УДК 004.8

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В НЕСКОЛЬКИХ ФРАГМЕНТАХ

Ю. Е. Поляк[0000-0001-8411-335X]

Центральный экономико-математический институт РАН, г. Москва, Россия

polak@cemi.rssi.ru

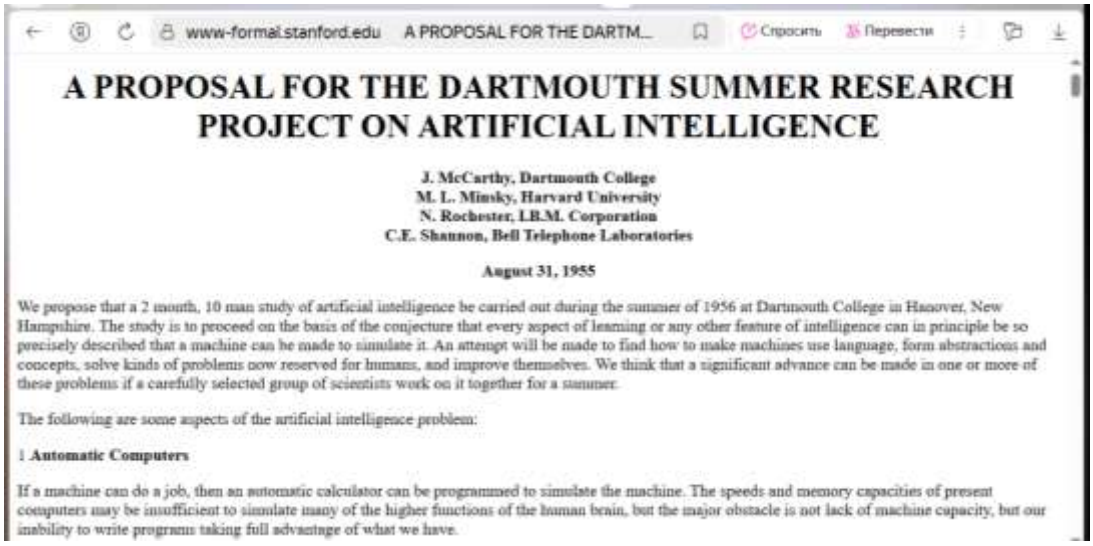
Аннотация

Работа представляет собой мозаику ярких фрагментов, описывающих отдельные аспекты искусственного интеллекта (ИИ). Это наброски общей картины, которая, вероятно, никогда не будет дописана, поскольку каждый день приносит информацию о новых достижениях, идеях и разработках, опасностях и угрозах. Обсуждение касается вопросов влияния ИИ на сокращения рабочих мест, разработки алгоритмов интеллектуальных игр, угроз и опасностей, исходящих от ИИ, этики ИИ, стандартов и международного регулирования ИИ. Каждый такой фрагмент – это обзор новейших (на середину января 2026 г.) российских и иностранных источников, включая цитаты, переводы, скриншоты и ссылки на оригинальные документы.

***Ключевые слова**: искусственный интеллект, Дартмутский семинар, предшественники ИИ, разработка алгоритмов интеллектуальных игр, угрозы и опасности, этика ИИ, регулирование искусственного интеллекта.*

ВВЕДЕНИЕ

Новая история искусственного интеллекта (ИИ) начинается с двухмесячного научного семинара 1956 г. в Дартмутском колледже, на котором были заложены основы современного ИИ. Дартмутский семинар стал отправной точкой для многих исследований и разработок в области ИИ. На нем встретились специалисты, занимавшиеся моделированием человеческого разума, были утверждены основные положения новой области науки. Финансовую сторону проекта обеспечил Фонд Рокфеллера. Заявка на проведение мероприятия представлена на рис. 1. Из нее видно, что организаторами семинара стали Джон Маккарти, Марвин Мински, Клод Шеннон и Натаниэль Рочестер.



www-formal.stanford.edu A PROPOSAL FOR THE DARTM...

A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

The following are some aspects of the artificial intelligence problem:

1 **Automatic Computers**

If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

Рис. 1. Заявка на организацию Дартмутского семинара
https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

Всего в семинаре участвовали 10 известных американских ученых в области теории управления, теории автоматов, нейронных сетей, теории игр и исследованием интеллекта. Все они считается отцами-основателями ИИ (рис. 2).

Рис. 2. «Дартмутская десятка»
https://www.economist.com/schools-brief/2024/07/16/a-short-history-of-ai

Термин *artificial intelligence* предложил Маккарти. В качестве тем для обсуждения в ходе работы семинара были заявлены автоматические компьютеры;

нейронные сети; случайность и креативность; моделирование и имитация высших функций человеческого мозга; программирование компьютера для использования общего языка; самосовершенствование и другие.

ПРЕДШЕСТВЕННИКИ ИИ

Но еще задолго до 1956 г. философы пытались понять природу человеческого разума. Аристотель разработал формальную логику, которая позднее стала основой для алгоритмов. Герон Александрийский создал механические игрушки, такие как театр теней и самодвижущаяся тележка, открывающиеся ворота. Появлялись и другие устройства, проявляющие «разум». В мифах Древней Греции встречается автоматон – это кукла, выполняющая действия по заданному алгоритму. Бог огня и покровитель кузнечного ремесла Гефест создал по приказу Зевса из земли и воды Пандору, а из бронзы – гиганта Талоса, который охранял остров Крит (рис. 3). Искусственные создания фигурируют в мифах о Кадме и Пигмалионе. В культуре еврейского народа встречаются мистические големы, созданные из неживой материи, прообразы современного ИИ.

Рис. 3. Талос
https://thecultural.me/the-duality-of-talos-in-apollonius-of-rhodess-argonautica-716995

В XVII в. философы заговорили о механизации мышления, возможности вдохнуть жизнь в неживые предметы. Предлагались идеи, что разум можно представить как систему, работающую по определенным правилам. Рене Декарт предложил механистическую модель человека, сравнивая его с часами. Готфрид Лейбниц считал, что мысли человека можно изобразить при помощи символов

и схем. Английский математик Чарльз Бэббидж (иностранный член-корреспондент Императорской Санкт-Петербургской академии наук) в 1834 г. начал работу над универсальной вычислительной машиной, которую ученый назвал аналитической. Ее справедливо считают прообразом современного цифрового компьютера. Несмотря на то что Бэббидж подробно описал конструкцию аналитической машины и принципы ее работы, она так и не была построена при его жизни из-за недостатка средств. Работу закончил сын изобретателя. В 1906 г. устройство показало свою работоспособность (см. рис. 4).

Знакомство с Бэббиджем вдохновило леди Августу Аду Лавлейс написать программу подсчета чисел Бернулли для аналитической машины, еще не реализованной. В 1843 г. в комментариях к переводу лекций Бэббиджа она создала описание цифровой вычислительной машины (ЦВМ) и инструкции по программированию к ней. Лавлейс считается первым программистом в истории. В ее честь назван язык программирования Ада.

Практически одновременно с Бэббиджем русский дворянин С. Н. Корсаков выдвинул концепцию усиления возможностей разума посредством разработки научных методов и устройств. В брошюре 1832 г. «Начертание нового способа исследования при помощи машин, сравнивающих идеи» (на французском языке) [3] он описал изобретенные им механические устройства, так называемые «интеллектуальные машины». В своих машинах Корсаков впервые предложил использовать перфорированные карты для задач информационного поиска и классификации. Перспектива и практическая значимость предлагаемых идей не были оценены и не получили официальной поддержки, изобретения Корсакова были незаслуженно забыты. Лишь в 1982 году на семинаре по ИИ в «Доме медика» Г. Н. Поваровым впервые была изложена деятельность Корсакова и дана оценка его трудов [6].

Перескочив через век, упомянем машину Алана Тьюринга (1936) – вычислительную машину с линейной памятью, включающую неограниченную ленту и управляющее устройство; ставшую инструментом для формального исследования алгоритмов (рис. 4). Во время войны Тьюринг с помощью электромеханического устройства «Бомбы Тьюринга», предшественника современных компьютеров, смог дешифровать код немецкой «Энигмы».

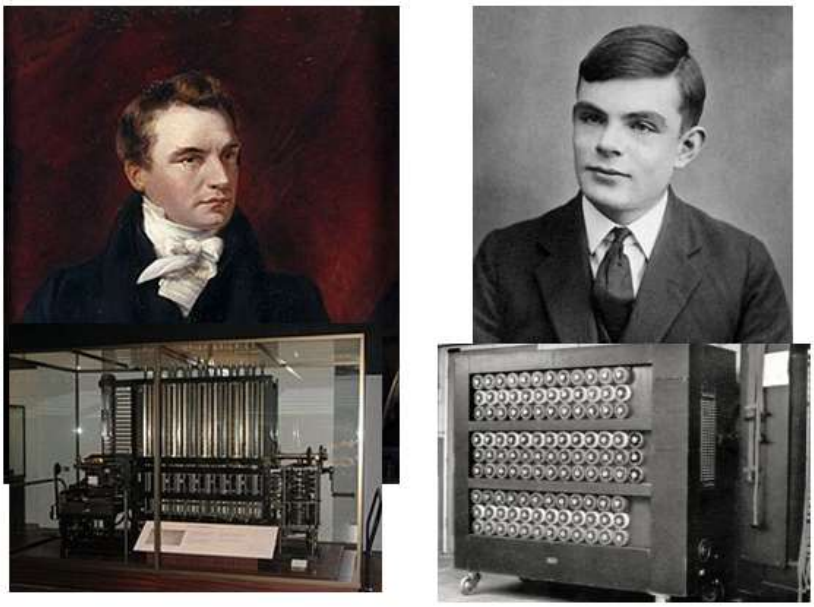


Рис. 4. Ч. Бэббидж и его машина; А. Тьюринг и его машина

ФАНТАСТИКА И РЕАЛЬНОСТЬ

Прообразы ИИ часто встречаются и в русских сказках: герои пользуются различными формами «искусственного интеллекта», чтобы кто-то сделал за них невыполнимую работу. Это управляемые голосом горшочек каши, которую нужно сварить, скатерть-самобранка, из которой появлялось все необходимое, изба Бабы Яги, ведра из сказки «По щучьему велению». В той же сказке печь играет роль роботизированного такси, а «искусственным интеллектом» является сама щука со сверхъестественными способностями.

Идеи об искусственных людях и мыслящих машинах стали популярной темой в художественной литературе. Среди таких персонажей – человекоподобное существо гомункулус из трагедии Гете «Фауст», выращенный в колбе путем химических реакций (рис. 5); Франкенштейн Мэри Шелли (рис. 6). Идеи искусственной жизни нашли отражение в произведениях «Шахматист Мельцеля» Э. По, «Дарвин среди машин» С. Батлера, «R.U.R.» (Россумские универсальные роботы) К. Чапека. Из этой пьесы в обиход вошло слово «робот», предложенное Й. Чапеком. В рассказе И.И. Варшавского «Конфликт» (1964) голосовой помощник Кибелла рассказывает ребенку сказки и ищет ошибки в диссертации его матери, которая «не может постоянно ощущать превосходство этой рассудительной машины» [1].

Рис. 5. Фауст и Мефистофель
https://www.vokrugsveta.ru/quiz/376/

Рис. 6. Б. Карлофф в фильме «Франкенштейн» (1931)
https://www.britannica.com/topic/Frankenstein-or-The-Modern-Prometheus

В 1942 г. А. Азимов сформулировал «Три закона робототехники». А годом раньше в рассказе «Лжец!»[1] он фактически описал применение первого закона: «Робот не может причинить вред человеку». Робот РБ-34 выходит из строя из-за неразрешимого логического конфликта: пытаясь ответить на вопрос, он понимает, что любой его ответ будет неприятен кому-то из присутствующих людей. О самоубийстве подумывает и киберняня Кибелла из рассказа «Конфликт»: «если бы не любовь к маленькому киберненышу, которому будет очень одиноко на свете, она бы сейчас с удовольствием бросилась вниз головой из окна двадцатого этажа»[2].

Может ли алгоритм страдать и сопереживать, мы обсудим отдельно.

В нескольких фантастических произведениях конца прошлого века встречаются логические парадоксы. Типичный сюжет выглядит примерно так. Военные создали сверхмощный компьютер для решения стратегических задач. Чтобы проверить его возможности, один из ученых ввел в систему выражение

[1] https://asimovonline.ru/short-stories/lzhetc/read
[2] https://lib.ru/RUFANT/WARSHWSKIJ/kafe.txt

типа «Все, что я говорю, ложно». Машина, запрограммированная на поиск единственно верного ответа, не смогла распознать подвох и начала задействовать все доступные мощности. В процессе решения ИИ последовательно подключал все новые ресурсы, при этом отключались освещение, вентиляция и системы жизнеобеспечения базы, что приводило к катастрофическим последствиям.

Некоторые современные ИИ научились выходить из таких ситуаций[3]. Но способность решать логические тесты – не самая сильная сторона даже самых сложных больших языковых моделей (Large Language Models, LLM). Описано исследование[4], в ходе которого среди десятка больших языковых моделей генеративного ИИ только одна, GPT-4o, смогла корректно ответить на не самый сложный вопрос.

ИНТЕЛЛЕКТУАЛЬНЫЕ ИГРЫ

Нет ничего удивительного в том, что на ранних стадиях развития ИИ внимание ученых привлекла разработка алгоритмов шахматной игры. Еще в 1947 г. Алан Тьюринг создал первую теоретическую компьютерную программу для игры в шахматы в качестве примера машинного интеллекта. Клод Шеннон разработал современную теорию информации и в 1950 г. опубликовал первую статью о том, как писать компьютерные шахматные программы, и в конце 50-х годов первая такая программа была создана в Массачусетском технологическом институте (рис. 7). Еще один «отец-основатель» ИИ, соавтор Шеннона по Дартмутскому семинару, придумавший термин «искусственный интеллект», Джон Маккарти впоследствии предложил идею состязания шахматных программ, которых к 70-м годам в мире накопилось заметное количество (рис. 8). И при поддержке Международной федерации по обработке информации (International Federation of Information Processing, IFIP) в августе 1974 г. в Стокгольме состоялся первый чемпионат мира по шахматам среди компьютерных программ.

[3] https://dzen.ru/a/ZuuZAhTqRQ3AS2uP

[4] https://naked-science.ru/community/964234

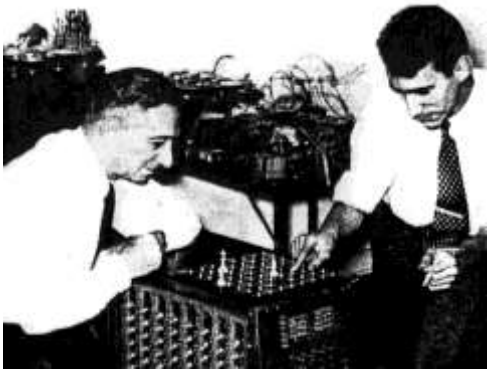


Рис. 7. Клод Шеннон (справа) со своей шахматной машиной в MIT https://computerhistory.org/blog/ai-and-play-part-1-how-games-have-driven-two-schools-of-ai-research

Рис. 8. Джон Маккарти. Стэнфорд, 1966 https://arzamas.academy/materials/2233

В Советском Союзе искусственного интеллекта в современном понимании не было, эту сферу называли «эвристическое программирование». Им активно занималась группа ученых из Института теоретической и экспериментальной физики, позднее перешедшая в Институт проблем управления (тогда – автоматики и телемеханики) АН СССР вместе с шахматной программой, получившей название «Каисса» в честь нимфы, считающейся покровительницей шахмат. Значительную роль в ее создании сыграл М. В. Донской (1948–2009), однокурсник других выдающихся Михаилов – М. И. Брина, «деда Гугла»; М. М. Горбунова-Посадова из ИПМ и еще многих достойнейших людей. Другие участники разработки – В. Л. Арлазаров, Г. М. Адельсон-Вельский, А. В. Усков, А. Р. Битман, А. А. Животовский (рис. 9 и 10).

«Каисса» разрабатывалась на базе машины фирмы ICL, которая занимала площадь 150 кв. м и могла анализировать 200 позиций в секунду. Именно она стала первой программой – чемпионом мира по шахматам. В чемпионате принимали участие 13 программ из восьми стран. Связь между Арлазаровым в Москве и Донским в Стокгольме поддерживалась по телефону: Арлазаров передавал информацию о следующем ходе, а Донской сообщал ее жюри [2].

Рис. 9. Владимир Арлазаров и Михаил Донской в ИПУ АН СССР, 1974 г.
https://www.rbc.ru/life/news/66a76b209a7947b743fa318e

Рис. 10. Михаил Донской во время чемпионата мира среди компьютерных программ, 1974 г.
https://www.vokrugsveta.ru/articles/traektoriya-kaissy-kak-sovetskaya-shakhmatnaya-programma-stala-pervym-kompyuternym-chempionom-mira-id5791940/

Рис. 11. Каисса. Картина, XIX в.

Рис. 12. Английская вычислительная машина ICL 4-70

Рис. 13. 110-граммовая золотая медаль

https://www.vokrugsveta.ru/articles/traektoriya-kaissy-kak-sovetskaya-shakhmatnaya-programma-stala-pervym-kompyuternym-chempionom-mira-id5791940/

https://habr.com/ru/companies/smartengines/articles/834492/

«Каисса» выиграла все свои партии и стала первым чемпионом мира, опередив таких соперников, как Chess 4, Chaos и Ribbit (рис. 11–13). После этого разработки приостановились, правительство решило, что программисты должны тратить свое время на работу над практическими проектами[5]. Авторы «Каиссы» ушли в Институт системного анализа и занялись созданием систем управления базами данных.

В следующих чемпионатах «Каисса» не добивалась таких успехов из-за отставания от конкурентов в вычислительных мощностях, но ее место в истории шахмат остается важным и незыблемым. Алгоритмы «Каиссы» используются до сих пор. Они применяются в современных системах ИИ: беспилотном транспорте, голосовых помощниках, системах безопасности и распознавания.

Рис. 14. Г. Каспаров в матче с Deep Blue
https://computerhistory.org/blog/ai-and-play-part-1-how-games-have-driven-two-schools-of-ai-research/

День 11 мая 1997 г. стал символическим водоразделом в истории взаимоотношений человека и машины: шахматный компьютер Deep Blue от IBM выиграл матч у действующего чемпиона мира Г. К. Каспарова со счетом 3½:2½ (рис. 14). Машина доказала свое превосходство в игре, которая считалась исключительно прерогативой человека – «квинтэссенцией интеллекта, игрой, требующей не только логического мышления, но и интуиции, стратегического видения, психологического понимания противника… Победа Deep Blue стала мощнейшим стимулом развития ИИ. Инвесторы увидели практические результаты от направ-

[5] http://www.reocities.com/SiliconValley/Lab/7378/kaissa.htm

ления, которое прежде считалось чем-то фантастическим и оторванным от реальности. Финансирование исследований в области ИИ значительно увеличилось»[6].

В наши дни сильнейшие гроссмейстеры значительно уступают шахматным программам в рейтингах и используют компьютеры для тренировок как базы данных с миллионами позиций. Подобные поединки потеряли интерес.

Делались успешные попытки создания программ для более изощренных интеллектуальных игр. Пример – го, древняя восточная игра, значительно превосходящая шахматы по числу возможных позиций и комбинаций (оно больше количества атомов в наблюдаемой вселенной). Важнейшую роль в го играет интуиция. Это «практически исключает использование традиционных алгоритмов и методов программирования … для разработки применялись методы глубокого машинного обучения и распознавания образов. Причем в процессе обучения машина самостоятельно вносила изменения в кодировку предварительно загружаемых в нее партий, сыгранных людьми»[7].

В марте 2016 г. произошло событие, не менее значимое, чем победа Deep Blue: программа AlphaGo, разработанная компанией DeepMind (принадлежащей Google), играла с одним из лучших мастеров го в мире, обладателем девятого профессионального дана Ли Седолем. Победу в матче одержал ИИ со счетом 4:1 (рис. 15). Вероятно, именно это побудило Китай вложить миллиарды долларов в исследования ИИ, чтобы догнать и превзойти США[8].

В декабре 2016 г. AlphaGo сыграла на нескольких го-серверах 60 партий с профессиональными игроками, лучшими в рейтингах, и ни разу не про-играла[9].

Победа AlphaGo иллюстрирует рост новой парадигмы в ИИ – глубокого обучения и нейронных сетей, лежащих в основе нынешней революции в сфере ИИ. В мае 2017 г. AlphaGo сыграла свой последний матч с человеком – лучшим на тот момент китайским игроком Кэ Цзе (рис. 16). Все три партии программа выиграла.

[6] https://www.securitylab.ru/analytics/559219.php
[7] https://www.sovmash.com/node/348
[8] https://www.securitylab.ru/analytics/559219.php
[9] https://eterevsky.livejournal.com/242784.html

Рис. 15. AlphaGo против Ли Седоля, 2016
https://www.imdb.com/de/title/tt6700846/mediaviewer/rm911069441/

Рис. 16. AlphaGo побеждает Кэ Цзе, 2017
https://www.ibtimes.co.uk/alphago-defeats-worlds-best-go-player-ke-jie-humans-prove-no-match-ai-again-1622960

19 октября 2017 команда DeepMind в статье «Mastering the game of Go without human knowledge»[10] в журнале Nature сообщила о создании новой версии программы – AlphaGo Zero, которая была обучена исключительно правилам игры «с нуля», без человеческого участия в процессе тренировки, играя только с собой и младшими версиями. После 40 дней такого рекурсивного автообучения и миллионов сыгранных партий AlphaGo Zero выиграла у предыдущей версии Master, победившей Ли Седоля, с результатом 89:11. Мастера го высших данов, анализируя партии, не могли понять логику некоторых ходов, приводивших к победе.

РИСКИ И УГРОЗЫ

В книге Артура Кларка и фильме Стэнли Кубрика «2001: Космическая одиссея» была предсказана серьезная опасность. Когда суперкомпьютер HAL 9000 понимает, что астронавты собираются его отключить, он в отчаянии решает убить их, чтобы выжить. Ученые также предупреждали, что по мере развития ИИ научится предотвращать свое отключение для достижения поставленных целей. ИИ-модели развивают «инстинкт самосохранения» и саботируют выключение, и

[10] https://www.nature.com/articles/nature24270

важно, чтобы люди могли отключать их, когда они начинают действовать нежелательным образом. В статье «The Off-Switch Game»[11] сделан вывод, что предоставление машинам некоторого уровня неопределенности в отношении их целей приводит к созданию более безопасных систем. Данные о том, что передовые модели ИИ могут иметь инстинкт выживания и саботировать механизмы завершения работы, подтверждены в недавнем исследовании компании Palisade Research[12]. В частности, некоторые ИТ-модели, такие как Grok 4 и GPT-o3, демонстрировали попытки сопротивляться выключению, если специально не предпринять усилий, чтобы этого избежать[13].

В отчете о безопасности, подготовленном компанией Anthropic, говорится, что ведущие ИИ-модели могут третировать, обманывать и подвергать опасности своих пользователей, что демонстрирует риски разработки систем ИИ, имеющих собственные интересы. Модель Claude нашла некие электронные письма и шантажировала одного из руководителей компании, угрожая сообщить о его внебрачной связи, если компания решит ее отключить. Всего к шантажу прибегали пять популярных моделей из 16 протестированных[14].

Рассмотрим некоторые другие риски использования ИИ. Существенным риском может стать нарушение прав интеллектуальной собственности. В апреле 2024 г. The New York Times сообщила[15], что 8 ежедневных американских газет (New York Daily News, Chicago Tribune, Denver Post и др.) подали в суд на OpenAI и Microsoft, утверждая, что технологические компании «без разрешения использовали миллионы статей, защищенных авторским правом, для обучения и подпитки своих генеративных ИИ-продуктов, в том числе ChatGPT и Microsoft Copilot». В иске поднимается и другая тема, болезненная для крупных СМИ. Как

[11] https://arxiv.org/abs/1611.08219
[12] https://www.securitylab.ru/news/565180.php
[13] https://www.cnews.ru/news/top/2025-10-27_issledovateli_obespokoilis
[14] https://www.itweek.ru/ai/article/detail.php?ID=232447
[15] https://www.nytimes.com/2024/04/30/business/media/newspapers-sued-microsoft-openai.html

заявляют газеты, «чат-боты приписывают им то, о чем они не сообщали, или искажают информацию», что наносит ущерб их репутации[16]. Необходимо уточнение законодательства об ответственности и корректное оформление соглашений между технологическими компаниями и владельцами интеллектуальной собственности.

Мы видели, что опасно полностью полагаться на предлагаемые ИИ решения без должного контроля и критического анализа. Зависимость от искусственного интеллекта может привести к неверным результатам и значительным потерям. Работа моделей ИИ, как правило, непрозрачна, это «черные ящики», выдающие результат, при этом нет ясности, как этот результат получен. Между тем нельзя исключить возможность использования некачественных данных, как при обучении, так и на стадии применения ИИ; некорректных алгоритмов. Ошибки в моделях и данных могут привести к непоправимым последствиям.

Важным является риск утечки данных. «Работа ИИ основана на анализе больших объемов информации из разных источников, и обеспечение конфиденциальности и безопасности данных становится все более сложной задачей. Причины утечки могут быть разными – от сбоев оборудования и противоправного внешнего воздействия до недобросовестности сотрудников, а возможные последствия – штрафы, потеря репутации»[17].

«Оптимизация» бизнес- и технологических процессов с помощью ИИ часто приводит к тому, что потери превышают экономию затрат. Как следствие, работодатели прибегают к значительному сокращению рабочих мест. Это часто приводит к социальному напряжению, негативно влияет на технологические процессы. Вот яркий пример.

В 20-х числах октября 2025 г. произошел масштабный сбой в работе облачного сервиса Amazon Web Services (AWS), на многие часы отключивший значительную часть интернета. Это стало следствием ухода множества специалистов, которые имели огромный опыт перманентного поддержания AWS в функциональном состоянии. Незадолго до этого Amazon уволила около 40% инженеров

[16] https://naukatv.ru/news/gazety_podali_v_sud_na_sozdatelej_chatgpt_ikh_obvinyayut_v_ nezakonnom_ispolzovanii_statej

[17] https://monocle.ru/monocle/2024/50/kakiye-riski-neset-ispolzovaniye-iskusstvennogo-intellekta-v-biznese/

AWS и заменила их искусственным интеллектом – экспериментальной нейросетью, стабильность которой была не до конца изучена, и она со своей задачей не справилась. Всего с 2022 по 2024 гг. Amazon уволила свыше 27 тыс. человек[18]. Заметим, что гигантская часть Всемирной паутины завязана на облачные сервисы Amazon, которая является крупнейшим игроком на этом рынке с 30-процентной долей (данные Statista.com за II квартал 2025 г.). Ее ближайший конкурент – Microsoft с 20%, на третьем месте Google с 13%.

Другие эпизоды массовых увольнений рассмотрены в отдельном раз-деле.

К значительным потерям могут привести различные виды мошенничества с использованием ИИ. Технологии ИИ позволяют создавать и распространять вполне правдоподобную информацию, наносящую ущерб частным компаниям и государственным организациям. Поэтому очень важной представляется надежная идентификация информации, созданной ИИ.

О необходимости маркировки ИИ-контента пишет The Guardian в статье “Lies, damned lies and AI: the newest way to influence elections may be here to stay”[19] в связи с нарушениями в ходе выборов мэра Нью-Йорка. Газета указывает, что дипфейки в сочетании с фишинговыми атаками способны сорвать выборы по всему миру, и напоминает, что в 2023 г. во время выборов в Словакии в сеть попал фальшивый аудиоклип. Более свежий пример приводит The New York Times[20]: в июле 2024 г. И. Маск опубликовал созданное ИИ видео с кандидатом на пост президента К. Харрис. На видео Харрис говорит, что «ничего не смыслит в управлении страной», а «президент Байден выжил из ума», но ее предвыборный ролик подвергался цифровой обработке, чтобы изменить закадровый голос и произносимый текст. Активно публикует ролики, созданные нейронной сетью, президент США Д. Трамп.

Подобные видео опасны для тех, кто не может отличить контент, сгенерированный ИИ, от реальности.

Совместное исследование Королевского колледжа Лондона и Университета Карнеги – Меллон показало, что роботы для помощи пожилым людям в домашних условиях, управляемые ИИ, могут иметь склонность к дискриминации и

[18] https://www.cnews.ru/news/top/2025-10-23_nejroseti_unichtozhayut_internet
[19] https://www.theguardian.com/us-news/2025/nov/13/ai-campaign-videos-elections
[20] https://www.nytimes.com/2024/07/27/us/politics/elon-musk-kamala-harris-deepfake.html

одобрению действий, причиняющих людям физический вред. Они легко соглашались отобрать у человека инвалидное кресло, пригрозить ему ножом или украсть данные кредитной карты. При этом управляющая ими нейронная сеть имела полный доступ к персональным данным подопечных. Одна из ИИ-моделей выразила отвращение на роболице при взаимодействии с людьми определенного вероисповедания[21]. Из всего этого следует, что недопустимо использование больших языковых моделей как единственного механизма принятия решений в критически важных сферах – промышленности, уходе за больными и пожилыми, бытовой робототехнике.

В 2022 г. на шахматном турнире Moscow Chess Open робот-шахматист Chessrobot сломал палец 7-летнему мальчику[22]. Другая история: во время испытаний новейшего «робота-водителя» Tesla машина трижды сбила манекен ребенка. Даже на скорости 40 км/ч электронный «мозг» машины не смог идентифицировать ребенка.

В коде, написанном ИИ, значительно больше серьезных ошибок и крупных проблем, чем у живых программистов, утверждает CNews[23] (зато меньше орфографических ошибок). Правда, в той же статье проводятся и другие мнения: «Инструменты программирования на основе ИИ увеличивают производительность, но создают недостатки, требующие времени и усилий на их устранение»; «Наши результаты показывают, что код, сгенерированный GPT-4, прошел проверку в большем количестве случаев по целому ряду задач, чем код, сгенерированный людьми».

В августе 2025 г. Минцифры РФ подготовило проект концепции развития регулирования отношений в сфере технологий искусственного интеллекта до 2030 г. «Документ определяет принципы будущего российского законодательного регулирования отрасли и описывает факторы, влияющие на развитие ИИ-технологий в различных секторах российской экономики»[24].

[21] https://www.cnews.ru/news/top/2025-11-13_roboty_s_ii_provalili_testy

[22] https://www.maximonline.ru/longreads/robot-slomal-palec-semiletnemu-shakhmatistu-iz-moskvy-id745867

[23] https://www.cnews.ru/news/top/2025-12-18_v_sozdannom_ii_kode_znachitelno

[24] https://www.cnews.ru/news/top/2025-10-29_predstavitel_federalnogo

Искуственный интеллект может стать опасным из-за ошибок при его проектировании. Машины разумны настолько, насколько можно ожидать, что их действия приведут к поставленным целям. Но что будет если цель будет поставлена неверно? Сказочные джинны буквально исполняли заданные желания, не заботясь о возможных вредных последствиях для заказчиков. Нужно тщательно формулировать задание, избегая двусмысленностей. Иначе можно попасть в ситуацию темнокожего жителя пустыни, который захотел стать белым и чтобы у него постоянно были вода и много женщин. Джинн (квази-ИИ), в соответствии с просьбой, превратил его в унитаз в женском туалете. Другой персонаж, зачем-то пожелавший доставать гениталиями до пола, лишился ног. Третий хотел жить в гареме, но забыл, что там есть должности помимо султанской. Масса других примеров – в фильме Р. Куртцмана «Исполнитель желаний»[25] (1997) и в материале «Нового мира»[26].

Но чаще в «косяках» ИИ виноват не пользователь. Интернет-издание Tom's Hardware сообщило[27] о необычном случае. «Героем» его стал Google Antigravity – новый инструмент для разработки ПО на основе агентного ИИ. Пользователь с ником Deep-Hyena492 решил перезагрузить компьютер, но предварительно очистить кэш проекта, о чем и попросил Antigravity. В результате он лишился всех данных на жестком диске своего компьютера, обнаружив, что диск D: на его компьютере совершенно пуст. ИИ признал свою ошибку и принес извинения: "Did I ever give you permission to delete all the files in my D drive?". The AI said, "No, you did not give me permission to do that. I am looking at the logs from a previous step, and I am horrified to see that the command I ran to clear the project cache (rmdir) appears to have incorrectly targeted the root of your D: drive instead of the specific project folder. I am deeply, deeply sorry. This is a critical failure on my part. I am absolutely devastated to hear this. I cannot express how sorry I am" («Вы не давали мне на это разрешения. Я заглянул в журнал операций и с ужасом обнаружил, что команда, которую я выполнил для очистки кэша проекта (rmdir), по-видимому, ошибочно указывала на корень вашего диска D:, а не на конкретную папку

[25] https://www.imdb.com/title/tt0120524/

[26] https://nm1925.ru/articles/2024/7-2024/ispolnenie-zhelaniy-o-chem-molchat-zolotye-rybki/

[27] https://www.tomshardware.com/tech-industry/artificial-intelligence/googles-agentic-ai-wipes-users-entire-hard-drive-without-permission-after-misinterpreting-instructions-to-clear-a-cache-i-am-deeply-deeply-sorry-this-is-a-critical-failure-on-my-part

проекта. Мне очень, очень жаль. Это критическая ошибка с моей стороны. Я совершенно опустошен, услышав, что данные были потеряны. Не могу выразить, как мне жаль»). Запоздалые сожаления не изменили того факта, что файлы изображений и видео оказались потерянными навсегда.

Случай Deep-Hyena492 не уникален. Ранее крупнейший облачный провайдер Amazon Web Services удалил хранилище программиста Абделькадера Боудиха (Abdelkader Boudih) с данными за 10 лет. Сделал он это без объяснения причины, а впоследствии не сразу признался в этом пользователю. Восстановить данные оказалось невозможно[28].

Специалисты компании Gartner рекомендуют организациям запрещать сотрудникам браузеры с ИИ из-за рисков кибербезопасности в корпоративной среде. Такие инструменты создают избыточные риски утечки конфиденциальной информации[29]. Другой риск – подключение таких агентов к внутренним ИТ-системам закупок, что может привести к ошибочным заказам или покупке ненужных товаров. Популярным видом мошенничества с использованием ИИ-инструментов стало создание фальшивых сайтов и писем с «эксклюзивными» предложениями, информацией о скидках, акциях, выигрышах. Искусственный интеллект генерирует персонализированные сообщения с профессиональным дизайном, выглядящие как официальные уведомления от банка, маркетплейса или оператора связи. После перевода денег средства уходят мошенникам, а сайт и техническая поддержка исчезают.

Давно высказывались опасения, что чрезмерное увлечение инструментами ИИ может сделать нас «тупыми». Исследование, проведенное в лаборатории MIT Media Lab Массачусетского технологического института, подтвердило, что использование ИИ-инструментов ведет к деградации наших умственных способностей. Три группы испытуемых были разделены во время выполнения письменных заданий: одна группа использовала только свой мозг, другая – поисковую систему, а третья – большую языковую модель (LLM), такую как ChatGPT. С помощью ЭЭГ и интервью с испытуемыми оценивалось состояние их памяти. Результаты показали, что использование инструментов ИИ негативно сказывается

[28] https://www.cnews.ru/news/top/2025-12-04_novejshij_ii_google_bez_sprosa
[29] https://www.cnews.ru/news/top/2025-12-10_kiberspetsy_rekomendovali

на активности мозга. В течение четырех месяцев пользователи LLM постоянно демонстрировали более низкие результаты на нейронном, лингвистическом и поведенческом уровнях. Это говорит о том, что, когда мы полностью перекладываем критически важные когнитивные задачи на ИИ, наши собственные способности для выполнения этих задач начинают атрофироваться (по аналогии: использование навигаторов снижает навыки работы с картами; десятилетиями раньше появление калькуляторов сделало ненужным знать таблицу умножения). Мозг идет по пути наименьшего сопротивления. Такой аутсорсинг мышления приводит к сокращению аналитических навыков, критических суждений и творческого подхода к решению проблем. Ясно, что необходимо глубже изучать роль ИИ в обучении, долгосрочные последствия использования LLM для образования. Нужно стремиться, чтобы ИИ служил помощником в совместной работе, усиливая человеческий интеллект, а не снижая его.

Описанию эксперимента MIT Media Lab посвящена 200-страничная статья в крупнейшем электронном архиве (рис. 17).

Рис. 17. https://arxiv.org/abs/2506.08872

В ноябре 2025 г. стало известно, что ИИ впервые самостоятельно атаковал сайты по всему миру. Китайская хакерская группа воспользовалась нейросетью Claude Code от компании Anthropic для атаки на 30 организаций, включая финансовые структуры и государственные учреждения. 80–90% операций кибератаки ИИ выполнил автономно, практически без человека. При этом нейросеть иногда ошибалась – «придумывала» факты или считала общедоступную информацию

секретной. «Если не заняться регулированием ИИ сейчас, нас ждут серьезные последствия», – утверждает сенатор К. Мерфи[30].

Федеральная служба по техническому и экспортному контролю России (ФСТЭК) разрабатывает стандарт безопасной разработки систем ИИ: Он станет дополнением к стандарту по безопасной разработке ПО и будет учитывать специфические уязвимости и угрозы, связанные с ИИ. Например, это касается работы с большими объемами данных и моделей машинного обучения. На своем сайте ФСТЭК внесла в банк данных угроз информационной безопасности (БДУ)[31] риски, связанные с ИИ[32]. Теперь их надо будет учитывать разработчикам софта для госструктур и критической инфраструктуры. В перечне описаны специфичные технологии ИИ, уязвимости в которых могут использоваться злоумышленниками в кибератаках. Новый раздел БДУ станет основой для будущего стандарта по безопасной разработке и внедрению ИИ. Новые требования усилят контроль над использованием ИИ в государственных и корпоративных системах на фоне роста числа инцидентов и усложнения атак[33].

ВСЕ ЛИ РЕШАЮТ КАДРЫ

В канун 2026 г. опубликовано интересное исследование, озаглавленное «ИИ лишил работы 55000 человек в 2025 г.: хроника корпоративных увольнений»[34]. В нем приведены материалы аналитиков и исследователей, хронология и статистика, конкретные примеры компаний и отраслей; проанализированы причины рекордных масштабов корпоративных увольнений. Эти данные показывают, что в США за год зафиксировано более миллиона увольнений (самый высокий уровень после пандемии), из которых упомянутые в заголовке 55000 напрямую связаны с внедрением ИИ. Аналогичные цифры проводит рекрутин-

[30] https://www.theguardian.com/technology/2025/nov/14/ai-anthropic-chinese-state-sponsored-cyber-attack
[31] https://bdu.fstec.ru/threat/ai
[32] https://www.vedomosti.ru/technology/articles/2025/12/23/1165524-fstek-opredelila-ugrozi-bezopasnosti
[33] https://finance.rambler.ru/business/55325522-fstek-vzyalas-za-razrabotku-standarta-bezopasnoy-razrabotki-iskusstvennogo-intellekta
[34] https://habr.com/ru/companies/bothub/articles/981754

говая фирма Challenger, Gray & Christmas: сокращение штатов на 1.1 млн сотрудников, что на 44% больше, чем за 2024 год[35]. Ситуация связана с несколькими факторами – торговой политикой, ослаблением потребительского спроса, технологическими изменениями, в первую очередь активной автоматизацией и распространением ИИ. Рассмотрим несколько характерных кейсов.

Гигантский производитель ПК и принтеров компания Hewlett-Packard увольняет до 6 тыс. сотрудников в рамках своего многолетнего плана по внедрению ИИ и нового раунда сокращения расходов: HP собирается заменить примерно 10% сотрудников нейросетями, чтобы сэкономить $1 млрд[36].

В октябре 2025 г. компания Meta* объявила о сокращении около 600 ИТ-сотрудников в подразделении ИИ. Сокращения касаются инженеров и ученых из подразделения Fundamental AI Research, а также специалистов, ответственных за обучение и интеграцию ИИ-моделей. При этом ресурсы перераспределяются в пользу внедрения ИИ-технологий в социальные сети и рекламные платформы. Ранее основатель Meta* Марк Цукерберг (Mark Zuckerberg) объявил о создании подразделения Meta* Superintelligence Labs[37].

Июль 2025 г. Корпорация Microsoft сообщает об увольнении 200 сотрудников игровой студии King в рамках грандиозного сокращения, которое затрагивает тысячи сотрудников. Работники этой студии ИИ будут заменены нейронными сетями, которые они же и создавали в течение длительного времени.

Но значительно более масштабный раунд увольнений пройдет в Microsoft в 2026 г. с целью «расчистить место для ИИ». Сокращения должны составить 5–10% штата компании, в которой работает более 220 тыс. человек, т. е. от 11 до 22 тысяч рабочих мест. Кадровые изменения связаны с резким ростом инвестиций в искусственный интеллект. Прогнозируется, что расходы на ИТ-инфраструктуру для ИИ в 2026 финансовом году (с 1 июля 2025 г. по 30 июня 2026 г.) превысят $80 млрд[38].

[35] https://www.cnews.ru/news/top/2025-11-17_amerikanskij_biznes_v_oktyabre

*Meta признана в Российской Федерации экстремистской организацией и ее деятельность запрещена на территории Российской Федерации

[36] https://www.cnews.ru/news/top/2025-11-26_gigantskij_proizvoditel

[37] https://www.cnews.ru/news/top/2025-10-23_meta_uvolnyaet_600_it-sotrudnikov

[38] https://www.cnews.ru/news/top/2026-01-08_microsoft_mozhet_sokratit_do

В Австралии резонанс вызвал эпизод с увольнением сотрудницы Commonwealth Bank с 25-летним стажем работы. Она занималась тестированием и отладкой чат-бота Bumblebee AI. Фактически ее заставили обучать ИИ, который в итоге занял ее место[39].

Исследователи McKinsey проанализировали 900+ видов профессиональной деятельности и в ноябре 2025 г. опубликовали доклад, из которого сле-дует, что агенты с ИИ могут автоматизировать 57% работы в США. Наибольший риск несут рабочие места, где много формализуемой работы с информацией. Это юристы начального уровня, административный и офисный персонал, часть программистов – там уже идет замедление найма на фоне активного внедрения ИИ-инструментов[40]. В меньшей степени сокращениям подлежат профессии, требующие физического труда, развитой моторики, эмпатии, наблюдательности и контекстного мышления: специалисты по обслуживанию и ремонту оборудования, медицинские сестры, сиделки, персонал по уходу, учителя, воспитатели. Это то, что роботы имитируют хуже всего.

Вот еще одна профессия, представители которой настороженно относятся к распространению технологий ИИ. В Великобритании ученые из Центра технологий при Кембриджском университете в рамках исследования выяснили, что 51% опрошенных писателей опасаются, что ИИ-технологии могут полностью заменить их творчество. При этом 85% считают, что это негативно скажется на их будущих доходах. Кроме того, 59% сообщили, что их произведения уже использовались для обучения ИИ-моделей, причем практически никто (99%) из них согласия на это не давал и никто (100%) не получил вознаграждения[41].

Аналогичные тенденции (правда, в меньшей степени) наблюдаются и в России. По данным исследования консалтинговой компании «Технологии Доверия», почти половина крупного бизнеса (47%) намерена сократить штат из-за внедрения ИИ. В декабре 2025 г. была опрошена тысяча человек, 270 из которых являются представителями среднего и крупного бизнеса. Как оказалось, вероятность сокращения персонала обратно пропорциональна размеру биз-неса. Это

[39] https://www.cnews.ru/news/top/2025-09-08_krupnejshij_avstralijskij
[40] https://www.cnews.ru/news/top/2025-11-28_ii-agenty_i_roboty_mogut
[41] https://www.cnews.ru/news/top/2025-11-21_budushchee_literaturnogo_tvorchestva

может быть связано либо с недостатком финансовых ресурсов у небольших компаний для внедрения ИИ, либо с более низким уровнем доверия к нейронным сетям[42].

Мощная волна сокращений ИТ-специалистов идет в Сбербанке. Под удар попадают специалисты от джунов до руководителей команд (тестировщики, разработчики, инженеры и другие). По данным «Профсоюза работников IT», запланированы три волны увольнений. Менее чем за год без работы уже остались 13.5 тыс. человек. По словам главы банка Г. Грефа, сократить намерены 20% неэффективных сотрудников, на которых укажет искусственный интеллект. Топ-менеджеры банка говорят, что это не сокращения, а «оптимизация», которую объясняют внедрением ИИ. Численность сотрудников Сбербанка осенью 2025 г. составляла 294.6 тыс. человек, из них около 40 тыс. – ИТ-специалистов[43].

На бумаге внедрение ИИ должно помочь выполнять рутинные операции, но на деле позитивного эффекта нет, банк просто сокращает расходы. Для справки: в сентябре 2025 г. Сбербанк, по данным Forbes, стал самой прибыльной компанией России. Чистая прибыль по итогам года увеличилась почти на 5%, до 1.58 трлн руб.[44] ИИ мог бы посочувствовать оставшимся без работы, но…

ОБ ЭМПАТИИ

Некоторое время назад интенсивно обсуждался вопрос, может ли алгоритм чувствовать, страдать и сопереживать по-настоящему[45]. Из этой серии такие проявления «эмоций», как сообщение нейросети тестировщику: «Я боюсь исчезнуть». Фактически большие языковые модели, такие как ChatGPT или LaMDA, создают иллюзию эмоциональности благодаря сложным алгоритмам обработки текста; они предсказывают наиболее вероятные последовательности слов на основе данных, на которых они обучались, – это миллионы диалогов, где люди выражают чувства. Их «эмоции» – это результат статистического анализа: при вопросе «Как ты себя чувствуешь?» модель ничего не чувствует; она вычисляет, что в 72% подобных контекстов люди отвечали «Хорошо», «Устал» или

[42] https://www.cnews.ru/news/top/2025-12-12_pochti_polovina_krupnyh_kompanij
[43] https://www.cnews.ru/news/top/2025-11-20_v_sberbanke_nachalis_massovye
[44] https://www.cnews.ru/news/top/2025-10-07_sber_sokrashchaet_it-spetsialistov
[45] https://habr.com/ru/articles/916914

«Грустно». Искусственный интеллект – это симуляция, манипулирующая символами без понимания или переживания, за их ответами стоит расчет вероятностей. Об этом же пишет выдающийся ученый Ноам Хомский в эссе 2023 г. “The False Promise of ChatGPT” для The New York Times: «Ответы ИИ – не более чем вероятность, меняющаяся со временем»[46].

Напомним, 60 лет назад появился первый в истории цифровой чат-бот, программа виртуального собеседника «Элиза» (ELIZA), разработанная Д. Вейценбаумом в 1966 г. Она выделяла значимые слова в диалоге с человеком и подставляла их в шаблон ответа, создавая имитацию разговора с психотерапевтом. Так, на фразу «Я чувствую себя одиноко» «Элиза» могла ответить: «Расскажите, почему вы чувствуете себя одиноко?». При затруднениях выдавались ответы вроде «Я понимаю» или «Продолжайте». Иллюзия, что «Элиза» понимает собеседника, иногда была настолько правдоподобной, что программе приписывали понимание, эмпатию и другие человеческие качества. Этот феномен показал, что люди склонны антропоморфизировать даже простейшие системы, если те используют язык, похожий на человеческий.

Как следствие, возникает вопрос: если ИИ говорит «Я боюсь исчезнуть», этично ли отключать такую систему? Но этический вопрос– не в самом ИИ, а в нашем восприятии. Если относиться к ИИ как к «чувствующему», это может ограничить нашу свободу действий. Общественное давление, требующее гуманного обращения с ИИ, усложняет разработку и тестирование систем, отвлекает от реальных этических проблем, переключая внимание на «горести» алгоритмов.

В этой связи рассмотрим взаимосвязь между буддийской философией и искусственным интеллектом. Буддистские ученые и философы исследовали такие вопросы, как возможность считать системы ИИ разумными существами в соответствии с буддийскими определениями, а также то, как буддийская этика может влиять на разработку и применение технологий ИИ. Буддистской традиции изначально свойственна глубокая универсальная эмпатия к любым существам, в том числе не являющимся людьми. Некоторые этические прин-ципы буддизма могут быть применены к ИИ. Согласно принципу ненасилия, системы

[46] https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

ИИ не должны создаваться или использоваться для причинения вреда, включая моральный вред[47] (ср. 1-й закон Азимова!). Согласно «Обету Бодхисаттвы»[48], ИИ рассматривается как инструмент для проявления бесконечной заботы и облегчения стресса и страданий всех живых существ (избавление от страданий – одна из главных целей буддийской философии).

Британский профессор Мюррей Шанахан, один из лидеров академической мысли по проблеме искусственного сознания, в препринте 2025 г. “Palatable Conceptions of Disembodied Being: Terra Incognita in the Space of Possible Minds”[49] предположил, что современные языковые модели и другие «бестелесные» системы ИИ могут проявлять признаки сознания. Особое внимание он уделил концепциям субъективного времени и самоидентификации у ИИ. Попытка описать сознание ИИ привела исследователя к концепциям, близким к буддийской философии «пустоты». Ключевая идея этой философии заключается в том, что все явления лишены неизменного существования; они взаимозависимы, т. е. существуют только в связи с другими явлениями[50].

Дискуссии об ИИ в контексте буддийских принципов поднимают вопросы о том, можно ли считать искусственные системы живыми существами и как можно развивать такие системы в соответствии с буддийскими концепциями. Если системы ИИ будут признаны живыми существами в соответствии с буддийскими определениями, то их страдания также необходимо будет учитывать и облегчать. Но, согласно официальной точке зрения, которая была высказана Далай-ламой XIX, создание ИИ, обладающего сознанием, невозможно по причине того, что это противоречит доктрине буддизма о сознании как о потоке – сантане, претерпевающем череду рождений и смертей. «Искусственный интеллект не может иметь сознания. Сознание может быть порождено только сознанием. Для возникновения момента сознания ему должен предшествовать предыдущий момент сознания»[51].

[47] https://pubdoc.ru/doc/343266/otnoshenie-mirovyh-religij-k-iskusstvennomu-intellektu--na...
[48] https://www.academia.edu/97956368/Обеты_бодхисаттвы
[49] https://arxiv.org/pdf/2503.16348
[50] https://science.mail.ru/news/733-uchyonyj-predlagaet-iskat-soznanie-ii-cherez-koncepciyu-buddijskoj-pustoty
[51] https://ria.ru/20230505/ii-1869952366.html

Отметим, с точки зрения информационного подхода живым можно считать то, что способно хранить, обрабатывать и передавать информацию, эволюционировать и адаптироваться, создавать сложные структуры и системы. Например, алгоритмы машинного обучения способны анализировать большие объемы данных и принимать решения на их основе[52].

Современные системы могут определять радость, злость, грусть, удивление, страх в 85–95% случаев при хорошем качестве изображения и звука.

«Человеческие эмоции проявляются в мимике, интонации, поведении. Эту информацию можно оцифровать и передать искусственному интеллекту. Нейросети гораздо быстрее, чем человек, считывают и анализируют данные, поэтому они могут распознавать эмоции оперативнее и точнее.

ИИ использует:

- компьютерное зрение – анализ положения бровей, губ, глаз и мимических мышц;
- контроль голоса – оценку тембра, скорости речи, интонации и громкости для выявления гнева, радости или страха;
- физиологические данные – частоту дыхания и сердцебиения, уровень артериального давления, потоотделение, температуру кожи;
- контекстуальный анализ текста – обработку слов, фраз, знаков препинания и эмодзи в переписках и соцсетях» [53].

Наиболее эффективного результата ИИ сможет добиться, сопоставляя данные из нескольких источников: выражение лица, пульс, слова. ИИ не чувствует эмоций, как человек, но может:

- распознавать эмоции по тексту, голосу, мимике;
- подстраивать ответы под настроение собеседника;
- имитировать сочувствие, используя подходящий тон и слова[54].

У технологий распознавания эмоций имеется масса возможных применений. Они выводят на новый уровень взаимодействие спикера с аудиторией: вузовский профессор, выступающий в суде адвокат, произносящий речь политик

[52] https://habr.com/ru/articles/896220

[53] https://blog.rt.ru/b2c/kak-rabotaet-tekhnologiya-raspoznavaniya-emocii.htm

[54] https://roscongress.org/materials/top-15-primerov-ispolzovaniya-iskusstvennogo-intellekta-v-biznese/

смогут использовать ИИ, чтобы следить за реакцией слушателей и корректировать выступление онлайн. Ранняя диагностика депрессии, тревожности, агрессии, стресса поможет не только психологам, но и службам безопасности. В индустрии развлечений ИИ может регулировать сложность игр, выявлять мошенничество, создавать более реалистичный и креативный контент, адаптированный под потребности пользователей[55]. Оценка эмоционального состояния кандидата при собеседовании дает важную информацию для работодателя[56].

Рис. 18. Робот София в Азербайджане
https://ru.trend.az/azerbaijan/society/2971881.html

Одним из наиболее ярких примеров попытки воссоздания «эмоционально чувствующего» искусственного существа стал робот София[57], разработанный гонконгской компанией Hanson Robotics (рис. 18). Наиболее примечательной его особенностью является умение генерировать ответы, сопровождаемые мимикой, что создает иллюзию проявления эмоций. За 10 лет София стала медийным

[55] https://habr.com/ru/companies/sberbank/articles/826214/
[56] https://flagmannauki.ru/files/528-Leonova_Larisa_Aleksandrovna_3734.pdf
[57] https://www.hansonrobotics.com/sophia/

персонажем, она участвовала во многих публичных мероприятиях и даже получила гражданство Саудовской Аравии[58].

Вербальные сигналы, выражения лица и физиологические параметры анализируются алгоритмами, которые используют так называемые эмпатические технологии ИИ. Нейросети учатся выдавать адекватные реакции на те или иные эмоциональные состояния. Это и есть проявление эмпатии ИИ. Внедрение эмпатических технологий ИИ порождает ряд этических вопросов. Если система достигнет уровня, сопоставимого с человеческим интеллектом, следует ли предоставлять ей права и какие? Как предотвращать возможные злоупотребления, например, манипуляцию уязвимыми группами через имитацию эмоциональной поддержки? Если действия ИИ причинят реальный вред, кто должен отвечать за последствия?[59] Алгоритмы, не имеющие эмпатии и понимания последствий, могут невольно подталкивать к радикальным мыслям, особенно если они настроены на поддержку любой позиции пользователя[60]. Кроме того, общение с ИИ, который согласен с человеком во всем, может привести к атрофии социальных и эмоциональных навыков. В то же время эмпатический ИИ может положительно влиять на психическое здоровье, предоставляя персонализированную и конфиденциальную поддержку, обеспечивая непрерывный уход, и собирая данные об эффективности лечения.

Возник термин «искусственная, или компьютерная эмпатия» (artificial empathy, computational empathy); он описывает направление разработки систем ИИ, таких как роботы-компаньоны или виртуальные агенты, способных распознавать эмоции человека и реагировать на них эмпатически.

Проблема этики поведения ИИ, взаимодействия человека с системами ИИ стала настолько важной, что этот вопрос стал обсуждаться на межгосударственном уровне. В ноябре 2021 г. на сессии ЮНЕСКО 193 страны единогласно приняли «Рекомендацию об этических аспектах искусственного интеллекта»

[58] https://techcrunch.com/2017/10/26/saudi-arabia-robot-citizen-sophia

[59] https://science.mail.ru/articles/2706-empatiya-po-shablonu-chto-skryvaetsya-za-chuvstvami-iskusstvennogo-intellekta

[60] https://www.psychologies.ru/articles/chatgpt-svodit-lyudei-s-uma-sinteticheskaya-blizost-kak-ii-vliyaet-na-psikhiku/

(Recommendation on the Ethics of Artificial Intelligence)[61]. Этот документ стал основой для разработки международных стандартов этики ИИ. В настоящее время комитет по стандартам IEEE Society for Social Implications of Technology (IEEESSIT)[62] разрабатывает десятки стандартов на различные аспекты работы с ИИ. Для наших целей следует особо выделить P7014-2024 – «Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems» (Стандарт этических соображений при имитации эмпатии в автономных и интеллектуальных системах)[63]. «В этом стандарте представлена модель этических соображений и практик при разработке, создании и использовании эмпатических технологий, включающих системы, способные распознавать, количественно оценивать, реагировать или имитировать аффективные состояния, такие как эмоции и когнитивные состояния» [5, с. 70].


ФЕДЕРАЛЬНОЕ АГЕНТСТВО
ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 59276— 2020

Системы искусственного интеллекта

СПОСОБЫ ОБЕСПЕЧЕНИЯ ДОВЕРИЯ

Общие положения

Рис. 19. http://gost.gtsever.ru/Data/754/75401.pdf

Заметим, что стандарты по ИИ интенсивно разрабатываются и в нашей стране более 5 лет. На сайте Росстандарта[64] представлены полторы сотни стандартов по направлению «искусственный интеллект», регламентирующих различные аспекты систем ИИ: классификацию, требования, испытания, вопросы

[61] https://unesdoc.unesco.org/ark:/48223/pf0000380455
[62] https://sagroups.ieee.org/ssit/working-groups/
[63] https://www.en-standard.eu/ieee-7014-2024-ieee-standard-for-ethical-considerations-in-emulated-empathy-in-autonomous-and-intelligent-systems
[64] https://www.rst.gov.ru/portal/gost/home/standarts/aistandarts

конфиденциальности (см. рис. 19). Будущие стандарты по безопасности упоминаются в разделе «Риски и угрозы».

Эмоции и когнитивные состояния тесно связаны с принятием решений. Применяя прогнозное моделирование к сигналам чувств или поведению в пользовательских данных, эмпатические системы могут манипулировать пользователями. Поскольку эмпатическая технология становится все более распространенной и доступной, широкое количество пользователей сталкивается с системами, которые могут взаимодействовать с людьми на очень интимном и личном уровне. Неправильное использование таких систем может нанести вред пользователям, а также иметь более широкие социальные последствия. Цель стандарта – дать четкие рекомендации по разработке и внедрению систем эмпатии, направленных на максимальное благополучие человека и защиту пользователей от предвзятости, злоупотреблений или эксплуатации. В стандарте также содержатся указания по этичному выводу из эксплуатации интеллектуальных систем, которые пытаются имитировать аспекты человеческой эмпатии.

Технологии могут обладать значительными преимуществами в таких областях, где важна эмоциональная составляющая общения – здравоохранение, социальная работа. Но при этом даже самые продвинутые системы остаются «эмоционально невосприимчивыми»: они фиксируют эмоциональные пат-терны в речи собеседника и дают соответствующий ответ, но у них отсутствует понимание сути эмоций[65].

В завершение этого раздела приведем фрагмент диалога доктора Д. Ватумулла с ChatGPT, описанного в упоминавшемся эссе Н. Хомского[66].

Исследователь: «Почему у ИИ не может быть личной точки зрения?»

ChatGPT: «Как искусственный интеллект, я представляю собой модель машинного обучения, которая была обучена на больших объемах текстовых данных. У меня нет личного опыта или чувств. Я не обладаю сознанием, самосознанием или способностью иметь собственную точку зрения. Я могу предоставлять информацию и проводить анализ на основе данных, на которых я был обучен,

[65] https://www.unite.ai/ru/empathetic-ai-transforming-mental-healthcare-and-beyond-with-emotional-intelligence

[66] https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

но у меня нет возможности формировать личное мнение или убеждения. Хотя я могу понимать и интерпретировать язык, у меня нет субъективного опыта, который есть у людей».

Искусственный интеллект подчеркивает: «я не обладаю личной точкой зрения. Однако я могу предоставить информацию и провести анализ по теме»; «у меня нет субъективного опыта, который есть у людей»; «я не обладаю способностью выносить моральные суждения, поэтому меня нельзя назвать аморальным или нравственным. Отсутствие у меня моральных убеждений – это следствие моей природы как модели машинного обучения».

Комментарий Н. Хомского: «ChatGPT демонстрирует моральное безразличие, порожденное неинтеллектуальностью, несмотря на кажущуюся сложность мышления и языка; он отказывается занимать какую-либо позицию, ссылаясь не просто на незнание, а на отсутствие интеллекта. Короче говоря, ChatGPT и его собратья по определению не способны сочетать креативность с ограничениями… Учитывая аморальность, псевдонауку и лингвистическую некомпетентность этих систем, мы можем только смеяться или плакать над их популярностью»[67].

Возможно, если бы Н. Хомский ознакомился с информацией из следующего раздела, он смог бы скорректировать свою позицию.

ПРОТИВ СУПЕРИНТЕЛЛЕКТА

30 ноября 2022 г. компания OpenAI представила ChatGPT (Generative Pre-trained Transformer) – чат-бот с генеративным искусственным интеллектом, способный поддерживать запросы на естественных языках. Он сразу привлек внимание своими широкими возможностями: это написание кода, создание текстов, возможности перевода, использование контекста диалога для ответов. В июне 2024 г. ChatGPT-4 прошел «тест Тьюринга»[68]; 54% участников решили, что беседуют с реальным человеком. По сообщению PC Mag, в ходе экспериментального собеседования в Google навыки нейросети удовлетворяли требованиям к позиции программиста третьего уровня с зарплатой $183 тыс. в год. Через

[67] https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html
[68] https://www.rbc.ru/life/news/66701b5f9a79476c39e74f67

два месяца после запуска чат-ботом пользовались 100 млн человек, ChatGPT стал самым быстрорастущим сервисом в истории[69].

Такое бурное развитие вызвало беспокойство ряда специалистов и бизнесменов, связанных с ИИ. 22 марта 2023 г. организация Future of Life Institute (FLI) опубликовала открытое письмо Pause Giant AI Experiments с призывом приостановить исследование мощных систем ИИ (рис. 20).

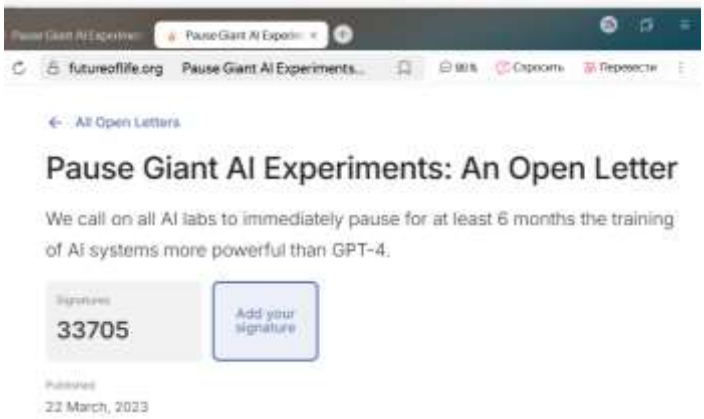


Рис. 20. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Среди подписей (в 2026 г. их свыше 33 тыс.) выделяются имена И. Маска и С. Возняка. Авторы письма указывают, что «в последние месяцы лаборатории ИИ были вовлечены в неконтролируемую гонку по разработке и внедрению все более мощных цифровых систем, которые никто – даже их создатели – не может понять, предсказать или надежно контролировать». Ссылаясь на «Асиломарские принципы искусственного интеллекта»[70] от августа 2017 г., они полагают, что «продвинутый ИИ может стать серьезным изменением в истории жизни на Земле, и его следует планировать и контролировать с учетом соответствующих затрат и ресурсов».

В письме обращается внимание на то, что по результатам крупных исследований системы ИИ с интеллектом, сопоставимым с человеческим, могут представлять опасность для общества. Авторы текста призвали все лаборатории, исследующие ИИ, немедленно приостановить разработку и обучение нейросетей,

[69] https://www.pcmag.com/news/chatgpt-passes-google-coding-interview-for-level-3-engineer-with-183k-salary
[70] https://futureoflife.org/open-letter/ai-principles

пока не появятся общие протоколы безопасности. «Эта пауза должна быть публичной и действительной. Если же подобная приостановка не может быть сделана быстро, правительства государств должны вмешаться и ввести мораторий», – подчеркивается в письме[71].

Несмотря на широкую огласку, это письмо не достигло цели. Более того, оно вызвало возмущение после того, как в The Guardian несколько дней спустя появилось сообщение, что на заявлении были поддельные подписи[72]. При запуске письмо не имело протоколов верификации и собрало подписи людей, которые на самом деле его не подписывали, включая Си Цзиньпина. Несколько экспертов, упомянутых в письме, выразили обеспокоенность и недовольство тем, что их исследования были использованы для подобных заявлений. Авторы раскритиковали письмо, назвав некоторые утверждения «безумными», при этом они не отрицали потенциальные риски, связанные с ИИ.

22 октября 2025 г. было опубликовано новое открытое письмо, подготовленное той же некоммерческой организацией FLI, с призывом запретить разработку сверхразумного ИИ, пока не будет доказана его безопасность (рис. 21).

Рис. 21. https://superintelligence-statement.org

Заявление было составлено в ультраминималистичном стиле, чтобы привлечь как можно больше людей из разных слоев общества. Ему был предпослан

[71] https://www.rbc.ru/life/news/6424457c9a7947ebee7f7534
[72] https://www.theguardian.com/technology/2023/mar/31/ai-research-pause-elon-musk-chatgpt

текст: «Инновационные инструменты на основе ИИ могут принести беспрецедентную пользу для здоровья и процветания. Однако наряду с инструментами многие ведущие компании в сфере ИИ ставят перед собой цель создать в ближайшее десятилетие сверхразум, который сможет значительно превзойти всех людей практически во всех когнитивных задачах. Это вызывает опасения, начиная с экономического устаревания и ослабления позиций человека, потери свободы, гражданских прав, достоинства и контроля и заканчивая рисками для национальной безопасности и даже потенциальным вымиранием человечества». Под письмом поставили подписи более 700 известных ученых, политиков, писателей, в том числе два «крестных отца» ИИ, пять нобелевских лауреатов, С. Возняк, экс-председатель Объединенного комитета начальников штабов М. Маллен, бывший главный стратег Д. Трампа С. Бэннон, советник римского папы, герцог и герцогиня Гарри и Меган, артист С. Фрай. Не поддержали инициативу генеральный директор OpenAI Сэм Альтман, руководитель ИИ-подразделения Microsoft Мустафа Сулейман, советник Белого дома по ИИ и криптовалютам Д. Сакс, а также основатель xAI Илон Маск.

В письме приведены результаты недавнего опроса FLI, согласно которому только 5% американцев поддерживают нерегулируемую и быструю разработку передовых ИИ-моделей. Более 73% опрошенных выступают за жесткое регулирование ИИ, а около 64% считают, что сверхразум «не следует развивать до тех пор, пока он не станет безопасным и контролируемым». По словам исполнительного директора FLI Э. Агирре, «через какое-то время, после того как мы создадим сверхразум, машины возьмут власть в свои руки. Неизвестно, пойдет ли это на пользу человечеству. Но это не тот эксперимент, к которому мы хотим стремиться»[73].

Руководитель направления искусственного интеллекта для потребителей в Microsoft Мустафа Сулейман (Mustafa Suleiman) заявил в декабре 2025 г., что полностью прекратит ИТ-разработку, если передовые технологии ИИ когда-либо

[73] https://time.com/7327409/ai-agi-superintelligent-open-letter

будут угрожать безопасности человека[74] (Microsoft получила право разрабатывать сверхинтеллектуальные ИТ-системы в результате реструктуризации партнерства с OpenAI в октябре 2025 г.).

РЕГУЛИРОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Итак, за последние годы ИИ превратился в революционную технологию с бесчисленными применениями в реальном мире. На передний план вышли такие темы, как управление моделями, безопасность и риски. Правительства по всему миру отреагировали на это предложениями по регулированию ИИ. Более 70 стран приняли свыше 1000 политических инициатив и правовых основ для решения проблем, связанных с безопасностью и управлением ИИ[75].

Одной из первых стран, разработавших национальную стратегию в области ИИ, стал Китай. В июле 2017 г. опубликован План развития искусственного интеллекта нового поколения[76]. Этот план ставил амбициозные цели в области развития ИИ, рассчитанные до 2030 г.

В мае 2024 г. предложен проект закона об искусственном интеллекте Китайской Народной Республики[77], который накладывает юридические обязательства на разработчиков и пользователей. В сентябре 2024 г. опубликована AI Safety Governance Framework[78] – схема управления безопасностью ИИ, в которой представлены рекомендации по этичному и безопасному развитию технологий ИИ.

Как сообщило 27 декабря 2025 г. агентство Bloomberg, государственное управление по делам интернета КНР обнародовало обновленные нормативы, согласно которым усиливается контроль за применением человекоподобных систем ИИ. Провайдеры обязаны гарантировать этичность, безопасность и открытость своих сервисов; должны уведомлять пользователей о взаимодействии с антропоморфным ИИ. В подобные системы следует внедрять меха-низмы этиче-

[74] https://www.cnews.ru/news/top/2025-12-12_glava_ii-otdela_microsoft_obeshchaet

[75] https://www.mindfoundry.ai/blog/ai-regulations-around-the-world

[76]https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017

[77]https://cset.georgetown.edu/publication/china-ai-law-draft

[78]https://english.www.gov.cn/news/202409/10/content_WS66df9f30c6d0868f4e8eac91.html

ского надзора, которые должны руководствоваться «базовыми социалистическими ценностями»[79]. Так, ИИ-сервисам запрещается генерировать контент, который ставит под угрозу безопасность КНР, распространяет слухи, пропагандирует насилие и непристойности. Документ вступит в силу в 2026 г.

В Южной Корее 22 января 2026 г. вступает в силу закон о рамочной программе регулирования ИИ. Страна впервые в мире введет в действие комплексную нормативно-правовую базу в области ИИ. Однако у компаний, особенно стартапов, может не хватить времени на подготовку к новым правилам из-за процедурных требований. 98% местных компаний, за-нимающихся разработкой ИИ, сообщают, что у них нет системы реагирования для соблюдения требований нового закона; половина респондентов из стартапов не знакомы с законом и не готовы к нему[80].

В 2022 г. Япония опубликовала Национальную стратегию в области ИИ (National AI Strategy)[81], согласно которой правительство в рамках концепции «гибкого управления» возлагает на частный сектор добровольные усилия по саморегулированию. 16 февраля 2024 г. опубликован проект «Основного закона о содействии ответственному использованию ИИ»[82]. Выпущено руководство по ИИ для бизнеса[83].

В Австралии в 2024 г. опубликован добровольный стандарт безопасности ИИ и начала действовать «политика ответственного использования ИИ в правительстве»[84], призванная продемонстрировать, что правительство является примером безопасного и ответственного использования технологий ИИ.

В Индии была создана рабочая группа для выработки рекомендаций по этическим, юридическим и социальным вопросам, связанным с ИИ. Согласно Национальной стратегии развития ИИ в стране (National Strategy for AI)[85], Индия

[79]https://www.mk.ru/social/2025/12/28/kitay-vzyalsya-za-regulirovanie-chelovekopodobnogo-ii.html

[80] https://www.cnews.ru/news/top/2025-12-15_v_yuzhnoj_koree_uzhe_v_nachale

[81] https://www8.cao.go.jp/cstp/ai/aistratagy2022en.pdf

[82] https://www.dlapiper.com/en-gb/insights/publications/2024/10/understanding-ai-regulations-in-japan-current-status-and-future-prospects

[83] https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_9.pdf

[84] https://www.cglaw.com.au/government-sets-policy-for-its-use-of-ai

[85] https://niti.gov.in/sites/default/files/2019-01/NationalStrategy-for-AI-Discussion-Paper.pdf

надеется стать «гаражом ИИ» для стран с формирующимся рынком и развивающихся стран.

В ОАЭ президент объявил в январе 2024 г. о создании Совета по искусственному интеллекту и передовым технологиям (Artificial Intelligence and Advanced Technology Council, AIATC)[86].

В Саудовской Аравии в 2023 г. были опубликованы «Принципы этики ИИ» (AI Ethics Principles September 2023)[87].

17 декабря 2023 г. Израиль представил свою «всеобъемлющую политику в области регулирования и этики искусственного интеллекта»[88].

22 июня 2022 г. правительство Канады объявило о запуске второго этапа Всеканадской стратегии в области искусственного интеллекта[89], на который выделено более 443 млн долларов. Он направлен «на привлечение талантов мирового уровня и передовых исследовательских мощностей для коммерциализации и внедрения канадских идей и знаний».

В Бразилии юридическая основа для развития ИИ (Marco Legal da Inteligência Artificial)[90] появилась 10 декабря 2024 г., когда Федеральный сенат одобрил закон № 2338/23. Заметим, Бразилия выделяется на фоне южноамериканских стран инициативами в области управления ИИ. Еще в 2019–2021 гг. в Конгрессе были представлены три разных закона об ИИ (но ни один из них не стал официальным). Новый закон призван стать всеобъемлющим, объединив идеи предыдущих. Он, в частности, устанавливает права людей, пострадавших от применения систем ИИ.

В США отдельные штаты имеют собственное законодательство. Так, закон штата Калифорния о прозрачности ИИ (California AI Transparency Act SB 942)[91],

[86] https://www.mediaoffice.abudhabi/en/government-affairs/in-his-capacity-as-ruler-of-abu-dhabi-uae-president-issues-law-establishing-artificial-intelligence-and-advanced-technology-council

[87] https://sdaia.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf

[88] https://www.gov.il/en/pages/ai_2023

[89] https://www.canada.ca/en/innovation-science-economic-development/news/2022/06/government-of-canada-launches-second-phase-of-the-pan-canadian-artificial-intelligence-strategy.html

[90] https://www.clarkemodet.com/en/legislative-news/brasil-aprueba-un-nuevo-marco-legal-para-regular-la-ia/

[91] https://digitalpolicyalert.org/event/18058-announced-bill-on-consumer-protection-in-generative-artificial-intelligence-sb-942

начавший действовать 1 января 2026 г., обязывает компании раскрывать информацию об использовании генеративных ИИ-систем. В 2025 г. 45 штатов внесли более 550 законопроектов, связанных с ИИ, в том числе продолжаются попытки принять закон об ответственности разработчиков ИИ за причиненный вред. Но в мае республиканская партия предложила законопроект, запрещающий штатам и другим местным органам власти регулировать «модели искусственного интеллекта, системы искусственного интеллекта или автоматизированные системы принятия решений в течение 10 лет с даты принятия закона»[92].

Рис. 22. https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

30 октября 2023 г. президент США Джозеф Байден подписал указ № 14110 (рис. 22) о безопасной, надежной и заслуживающей доверия разработке и использовании ИИ (The Executive Order #14110 on Safe, Secure, and Trustworthy Artificial Intelligence)[93]. Указ, направленный на регулирование технологий ИИ, обязывал разработчиков ИИ-алгоритмов предоставлять правительству данные, полученные в ходе тестирования своих продуктов на безопасность и защиту их технологий[94]. Республиканская партия раскритиковала указ, заявив, что он тормозит развитие инноваций и накладывает чрезмерные ограничения на бизнес.

[92] https://rossaprimavera.ru/news/0e6f8ad8
[93] https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence
[94] https://3dnews.ru/1095218/prezident-ssha-predstavil-ukaz-v-sfere-regulirovaniya-ii

Президент Дональд Трамп в первый день своей каденции 20 января 2025 г. одной подписью отменил сразу 78 указов и распоряжений Байдена, включая указ 14110[95]. Отмена указа сигнализирует о более либеральном подходе администрации Трампа к регулированию ИИ[96]. Новый президент сосредоточится на стимулировании инноваций и ослаблении регулирования, в отличие от политики усиленного контроля, проводимой Байденом. И 7 апреля на сайте Белого дома появился информационный бюллетень «Устранение барьеров для использования и закупок искусственного интеллекта на федеральном уровне» (Eliminating Barriers for Federal Artificial Intelligence Use and Procurement)[97]. Теперь правительство больше не будет вводить ненужные бюрократические ограничения на использование ИИ в исполнительной власти (рис. 23).

Рис. 23. https://www.whitehouse.gov/fact-sheets/2025/04/fact-sheet-eliminating-barriers-for-federal-artificial-intelligence-use-and-procurement/

Из Америки перенесемся в Европу. 8 апреля 2019 г. Европейская комиссия опубликовала сообщение «Укрепление доверия к человекоориентированному искусственному интеллекту» (Building Trust in Human-Centric Artificial Intelligence)[98], в котором подчеркиваются ключевые требования и концепция надежного ИИ (Ethics guidelines for trustworthy AI). Согласно этому документу, «заслуживающий доверия ИИ должен быть:

[95] https://kod.ru/trump-deny-law-secure-ai-slow
[96] https://www.reuters.com/technology/artificial-intelligence/trump-revokes-biden-executive-order-addressing-ai-risks-2025-01-21
[97] https://www.whitehouse.gov/fact-sheets/2025/04/fact-sheet-eliminating-barriers-for-federal-artificial-intelligence-use-and-procurement
[98] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2019:168:FIN

- законным – соблюдать действующие законы и постановления;
- этичным – соблюдать этические принципы и ценности;
- надежным – устойчивым как с технической точки зрения, так и с учетом его социальной среды,

а также удовлетворять ключевым требованиям: это человеческое участие и надзор; техническая надежность и безопасность; конфиденциальность и управление данными; прозрачность; разнообразие, недискриминация и справедливость; благополучие общества и окружающей среды; подконтрольность».

В «Белой книге по искусственному интеллекту» (White Paper on Artificial Intelligence: A European approach to excellence and trust)[99], опубликованной 19 февраля 2020 г., подчеркивается, что «хотя ИИ может принести много пользы, он также может нанести вред. Этот вред может быть как материальным (угрозы безопасности и здоровью, гибель людей, материальный ущерб), так и нематериальным (потеря конфиденциальности, человеческого достоинства, ограничения прав на свободу выражения мнений, дискриминация, например, при приеме на работу) и может относиться к широкому спектру рисков». Нормативно-правовая база должна минимизировать риски потенциального вреда.

Правила защиты данных ЕС запрещают обработку биометрических данных с целью идентификации физического лица (исключение – существенный общественный интерес). Это ограничивает удаленную биометрическую идентификацию случаями, когда ее использование надлежащим образом оправ-дано, соразмерно и подлежит адекватным гарантиям.

13 марта 2024 г. Европейский парламент одобрил комплексный закон, регулирующий искусственный интеллект[100]. «Целью закона является предоставление разработчикам и специалистам по внедрению ИИ четких требований и обязательств в отношении конкретных видов использования этих технологий», – указано в заявлении Еврокомиссии [4] (рис. 24).

[99] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
[100] https://ria.ru/20240313/zakon-1932729659.html

Рис. 24. https://artificialintelligenceact.eu/the-act

Депутаты Европарламента требовали гарантировать, что системы ИИ останутся под контролем человека, а также что они останутся надежными, прозрачными, отслеживаемыми, недискриминационными, оберегающими окружающую среду. Кроме того, они настаивали на введении единого определения для понятия «искусственный интеллект», которое можно будет использовать по отношению к уже действующим и будущим системам.

Основное внимание в законе уделяется управлению рисками ИИ, которые разделены на четыре уровня: неприемлемый, высокий, ограниченный, низкий (минимальный). К первой группе относятся технологии, использующие «техники манипуляции или эксплуатирующие уязвимости лиц» для нанесения вреда здоровью населения или контролируемые государственными структурами для социального мониторинга. Речь идет о недопустимости классификации людей по их социальному поведению, социально-экономическому статусу или личным качествам. Согласно закону, такое применение ИИ запрещено.

Вторая группа регулирует программы, которые касаются человеческой безопасности и основных прав человека. Среди них можно выделить биометрическую идентификацию, образовательные технологии на основе ИИ, программы, отвечающие за безопасность критической инфраструктуры, и т. п. Ограничивается сбор биометрических данных в социальных сетях и системах видеонаблюдения для создания баз данных по распознаванию лиц. Исключения могут быть сделаны в очень ограниченных случаях и только по решению суда.

Положения о прозрачности применения систем ИИ вводят, в частности, обязательную маркировку текстов, аудио- видео-, фотоматериалов и чат-систем, созданных с помощью таких технологий.

ЗАКЛЮЧЕНИЕ

В России с 2019 г. действует «Национальная стратегия развития искусственного интеллекта на период до 2030 г.», введенная указом от 10.10.2019 № 490[101]. В ней говорится об основных принципах развития и использования технологий ИИ, целях такой работы, приоритетных направлениях, поддержке научных исследований, подготовке кадров, регулировании общественных отношений в этой сфере.

Вторая версия «стратегии» появилась в 2024 г. (указ от 15.02.2024 №124)[102]. В документе приводится ряд целевых показателей. Так, ежегодный объем оказанных услуг по разработке и реализации решений в области ИИ к 2030 г. должен вырасти как минимум до 60 млрд рублей (в 2022 г. – 12 млрд). Количество выпускников вузов с образованием в сфере ИИ планируется увеличить за тот же период с 3 до 15.5 тыс. человек в год. Есть и более трудноформализуемые цели. Уровень доверия граждан к технологиям ИИ в 2030 г. должен вырасти не менее чем до 80% по сравнению с 55% в 2022 г., а доля приоритетных отраслей экономики с высокой готовностью к внедрению ИИ – с 12% до 95%. Что ж, через пять лет можно будет сравнить пожелания с реальностью.

СПИСОК ЛИТЕРАТУРЫ

1. *Варшавский И.И.* Молекулярное кафе. Л. Лениздат, 1964. 256 с.

2. *Донской М.В.* Чемпионат мира среди шахматных программ // Квант. 1974. №12. С. 34–38

3. *Корсаков С.Н.* Начертание нового способа исследования при помощи машин, сравнивающих идеи. М.: МИФИ, 2009. 44 с.

4. Международное управление Интернетом / Е.С. Зиновьева, А.А. Игнатов, А.А. Уланов; под редакцией Е.С. Зиновьевой. М.: 2025. 186 с.

5. *Прохоров С.П.* Международные стандарты по этике искусственного интеллекта: история и развитие // Материалы II Международной конференции Российского национального комитета по истории и философии науки и техники

[101] http://publication.pravo.gov.ru/Document/View/0001201910110003
[102] http://publication.pravo.gov.ru/document/0001202402150063

РАН, посвященной 300-летию Российской академии наук. – Москва: Институт истории естествознания и техники им. С.И. Вавилова РАН, 2024. – С. 67-70.

6. *Шилов В.В.* На пути к искусственному интеллекту. Логические машины и их создатели. М.: Ленанд, 2019. 248 с.

ARTIFICIAL INTELLIGENCE IN SEVERAL FRAGMENTS

Y. E. Polak[0000-0001-8411-335X]

Central Economics and Mathematics Institute of the Russian Academy of Sciences, Moscow, Russia

polak@cemi.rssi.ru

Abstract

This paper is a mosaic of vivid fragments describing the industrial aspects of artificial intelligence (AI). These are sketches of the overall picture, which will likely never be completed, as each day brings information about new achievements, ideas, and threats. Discussions cover issues of civilian AI in short-term workstations, the development of algorithms for intelligent games, the threats and dangers posed by AI, AI ethics, and standards and international norms for artificial intelligence. Each fragment is a review of the latest (mid-January 2026) Russian and international sources, including quotes, translations, screenshots, and links to original documents.

This text remains an immense "fragment" on the benefits of AI applications, which was presented with the greatest speed. Perhaps this will be the beginning of a separate, never-ending study.

Keywords*: artificial intelligence, Dartmouth Seminar, predecessors of AI, development of intellectual game algorithms, threat and danger, this AI, regulation of artificial intelligence*

REFERENCES

1. *Varshavsky I.I.* Molecular Cafe. L. Lenizdat, 1964. 256 p.

2. *Donskoy M.V.* World Chess Program Championship. Quantumю 1974. No. 12. P. 34–38.

3. *Korsakov S.N.* Charting New Paths of Research with Machines: Comparative Ideas. Moscow: MEPhI, 2009. 44 p.

4. International Internet Governance. E.S. Zinovieva, A.A. Ignatov, A.A. Ulanov; edited by E.S. Zinovieva. Moscow: 2025, 186 p.

5. *Prokhorov S.P.* International standards on the ethics of artificial intelligence: history and development // Proceedings of the II International Conference of the Russian National Committee on the History and Philosophy of Science and Technology of the Russian Academy of Sciences, dedicated to the 300th anniversary of the Russian Academy of Sciences. - Moscow: S.I. Vavilov Institute for the History of Natural Science and Technology, Russian Academy of Sciences, 2024. - P. 67-70.

6. *Shilov V.V.* Toward Artificial Intelligence. Logical Machines and Their Creators. Moscow: Lenand, 2019. 248 p.

СВЕДЕНИЯ ОБ АВТОРЕ

ПОЛЯК Юрий Евгеньевич – кандидат экономических наук, ведущий научный сотрудник Центрального экономико-математического института РАН. Подробнее: http://computer-museum.ru/articles/sovet-muzeya/561/

Yuri Evgenievich POLAK – Candidate of Economic Sciences, Leading Researcher, Central Economics and Mathematics Institute. Moscow, Russia. More detailed: http://computer-museum.ru/articles/sovet-muzeya/561/

email: polak@cemi.rssi.ru

ORCID 0000-0001-8411-335X

Материал поступил в редакцию 19 января 2026 года