа.

2. Включать в обучающие модели только те наборы данных, на которые распространяются соответствующие гарантии их формирования, тем самым снижая потенциальные риски, возникающие из-за низкого качества данных, предвзятости и ненадежности источников.

3. Использовать консультации экспертов, механизмы документирования и широкое тестирование на этапах проектирования и разработки модели, чтобы обеспечить поддержание надлежащих уровней эффективности, предсказуемости, интерпретируемости, улучшаемости, безопасности и кибербезопасности на протяжении всего жизненного цикла модели.

4. Соблюдать соответствующие стандарты сокращения энергопотребления, повышения энергоэффективности и минимизации отходов. Кроме того, поставщики должны разрабатывать базовые модели, предусматривая возможности измерения потребления энергии и ресурсов и, где это технически возможно, предусматривать встроенные показатели воздействия модели на окружающую среду на протяжении всего жизненного цикла.

5. Способствовать соблюдению нижестоящими поставщиками требований закона путем предоставления обширной технической документации и удобных для пользователя инструкций по использованию.

6. По аналогии с системами высокого риска, создавать систему управлением качеством для обеспечения соответствия требованиям и документирования этого соответствия.

7. Зарегистрировать базовые модели в базе данных ЕС, следуя инструкциям, приведенным в приложении к Закону.

8. Вести техническую документацию в течение 10 лет с момента поступления базовой модели на рынок и хранить ее копию в компетентных национальных органах.

9. Дополнительные обязательства для поставщиков генеративных систем ИИ:

a. Соблюдать обязательства по обеспечению прозрачности.

b. Следить за тем, что при обучении, проектировании и разработке систем генеративного ИИ контент, генерируемый такими системами, не нарушает законодательство ЕС.

c. Опубликовать подробное резюме использования обучающих данных, защищенных законом об авторском праве.

6.4. Обсуждение проекта

Краткий обзор хода создания и обсуждения проекта EU AIA до весны 2023 г. проведен российским юристом с общих позиций [30].

Критическим моментом в обсуждении стала сессия Европарламента в июне 2023 г., в ходе которой в проекте закона появился ряд принципиально новых предложений, касающихся ограничений на создание и использование базовых моделей и генеративного ИИ. Мир успел измениться, пока готовился закон. Обсуждение привело к почти немедленной реакции – появилось открытое письмо: "Open letter to the representatives of the European Commission, the European Council and the European Parliament" с подзаголовком "Artificial Intelligence: Europe's chance to rejoin the technological avant-garde", подписанное 150 европейскими венчурными капиталистами и руководителями технологических компаний [31], информация была перепечатана многими изданиями. Риторика письма не адресуется к конкретным положениям законопроекта, стилистически письмо аналогично алармистским призывам компаний, относящимся к опасностям ИИ. Однако в данном случае компании озабочены тем торможением в развитии ИИ в Европе и проигрыше американским конкурентам, к которому, по их мнению, может привести принятие закона:

"*...законопроект поставит под угрозу конкурентоспособность и технологический суверенитет Европы без эффективного решения проблем, с которыми мы сталкиваемся и будем сталкиваться в будущем*.

Это особенно верно в отношении генеративного искусственного интеллекта. Согласно версии, недавно принятой Европейским парламентом, базовые модели, независимо от вариантов их использования, будут жестко регулироваться, и компании, разрабатывающие и внедряющие такие системы, столкнутся с непропорциональными затратами на соблюдение требований и непропорциональными рисками ответственности. Такое регулирование может привести к тому, что высокоинновационные компании перенесут свою деятельность за рубеж, инвесторы выведут свой капитал из разработки европейских базовых моделей и европейского искусственного интеллекта в целом. Результатом

стал бы критический разрыв в производительности труда по обе стороны Атлантики.

<...> желание закрепить регулирование генеративного ИИ в законодательстве и следовать жесткой логике соблюдения требований является бюрократическим подходом, неэффективным для достижения своей цели.

<...> европейское законодательство должно ограничиться изложением общих принципов подхода, основанного на оценке риска. Внедрение этих принципов должно быть поручено специализированному регулирующему органу, состоящему из экспертов на уровне ЕС. <...>

<...> Создание трансатлантической структуры также является приоритетом <...>, чтобы создать юридически обязывающие равные условия игры".

На этом дискуссия со стороны бизнеса не закончилась. Слышны голоса и в поддержку Закона, в частности регулирования генеративного ИИ [32]. Еще раньше утверждалось, что Закон хотя и не выгоден большим монополиям, но необходим малому инновационному бизнесу [33].

7. ИНИЦИАТИВА G7

На ежегодном саммите G7, прошедшем в Хиросиме в мае 2023 г., по предложению Японии был запущен процесс Hiroshima AI Process (HAP). "Большая семерка" договорилась о создании министерского форума для стимулирования добросовестного использования искусственного интеллекта. HAP – это попытка G7 определить дальнейший путь регулирования искусственного интеллекта. HAP создает форум для международных дискуссий по инклюзивному управлению ИИ и интероперабельности для достижения общего видения и цели создания надежного ИИ на глобальном уровне. HAP будет работать в тесной связи с организациями, включая Организацию экономического сотрудничества и развития (OECD) и Глобальное партнерство по искусственному интеллекту (GPAI). К концу 2023 года будет подготовлен проект первого документа HAP. На сайте OECD размещен подготовительный документ [34].

30 октября 2023 г. министры G7 подписали кодекс ИИ – Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems [35]. Кодекс состоит из преамбулы и одиннадцати рекомендаций, формулируемых в повелительном наклонении. Он адресован организациям, ведущим исследования и разработки ИИ (что указано в названии), а также организациям гражданского общества, частного и/или государственного сектора. Кодекс базируется на существующих принципах искусственного интеллекта ОЭСР и учитывает последние разработки в области систем искусственного интеллекта; он призван помочь воспользоваться этими достижениями и ответить на риски и вызовы, связанные с ними. Он применяется на этапах проектирования, разработки, развертывания и использования продвинутых систем искусственного интеллекта.

В тот же день 30 октября 2023 г. Президент Европейской комиссии Урсула фон дер Ляйен присоединила свою подпись к документу и заявила [36]:

"*Уже являясь лидером в области регулирования, основанном на принимаемом нами Законе об искусственном интеллекте, ЕС также вносит свой вклад в создание и управление безопасным ИИ на глобальном уровне. Я рада приветствовать международные руководящие принципы G7 и добровольный кодекс поведения, отражающие ценности ЕС по продвижению заслуживающего доверия искусственного интеллекта. Я призываю разработчиков искусственного интеллекта подписать и внедрить этот Кодекс поведения как можно скорее*".

Пока трудно судить, будет ли принятый документ реальным регулятором (в том числе на национальном уровне) научно-технологических разработок, экономики, потребления, будет ли он реально формулировать исследовательские приоритеты, или останется на уровне декларации, стимулирующей общественное обсуждение.

8. В ПОИСКАХ ВЫХОДА. ПЕРСПЕКТИВЫ, ПРЕДЛОЖЕНИЯ. НАУЧНЫЕ ИССЛЕДОВАНИЯ И ПУБЛИКАЦИИ, ОБРАЗОВАНИЕ, ПРОСВЕЩЕНИЕ

Видимо, основными группами, участие которых может снизить риски негативного влияния расширения ИИ являются:

1. Правительства, межнациональные организации – путем введения ограничений и преференций.

2. Корпорации, их группы и саморегулирующиеся ассоциации разработчиков и производителей – путем осознания и реализации моральной ответственности.

3. Ассоциации потребителей – формируя общественное мнение.

4. Граждане – путем осознанного выбора и соблюдения закона, в том числе по вопросам, по которым не может быть традиций и исторических норм.

В предыдущих разделах мы касались, прежде всего, двух первых групп. Однако их представители сами подчеркивают необходимость вовлечения в проблему снижения рисков всего общества.

В то же время опросы общественного мнения, например, [37, 38] показывают негативную установку общественного большинства по отношению к ИИ, последним достижениям в этой обла-

сти, а часто и по отношению к дальнейшему развитию цифровых технологий вообще. Но это ясно и без опросов.

Проведенный в 2023 г. опрос 14 тыс. потребителей и компаний в 25 странах показал, что три четверти из них озабочены этичностью использования ИИ. Опрос 1500 австралийцев показал, что около 60% из них считает, что ИИ больше приносит проблем, чем решает.

В данном разделе мы рассмотрим перспективы формирования – просвещения, образования профессионалов и просто граждан, которое может в своей деятельности и мировосприятии фактически противостоять вредоносным событиям и процессам типа тех, о которых говорит Ю.Н. Харари [25].

Можно считать, что мы говорим о новом Веке просвещения. При этом нам кажется полезным выходить за рамки наиболее популярных сегодня дискуссий, рассматривать ситуацию в более широком контексте.

8.1. Достоверность научных публикаций

В эпоху, предшествующую развитию ИИ, исследователи различных областей науки устанавливали требования, которые могут содействовать достоверности объявляемых, публикуемых результатов и даже гарантировать такую достоверность. Примером этого является появление естественно-научного метода Галилея и др., позитивизм XX в. и т.п.

Однако серьезные проблемы с достоверностью оставались. Это видно даже на примере математики, где, казалось бы, имеются “объективные критерии” правильности доказательств. Даже здесь имеются обескураживающие случаи. Поясним, в чем состоит одна из проблем. Уже представленное доказательство можно проверить. Априори проверка должна быть делом существенно более простым, чем построение, придумывание доказательства и его аккуратная запись. Однако часто первоначально публикуемое сложное доказательство содержит “недочеты” – в нем могут быть упущены случаи, не согласованы определения и обозначения. Чтение такого доказательства может оказаться сопоставимым по трудности с его записью, по крайней мере – требовать понимания общей конструкции, но дефекты изложения препятствуют такому пониманию. В результате традиционный, “надежный” механизм верификации – рецензирование (peer review) дает сбои, рецензент смотрит на правдоподобность и важность результата, возможно, на авторитетность автора и “пропускает” работу. Одни из авторов (Г. Перельман) вообще отказываются от механизма рецензирования, считая, что серьезной гарантией правильности является использование результатов и техники в сообществе. Другие (В. Воеводский) считают необходимым всю математику и деятельность математиков сделать более машинно верифицируемыми.

Впечатляющий современный обзор проблемы доказательства в математике имеется в работах Н. А. Вавилова, где, в частности, показана расплывчатость самого современного понятия доказательства [39].

Если говорить о позитивном применении ИИ, то постепенно растет число случаев, когда при написании текста доказательства происходит его локальная верификация средствами ИИ. Видимо, будет идти совершенствование механизмов рецензирования в направлении создания научных блок-чейнов.

Сейчас, однако, ситуация с миром текстов вообще становится намного более драматичной. Этому, собственно, и посвящено упомянутое выше известное выступление Харари [25], где он говорит о том, что ИИ “взломал”, “хакнул” человеческий язык. Все большее количество текстов, в том числе – полезных, порождается генеративным ИИ. При этом сам подход генеративного ИИ не адресуется к научной или исторической истине или к точности цитирования, корректности заимствования и т.д. Эти априорные человеческие ценности, воспитываемые в обществе, никто не закладывал в большие языковые модели ИИ. Более того, в некоторых ситуациях адресация к истине невозможна, бессмысленна и даже вредна, например, при создании художественного произведения. Даже если мы отнесем к исходному большому массиву языковых данных только научные истины, нет гарантии, что порожденный текст “научной статьи” будет верифицируем, он может быть просто “очень похож” на доказательный, но, например, содержать ничем не обоснованное число, а это число – физически невозможно или, наоборот, составляло бы предмет пока не сделанного научного открытия.

Таким образом, с одной стороны, встает проблема “встраивания” в системы генеративного ИИ механизмов автоматической верификации. Как это сделать – пока не ясно. С другой стороны, возникает сверхактуальная проблема защиты научного знания (сегодняшнего и будущего) от дополнительного загрязнения, порождаемого использованием генеративного ИИ.

8.2. Ассоциации потребителей

Как показывает опрос Форбса [38], большинство людей уже сейчас используют ChatGPT вме-

сто поисковика, то есть предпочитают возможно сомнительный пересказ первичной информации.

Имеется шанс вести просветительскую работу, снижающую риски ИИ, через ассоциации потребителей. Пример такого просвещающего документа – [40].

8.3. Ковчег

Предлагается начать выстраивать блок-чейн науки на основе, с одной стороны, опыта блокчейнов финансов и произведений искусства (в том числе – цифрового, но не только). С другой стороны, при организации системы доверия предлагается использовать уже существующие, традиционные механизмы научной корпорации, опирающиеся на исследователей и их авторитет. В России это представители структур Академии наук и вузов: заведующие отделами, кафедрами, председатели советов по защитам диссертаций, главные редакторы научных журналов. В свою очередь они связаны с большим числом докторов и кандидатов наук, также обладающих научным авторитетом, у них есть увлеченные наукой аспиранты и студенты, ученики, работающие в различных сферах. При всех своих недостатках традиционная научная иерархия воспроизводит научные кадры и может обеспечивать квалифицированную оценку достоверности информации, в первую очередь – подбор и установление аутентичности существующих информационных источников и контроль за появляющимися научными публикациями результатов исследований. В качестве единого, распределенного и защищенного хранилища предлагается использовать Ковчег знаний МГУ [41, 42]. В Ковчеге будут накапливаться публикации, в том числе, например, массивы экспериментальных данных и записи курсов лекций, классические учебники и литературные произведения, достоверность которых проверена и подтверждена людьми.

В романе Рея Брэдбери символом уничтожения человеческой культуры было сжигание книг. Сегодня мы приобретаем еще более мощный символ в форме потопа псевдознания. Ковчег становится инструментом и символом спасения культуры.

8.4. Энциклопедия

Общеизвестна и общеупотребительна Википедия и ее аналоги. Благодаря цифровым технологиям издание такого сорта мгновенно становится доступным повсеместно, оно всегда с тобой. Ее полезность по сравнению с энциклопедиями прошлого существенно повышается также за счет богатой системы ссылок, в том числе – доступных во вне, мультимедийности и постоянной актуализации. Кроме того, благодаря гибкости цифрового формата можно иметь в энциклопедии слои, предназначенные для различных пользователей – в частности, по исходному знанию области и общекультурному багажу пользователей. Например, одна и та же статья о каком-то математическом термине может иметь слой для школьника, слой для взрослого нематематика и слой для профессионального математика.

Наконец, для википедий отсутствует издательский цикл многотомного и многолетнего издания. Может достигаться высокая степень параллелизма в ее создании и актуализации.

Сегодня создаваемая, редактируемая и рецензируемая людьми цифровая энциклопедия в интернете может выстраиваться как обращенный ко всему населению портал Ковчега, через который можно дойти до всего знания, которое в Ковчеге сохранено. Конечно, это знание является общественным достоянием и хранится на основании открытой лицензии, предусматривающей ссылку на источник, но не требует разрешения для использования.

Энциклопедия может быть вкладом научного сообщества и ИТ-индустрии в культуру народа и защиту ее от разрушения в силу разных причин, включая массовое некорректное использование генеративного ИИ. Энциклопедия может стать объясняющей, намного более диалогичной. Можно к ней подключить творческий ИИ, но именно для объяснения. При этом знанием является не пересказ ИИ, а именно текст Энциклопедии.

8.5. Общее образование и просвещение. Массовое сознание

Угрозы, подобные тем, о которых говорит Харари, могут быть снижены за счет повышения уважения к достоверному знанию, умения к нему обращаться и его использовать. Соответствующие установки и стереотипы могут формироваться уже в общеобразовательной школе. Здесь необходима преемственность – от детского сада до образования в течение всей жизни и прохождение этого пути в подготовке учителей [43, 44].

Такое формирование может и должно быть частью построения новой школы – адекватной потребностям и возможностям XXI века. В условиях, когда знание общедоступно, основным становится не хранение во внутренней, биологической памяти человека готовых ответов и алгоритмов, а ориентация в мире. Эта ориентация включает “большие идеи”, к которым можно отнести и общие модели деятельности. Их невозможно “выучить наизусть”, как “квадрат гипотенузы равен сумме квадратов катетов”, “для того чтобы найти четвертое число, нужно перемножить второе и третье число и произведение разделить на пер-

вое”. Большие идеи постигаются в ходе их постепенного самостоятельного открытия, в ходе применения в тех или иных частных ситуациях.

Критически важной большой идеей сегодня является общее умение задавать вопрос, обращенный к своей “внешней памяти” – интернету, а получив разнообразные ответы, извлечь из них то, что тебе нужно и полезно.

Как эту идею начать формировать у учащегося начальной школы (потом начинать уже поздно)? Эффективно это делать в составе воплощения более общей “большой идеи” самостоятельного открытия знания и его проверки. Или проверки знания, полученного от кого-то. Например, ученик в процессе решения осмысленных и интересных ему задач получает опыт нахождения в интернете (или его учебной симуляции) разнообразных, дополняющих друг друга или противоречащих друг другу ответов и советов.

9. ЗАКЛЮЧЕНИЕ

Настоящая статья представляет собой изложение общих взглядов автора на проблематику ИИ. Она основана на 60-летнем опыте включенности в исследования и разработки в этой области. Началом я считаю свое участие в создании в начале 1960-х гг. аналого-цифрового устройства, которое использовалось моей мамой Евгенией Тихоновной Семеновой (участником разработки ЭВМ “Стрела” в начале 1950-х гг.) в ее работе по распознаванию ограниченного набора устных русских слов. За этим следовало участие в разработке компилятора с Лиспа, участие в написании обзоров развития программирования, в проекте ЭВМ 5-го поколения Института кибернетики АН УССР, реализации концепций ситуационного управления. К тому же времени относится мое знакомство и начало дружбы с Дмитрием Александровичем Поспеловым, продолжавшейся до его смерти. Моими друзьями были Симур Паперт и Владимир Андреевич Успенский. Мне довелось работать с А.Н. Колмогоровым последние годы его жизни. Я встречался с Джоном МакКарти (благодаря академику А.П. Ершову), Марвином Минским (благодаря С. Паперту), Стивом Вольфрамом (благодаря Патрику Суппесу).

В начале 1990-х гг. в Институте новых технологий, который я создал вместе с Борисом Беренфельдом, Валентином Носкиным и Александром Мордуховичем, среди других исследовательских проектов я запустил работу над созданием первого русского спел-чекера. Ускорение в развитии цивилизации я испытывал непосредственно в своей жизни.

Совершенно очевидно, что в той же степени, в которой любая экономика пользуется мировыми достижениями ИТ, она разделяет и связанные с этим риски, особенно с учетом того, что эти достижения делают ситуацию мало зависящей от языка.

Приведем пару наугад взятых примеров. РГСУ создал Проект платформы социального скоринга “Мы” [45]. Название проекта копирует название знаменитой антиутопии Евгения Замятина, созданной около 1920 г. Многие сюжеты современных опасений, относящихся к ИИ, были предвидены и предчувствованы в серии антиутопий, начатой романом Замятина. Прежде всего – в “1984” Джорджа Оруелла. Будет ли социальный скоринг или распознавание лиц в метро законным во Франции? А в России?

Китай озабочен случаями криминальной имитации голоса, лица личности [46]. При этом, видимо, слухи о тотальном социальном скоринге, ведущемся китайскими властями, преувеличены.

Мы уверены, что и изучение общественной озабоченности применениями ИИ, и изучение попыток регулирования ИИ для нас могут быть весьма полезными, независимо от того, планируем ли мы с ними формально ассоциироваться.

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Работа выполнена при финансовой поддержке гранта № 23-Ш05-11 Междисциплинарных научно-образовательных школ Московского государственного университета им. М. В. Ломоносова.

СПИСОК ЛИТЕРАТУРЫ

1. Об использовании микрокалькуляторов в учебном процессе (Инструктивно-методическое письмо). НИИ содержания и методов обучения АПН СССР и Главное управление школ Министерства просвещения СССР // Математика в школе, 1982. 3. С. 6–8.
2. Постановление ЦК КПСС и Совета Министров СССР от 28 марта 1985 года № 271 “О мерах по обеспечению компьютерной грамотности учащихся средних учебных заведений и широкого внедрения электронно-вычислительной техники в учебный процесс”. Вопросы образования, 2005. 3. С. 341–346 (in Russian). http://vo.hse.ru/arhiv.aspx?catid=252&z=808&t_no=809&ob_no=854 (актуально на 3.09.2023).
3. *McCarthy J.* What is artificial intelligence? Stanford University, Stanford, USA, 2007. http://www-formal.stanford.edu/jmc/ (дата обращения: 31.10.2023).
4. *Выготский Л.С.* Инструментальный метод в психологии. Собр. соч. В 6 т. Т. 1. 1982.

http://elib.gnpbu.ru/text/vygotsky_ss-v-6tt_t1_1982/go,108;fs,1/ (дата обращения: 31.10.2023). Eng. translation: *Vygotsky L.S.* The instrumental method in psychology, 1981. https://www.marxists.org/archive/vygotsky/works/1930/instrumental.htm (дата обращения: 31.10.2023).

5. *Semenov A.L., Ziskin K.E.* Expanded Personality as the Main Entity and Subject of Philosophical Analysis: Implications for Education. Doklady Mathematics. 2023. V. 108. № 4. P. 331–341. ISSN 1064-5624. https://doi.org/10.1134/S1064562423700965
6. *Floridi L.* The fourth revolution: how the infosphere is reshaping human reality. Oxford University Press, UK, 2014. ISBN 978-0-19-960672-6. https://www.academia.edu/14408794/The_Fourth_Revolution_How_the_Infosphere_is_Reshaping_Human_Reality (дата обращения: 31.10.2023).
7. Кохлеарный имплантат. Википедия. https://ru.wikipedia.org/?curid=120843&oldid=132635763 (дата обращения: 31.10.2023).
8. *Платон*. Федр, 274c–275b. Сочинения в 4-х томах, т. 3. М.: Мысль, 1971. С. 455–542.
9. *Suleyman M.* My new Turing test would see if AI can make $1 million. MIT Technology Review, July 14, 2023. https://www.technologyreview.com/2023/07/14/1076296/mustafa-suleyman-my-new-turing-test-would-see-if-ai-can-make-1-million/ (дата обращения: 31.10.2023).
10. Образование — это не изучение фактов, а тренировка мышления. https://theocrat.ru/eng/content-64698/ (дата обращения: 31.10.2023).
11. Future of Life Institute. Официальный сайт. https://futureoflife.org/ (дата обращения: 31.10.2023).
12. R*ussell S., Dewey D., Tegmark M.* Research Priorities for Robust and Beneficial Artificial Intelligence. Future of Life Institute, 2015. https://futureoflife.org/data/documents/research_priorities.pdf?x89399 (дата обращения: 31.10.2023).
13. The Asilomar AI principles. Published August 11, 2017. https://futureoflife.org/open-letter/ai-principles/ (дата обращения: 31.10.2023).
14. Pause Giant AI Experiments: An Open Letter. Published March 22, 2023. https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (дата обращения: 31.10.2023).
15. As Six-Month Pause Letter Expires, Experts Call for Regulation on Advanced AI Development. Published September 21, 2023. https://futureoflife.org/ai/six-month-letter-expires/ (дата обращения: 31.10.2023).
16. FLI Recommendations for the UK Global AI Safety Summit, Bletchley Park, 1-2 November 2023. URL: https://futureoflife.org/wp-content/uploads/2023/09/FLI_AI_Summit_Recommendations.pdf (дата обращения: 31.10.2023).
17. Кодекс этики в сфере ИИ. Альянс в сфере искусственного интеллекта. https://ethics.a-ai.ru/ (дата обращения: 31.10.2023).
18. Альянс в сфере искусственного интеллекта. Официальный сайт. https://a-ai.ru/ (дата обращения: 31.10.2023).
19. Recommendation on the Ethics of Artificial Intelligence. UNESCO, 2021. 43 p. https://unesdoc.unesco.org/ark:/48223/pf0000381137 (дата обращения: 31.10.2023).
20. Statement of AI Risk. AI experts and public figures express their concern about AI risk. Center of IA Sagety. https://www.safe.ai/statement-on-ai-risk (дата обращения: 31.10.2023).
21. FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI. July 21, 2023. https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/ (дата обращения: 31.10.2023).
22. Ensuring Safe, Secure, and Trustworthy AI. https://www.whitehouse.gov/wp-content/uploads/2023/07/Ensuring-Safe-Secure-and-Trustworthy-AI.pdf (дата обращения: 31.10.2023).
23. *Jeans S.* The first AI Insight Forum: when tech titans, lawmakers, and critics converge. Daily AI, September 13, 2023. https://dailyai.com/2023/09/the-first-ai-forum-when-tech-titans-lawmakers-and-critics-converge/ (дата обращения: 31.10.2023).
24. DeepMind alignment team opinions on AGI ruin arguments. August 13, 2022. https://www.lesswrong.com/posts/qJgz2YapqpFEDTLKn/deepmind-alignment-team-opinions-on-agi-ruin-arguments (дата обращения: 31.10.2023).
25. AI and the future of humanity. Yuval Noah Harari at the Frontiers Forum. April 29, 2023. Montreux, Switzerland. https://www.youtube.com/watch?v=LWiM-LuRe6w (дата обращения: 31.10.2023).
26. *Jeans S.* Digital colonialism in the age of AI and machine learning. Daily AI, October 28, 2023. https://dailyai.com/2023/10/digital-colonialism-in-the-age-of-ai-and-machine-learning/ (дата обращения: 31.10.2023).
27. Open Letter to Generative AI Leaders. https://actionnetwork.org/petitions/authors-guild-open-letter-to-generative-ai-leaders (дата обращения: 31.10.2023).
28. Open Letter: Artists Using Generative AI Demand Seat at Table from US Congress. https://creativecommons.org/about/policy-advocacy-

copyright-reform/open-letter-artists-using-generative-ai-demand-seat-at-table-from-us-congress/ (дата обращения: 31.10.2023).

29. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Brussels, April 2, 2021. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN (дата обращения: 31.10.2023).

30. *Низов В.* EU AIA, или Как нужно организовывать законотворческий процесс. Портал zakon.ru, 17.05.2023. https://zakon.ru/blog/2023/05/17/eu_aia_ili_kak_nuzhno_organizovyvat_zakonotvorcheskij_process (дата обращения: 31.10.2023).

31. *Butcher M.* European VCs and tech firms sign open letter warning against over-regulation of AI in draft EU laws. TechCrunch.com, June 30, 2023. https://techcrunch.com/2023/06/30/european-vcs-tech-firms-sign-open-letter-warning-against-over-regulation-of-ai-in-draft-eu-laws/ (дата обращения: 31.10.2023).

32. European business leaders, startup founders, and investors call for regulating AI foundation models under the EU AI Act. Published October 09, 2023. https://www.ai-statement.com/ (дата обращения: 31.10.2023).

33. *Verdi G.* General-Purpose AI fit for European small-scale innovators. Published October 5, 2022. https://www.digitalsme.eu/general-purpose-ai-fit-for-european-small-scale-innovators/ (дата обращения: 31.10.2023).

34. G7 Hiroshima Process on Generative Artificial Intelligence (AI) Towards a G7 Common Understanding on Generative AI. EOCD. Published on September 07, 2023. 37 p. https://doi.org/10.1787/bf3c0c60-en. https://www.oecd.org/publications/g7-hiroshima-process-on-generative-artificial-intelligence-ai-bf3c0c60-en.htm (дата обращения: 31.10.2023).

35. Hiroshima Process International Code of Conduct for Advanced AI Systems. Published 30 October 2023. URL: https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems (дата обращения: 31.10.2023).

36. Hiroshima Artificial Intelligence Process: EU Commission welcomes voluntary code of conduct for AI developers. EU Today, October 30, 2023. https://eutoday.net/hiroshima-artificial-intelligence-process/ (дата обращения: 31.10.2023).

37. *Coghlan J.* Consumer surveys show a growing distrust of AI and firms that use it. Cointelegraph, August 29, 2023. https://cointelegraph.com/news/consumers-increase-distrust-artificial-intelligence-salesforce-survey (дата обращения: 31.10.2023).

38. *Haan K.* Over 75% Of Consumers Are Concerned About Misinformation From Artificial Intelligence. Forbs Advisor, Jul 20, 2023. https://www.forbes.com/advisor/business/artificial-intelligence-consumer-sentiment/ (дата обращения: 31.10.2023).

39. *Vavilov N.* Reshaping the metaphor of proof. Phil. Trans. R. Soc. A. 377, 20180279. https://doi.org/10.1098/rsta.2018.0279

40. A consumer checklist on protecting consumers from the risks of AI. The European Consumer Organization, April 2022. https://www.beuc.eu/sites/default/files/publications/beuc-x-2022-039_protecting_consumers_-from_the_risks_of_ai_-_a_consumer_checklist.pdf (дата обращения: 31.10.2023).

41. Ковчег знаний МГУ. Информационная система Московского государственного университета имени М.В. Ломоносова. https://arc.msu.ru/ (дата обращения: 31.10.2023).

42. *Горячко В.В., Бубнов А.С., Раевский Е.Н., Семенов А.Л.* Цифровой ковчег знания. Доклады РАН, Математика, информатика, процессы управления, 2022, 508. С. 128–133. Eng. translation: *Goryachko V.V., Bubnov A.S., Rayevskii E.V., Semenov A.L.* Digital Ark of Knowledge. Doklady Mathematics, 2022. V. 106. Suppl. 1. P. S113–S117. https://doi.org/10.1134/S1064562422060096. https://doi.org/10.31857/S2686954322070098

43. *Семенов А.Л., Абылкасымова А.Е., Поликарпов С.А.* Основания математического образования в цифровой век. Доклады РАН. Математика, информатика, процессы управления, 2023, т. 511. С. 3–12. Eng. translation: *Semenov A.L., Abylkassymova A.E., Polikarpov S.A.* Foundations of Mathematical Education in the Digital Age. Doklady Mathematics. 2023. V. 107. Suppl. 1. P. S1–S9. https://doi.org/10.1134/S1064562423700564 https://doi.org/10.31857/S2686954323700157

44. *Семенов А.Л., Абылкасымова А.Е.* Цифровые технологии как фактор преемственности в математическом образовании. Сб. Фундаментальные проблемы обучения математике, информатике и информатизации образования: сборник тезисов докладов международной научной конференции. 29 сентября–1 октября 2023 г. Елец: Елецкий государственный университет им. И.А. Бунина, 2023. ISBN 978-5-00151-379-7. С. 10–15.

45. *Алиев Д.Ф., Бородулина С.А.* Проект платформы социального скоринга (материалы к Ученому совету РГСУ). РГСУ Life, 07.09.2022. https://rgsu.net/about/activities/press_centre/life/platform/rgsu-life_20.html (дата обращения: 31.10.2023).

46. Authorities warn public against AI fraud World Internet Conference, China Daily, June 1, 2023. https://www.wicinternet.org/2023-06/01/c_891623.htm (дата обращения: 31.10.2023).

ARTIFICIAL INTELLIGENCE IN SOCIETY

Academician of the RAS A. L. Semenov[a,b,c]

[a] *Herzen University, St. Petersburg, Russia*
[b] *Institute of Education, HSE University, Moscow, Russia*
[c] *Lomonosov Moscow State University, Moscow, Russia*

The article is the author's review of the singularity in which events in the field of AI are developing. A general view is offered on the role of information technology revolutions as expanding the human personality. The current stage of personal expansion is considered, covering the last decade, and especially 2023. The most important and typical socially significant documents are considered, which express concern about the results of the development and application of AI, as well as documents that declare the conditions for a positive development of events: ethical principles, priority directions, requirements and restrictions. It takes a closer look at the pending European AI Act and how different groups are reacting to it. Cultural and historical factors are highlighted that can counteract the negative and catastrophic developments that become possible thanks to AI. Possible mechanisms for preserving genuine knowledge among professionals and disseminating it among the general public are analyzed.

Keywords: artificial intelligence, extended personality, concern about the negative impact of AI, interstate regulation, European AI Law, Hiroshima AI Process, Knowledge Ark, Encyclopedia