В. П. Часовских, В. А. Усольцев, Е. В. Кох

ЕСТЕСТВЕННЫЙ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ КАК ИНСТРУМЕНТ ПРЕОБРАЗОВАНИЯ ДАННЫХ

В. П. Часовских, В. А. Усольцев, Е. В. Кох

ЕСТЕСТВЕННЫЙ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ КАК ИНСТРУМЕНТ ПРЕОБРАЗОВАНИЯ ДАННЫХ

Монография

Киров
2025

УДК 004.8(075)
ББК 32.813я73
Ч-24

Авторы:
Часовских Виктор Петрович,
Усольцев Владимир Андреевич,
Кох Елена Викторовна

Рецензенты:
кафедра математических методов в экономике и социально-экономических наук
Уральского института фондового рынка
(протокол № 8 от 16 апреля 2024 г.);
***С. В. Поршнев,** профессор учебно-научного центра «Информационная безопасность»*
Института радиоэлектроники и информационных технологий — РтФ
Уральского федерального университета имени первого Президента России Б. Н. Ельцина
доктор технических наук, профессор

Ч-24 Часовских, В. П. Естественный и искусственный интеллект как инструмент преобразования данных [Электронный ресурс]: монография / В. П. Часовских, В. А. Усольцев, Е. В. Кох. – Электрон. текст. дан. (2,9 Мб). – Киров: Изд-во МЦИТО, 2025. – 1 электрон. опт. диск (CD-R). – Систем. требования: PC, Intel 1 ГГц, 512 Мб RAM, 2,9 Мб свобод. диск. пространства; CD-привод; ОС Windows XP и выше, ПО для чтения pdf-файлов. – Загл. с экрана.

ISBN 978-5-907974-77-7

Научное электронное издание

Естественный интеллект (человеческий) и искусственный интеллект (ИИ) представляют собой мощные инструменты для преобразования данных, хотя они функционируют по-разному и имеют различные сильные стороны. В монографии авторы определяют естественный интеллект, как функцию головного мозга человека. Рассматриваются основные характеристики головного мозга для использования в цифровом моделирование при построении ИИ. Концептуальная основа естественного интеллекта авторов основана на философии И. Канта. В монографии используется определение ИИ из «Национальную стратегию развития искусственного интеллекта в РФ». В монографии приводиться исследование технологий ИИ и утверждается, что современные технологии ИИ (слабый ИИ) образуют информационную технологию на базе цифровых вычислительных машин в архитектуре фон Неймана. Авторы приводят доказательства того, что мозг человека, его нейроны и сеть нейронов работают параллельно, процессы не алгоритмизированы, а как следствие используемые цифровые (вычислительные) модели искусственного нейрона представляют существенное упрощение биологического нейрона головного мозга. В монографии выполнен анализ фундаментальных проблем современного ИИ. Авторы отмечают, что собственно практические применено технологий ИИ является выдающимся достижение науки и практики информатика на базе цифровых вычислительных машин архитектуры фон Неймана и является мощным инструментом преобразования данных в полезную информацию и знания.

ISBN 978-5-907974-77-7

УДК 004.8(075)
ББК 32.813я73

Оглавление

ВВЕДЕНИЕ

Термин «искусственный интеллект» был формально предложен на историческом семинаре в Дартмуте (не Дортмунде) летом 1956 года.

Джон Маккарти, Марвин Минский, Клод Шеннон и другие основатели дисциплины рассматривали искусственный интеллект (ИИ) именно как направление создания программ для цифровых электронных вычислительных машин, которые могли бы демонстрировать поведение, считающееся интеллектуальным у людей.

В настоящее время важно, что исторически и практически ИИ возник и развивается как область программного обеспечения для цифровых электронных вычислительных машин (ЦЭВМ)).

В современном техническом и прикладном смысле ИИ неразрывно связан с ЦЭВМ.

Существующие рассуждения о возможности существования ИИ вне цифровых компьютеров относились скорее к теоретическим перспективам развития этой области в будущем, а не к текущей практической реальности или историческому определению.

Авторы считают, что следует уточнить некоторые аспекты (сферы) нашего исследования, а именно термины информатика, информационные технологии и место ИИ в них.

Общепринятое определение: Информатика — это наука о методах и процессах сбора, хранения, обработки, передачи, анализа и оценки информации с применением компьютерных технологий.

В нашей стране это определение закреплено в государственных образовательных стандартах и используется Российской Академией Наук. В международной практике соответствует понятию Computer Science.

Общепринятое определение: Информационные технологии — совокупность методов, производственных процессов и программно-технических

средств, интегрированных с целью сбора, обработки, хранения, распространения, отображения и использования информации в интересах ее пользователей.

В нашей стране это определение закреплено в ГОСТ 34.003-90 и международных стандартах ISO/IEC.

В образовательной сфере эти определения приняты Министерством науки и высшего образования РФ, а в мировой практике – такими организациями как IEEE (Institute of Electrical and Electronics Engineers) и ACM (Association for Computing Machinery).

Авторы считают, что в настоящее время ИИ является одним из ключевых разделов информатики как науки.

В контексте информационных технологий ИИ выступает как:

- инструментальный компонент современных информационных систем;
- технологическая парадигма, определяющая новые подходы к обработке информации;
- практическое воплощение теоретических разработок информатики;
- основа интеллектуальных информационных технологий, включающих:
  - системы поддержки принятия решений;
  - интеллектуальный анализ данных;
  - экспертные системы;
  - системы автоматического проектирования;
  - интеллектуальные пользовательские интерфейсы.

В современной структуре информационных технологий ИИ образуют самостоятельное направление разработки и внедрения, при этом они также интегрируются в традиционные информационные системы, расширяя их возможности и функциональность.

Из приведенных определений следует, что:

1. ИИ — это некоторая информационная технология;

2. ИИ — это то, что без цифровой электронной вычислительной машины существовать не может.

Естественный интеллект (ЕИ) — это термин, возникший по аналогии с ИИ для обозначения познавательных способностей, присущих живым существам, прежде всего человеку.

Этот термин появился после введения понятия ИИ как своеобразная оппозиция к нему.

Естественный интеллект занимает в области ИИ несколько ключевых позиций:

- Эталон и ориентир:
  - Человеческий интеллект служит моделью и критерием оценки для систем ИИ.
  - Тест Тьюринга и другие методы оценки ИИ основаны на сравнении с естественным интеллектом.
- Источник вдохновения:
  - Нейронные сети созданы по аналогии с нервной системой.
  - Генетические алгоритмы имитируют эволюционные процессы.
  - Когнитивные архитектуры моделируют структуру человеческого познания.
- Объект исследования:
  - Понимание естественного интеллекта часто является необходимым шагом для создания ИИ.
  - Нейронауки и когнитивные науки тесно взаимодействуют с ИИ.

В методологии ИИ выделяют два основных подхода:

- нисходящий (символьный)— моделирование высокоуровневых мыслительных процессов;
- восходящий (коннекционистский)— моделирование нейронных структур мозга.

Эти два подхода в разной степени опираются на понимание ЕИ.

Сопоставление ЕИ и ИИ поднимает фундаментальные вопросы:

- о природе сознания и мышления;

- о возможностях и ограничениях технического воспроизведения когнитивных функций;
- об этических границах при создании систем, имитирующих человеческий интеллект.

ЕИ является одновременно прототипом, ориентиром и источником вдохновения для исследований в области ИИ.

Взаимоотношение между ЕИ и ИИ остается центральной темой как в теоретических исследованиях, так и в практических разработках ИИ, в том числе и в исследовании авторов.

В настоящее время в мире происходит ускоренное внедрение технологических решений, разработанных на основе ИИ в различные отрасли экономики и сферы общественных отношений. ИИ превратился в одну из самых обсуждаемых и быстроразвивающихся технологий нашего времени.

Трудно придумать некоторую отрасль или вид деятельности, о которых можно сказать – ИИ не приемлем. Мы ежедневно слышим о ИИ в новостях, видим его работу в смартфонах, используем в повседневной жизни, даже не задумываясь об этом. ИИ уже изменил мир вокруг нас и продолжает трансформировать практически все сферы человеческой жизни.

Понятие об ИИ многие люди имеют лишь поверхностное, о том, что такое ИИ, как он работает, какие возможности ИИ открывает, несмотря на повсеместное распространение. Термин часто окружен мифами и заблуждениями.

Всем понятно, что ИИ является некоторой моделью естественного интеллекта (ЕИ) человека. И всем понято, что ИИ проявляет все свои возможности в среде цифровой электронной вычислительной машины (ЭВМ).

Авторы исследуют интеллект человека, как функционирование его мозга и называемый ЕИ, а также модели, техники и технологии ИИ. Авторы используют термины и определения, определяемые в Указах Президента РФ [1, 2], главным является определение ИИ.

«Искусственный интеллект – комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и по-

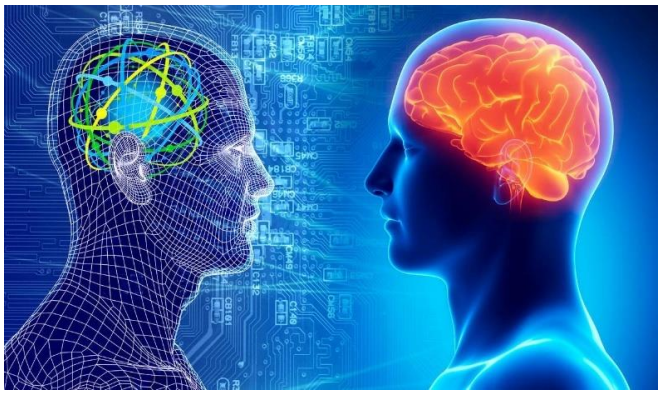
иск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе, в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений» [1, 2].

«Технологии искусственного интеллекта – технологии, основанные на использовании искусственного интеллекта, включая компьютерное зрение, обработку естественного языка, распознавание и синтез речи, интеллектуальную поддержку принятия решений и перспективные методы искусственного интеллекта» [1, 2].

В качестве определения ЕИ авторы применяют понятия И. Канта [3], в интерпретации Академика РАН А.И. Аветисяна [4].

Определим области нашего исследования (рис. 1).

ЭВМ
Алгоритм
Программы
Данные
Искусственный нейрон
Искусственная нейронная сеть



Головной мозг
Интеллект
Сознание
Нейрон
Нейронная сеть
Эмпирические знания
Априорные знания

Рисунок 1. Области исследования авторов

Авторы учитывают следующие ограничение:

1. ЭВМ — это цифровая электронная вычислительная машина, архитектура которой определяется фундаментальной моделью фон Неймана [15]. Все современные ЦЭВМ построены по модели фон Неймана.
2. На ЦЭВМ программы и данные хранятся в одной и той же памяти. Команды программы выполняются последовательно, одна за другой. реализующие некоторый алгоритм.
3. Процессор (ядро) ЦЭВМ все вычисления выполняет последовательно, определяется моделью Тьюринга.

ГЛАВА 1. ЕСТЕСТВЕННЫЙ ИНТЕЛЛЕКТ

1.1. Мозг человека и его характеристики, актуальные для ИТ

Мозг [13, 14] взрослого человека содержит около 100 млрд нейронов.

Длина всех нервных волокон в полушариях мозга составляет 500 тысяч километров.

Площадь коры головного мозга составляет около 20 квадратных метров.

Длина всех нейронных связей головного мозга более 1 миллиона километров – это почти 3 расстояния от Луны до Земли.

Скорость распространения нервного импульса достигает 288 км/ч, а к старости снижается примерно на 15%.

Объём памяти головного мозга составляет около 2.5–6 петабайт (с таким размером памяти можно больше 300 лет непрерывно смотреть фильмы на DVD).

Мозг составляет всего 2% тела человека, но потребляет 17% энергии тела и 20% кислорода. Мозг – чемпион по кровоснабжению. За 1 минуту через него проходит 1 литр крови. Каждый нейрон имеет собственную микроскопическую «кровеносную систему».

Мозговые нейроны посылают от 5 до 50 сообщений ежесекундно. Они бывают 3 видов:

- Чувствительные – органы чувств (глаза, уши, носа) доставляют им информацию, которую они должны отправить ЦНС;
- Эффекторные – передают информацию от других нейронов к мышцам, органам и разным железам;
- Вставочные – взаимодействуют с нервными клетками.

Мозг ежесекундно формирует более 1 миллиона нейронных связей.

Мозг человека состоит из нескольких областей (лобной, височной, теменной, затылочной). Каждая область отвечает за определенные функции. Например, лобная отвечает за планирование, принятие решений и контроль эмоций, а

височная область – за обработку зрительной информации. Работает мозг по принципу электрохимических сигналов. Когда мы воспринимаем информацию из окружающего мира, наши чувства передают сигналы электрическими импульсами к соответствующим частям мозга. Нейроны в мозге передают электрические импульсы друг другу через специальные соединительные точки, называемые синапсами. При активации нейрона он генерирует электрический импульс, который передается через аксон (длинный отросток нейрона) к синапсу. Там сигнал (импульс) переходит на следующий нейрон через химические вещества – нейромедиаторы. Эта передача сигнала называется синаптическим соединением.

Однако работа мозга не ограничивается только восприятием информации. Он также управляет движением, координацией, памятью, речью, эмоциями и т. д.

Главный орган, ответственный и за наш интеллект, способен изменяться и адаптироваться к новым условиям. Это свойство называется пластичностью мозга – оно позволяет ему формировать новые связи между нейронами и изменять существующие связи в зависимости от потребностей и опыта человека.

Калифорнийский технологический институт получил следующие результаты исследования мозга – человеческий мозг обрабатывает информацию со скоростью около 10 бит в секунду.

10 бит в секунду относится к сознательной обработке новой информации человеческим мозгом, а не к общей вычислительной мощности всего мозга или отдельных нейронов.

Различные уровни обработки информации в мозге.

- Сознательная обработка (~10 бит/с): относится к информации, которую мы осознанно воспринимаем и обрабатываем. Например, скорость осознанного чтения нового материала или принятия решений. Это «узкое горлышко» сознания, а не общая производительность мозга.
- Бессознательная обработка (значительно выше). Зрительная система обрабатывает ~10^7–10^8 бит/с. Общая пропускная способность всех сенсорных систем ~10^{11} бит/с.

Одиночный нейрон может генерировать до ~200 импульсов в секунду.

Каждый импульс несет примерно несколько бит информации. При $\sim 10^{11}$ нейронов и $\sim 10^{14}$ синапсов общая вычислительная мощность значительно выше.

Общая расчетная мощность мозга: различные оценки варьируются от 10^{14} до 10^{16} операций в секунду. По некоторым оценкам, мозг способен хранить $\sim 10^{15}$ бит информации. Таким образом, цифра в 10 бит/с представляет собой лишь «пропускную способность сознания», то есть скорость, с которой информация становится доступной для сознательного восприятия, а не общую вычислительную мощность мозга, которая многократно выше.

Мозг представляет собой биологическую структуру, обеспечивающую материальную основу для проявления интеллекта. Обеспечивает эту взаимосвязь от 80 до 90 миллиардов нейронов, триллионы синаптических соединений, различные специализированные области (неокортекс, гиппокамп, мозжечок и др.). Все нейроны головного мозга работают параллельно. Специализированные области мозга человека приведены в табл. 1.

Таблица 1. Некоторые специализированные области мозга человека [5, 6]

Неокортекс	Эта структура играет ключевую роль в формировании высших когнитивных функций, определяющих уникальность человеческого интеллекта. Неокортекс представляет собой уникальную биологическую систему, обеспечивающую большинство высших когнитивных функций человека. Его шестислойная структура, организованная в функциональные колонки и сложные сети, реализует принципиально важные вычислительные механизмы, позволяющие осуществлять абстрактное мышление, символическую обработку информации, планирование и творчество. Понимание неокортекса имеет фундаментальное значение как для неврологии и нейробиологии, так и для разработки более продвинутых систем искусственного интеллекта, способных приблизиться к гибкости, адаптивности и эффективности человеческого мозга.
Гиппокамп	Ключевые функции гиппокампа: 1. Формирование новых воспоминаний Эпизодическая память: отвечает за запоминание событий, их последовательности и контекста. Перевод кратковременной памяти в долговременную: информация сначала обрабатывается в гиппокампе, затем постепенно переносится в кору для постоянного хранения.

	Консолидация памяти: происходит преимущественно во время сна, особенно в фазе глубокого сна. 2. Пространственная навигация Когнитивные карты: содержит «клетки места» (place cells), которые активируются, когда человек находится в определенном месте окружающей среды Пространственная ориентация: помогает понимать, где мы находимся и как добраться до цели. 3. Эмоциональное регулирование Тесно связан с миндалевидным телом (амигдалой). Участвует в регуляции эмоциональных реакций и формировании эмоциональной памяти.
Мозжечок	Составляет около 10% общего объема мозга, но содержит почти 80% всех нейронов центральной нервной системы. Современные исследования показывают, что мозжечок играет гораздо более глубокую роль в работе мозга, чем считалось раньше, выходя далеко за пределы простой моторной координации и участвуя в сложных когнитивных и эмоциональных процессах.

В работах [28, 29] Роджер Пенроуз и Джон Лукас используют теоремы Гёделя о неполноте для аргументации против возможности полной алгоритмизации человеческого мышления.

Основные аргументы Лукаса:

1. Согласно теоремам Гёделя, для любой формальной системы, достаточной для арифметики, существуют истинные утверждения, которые невозможно доказать в рамках этой системы.
2. Компьютеры работают как формальные системы, следуя алгоритмам.
3. Человек способен «увидеть» истинность гёделевских предложений, недоказуемых в данной формальной системе.
4. Следовательно, человеческий разум не может быть сведен к алгоритму.

Основные проблемы в аргументации Лукаса:

1. **Проблема непоследовательности**. Лукас предполагает, что человеческий разум может распознать истинность любого геделевского предложения, но не предоставляет доказательство этой способности.
2. **Ограниченность когнитивных способностей**. Реальные люди имеют ограничения памяти и вычислительных возможностей, которые мешают им понять геделевские предложения для достаточно сложных формальных систем.

3. **Проблема механизма**. Лукас не объясняет, каким образом человеческий разум способен «видеть» истинность геделевских предложений.
4. **Неверное применение теоремы Гёделя**. Теорема применима к формальным системам, а не обязательно к любому физическому устройству, включая мозг.
5. **Игнорирование вероятностных и недетерминированных алгоритмов**. Аргумент Лукаса фокусируется на детерминированных алгоритмах, игнорируя другие вычислительные модели.

В аргументации против Лукаса и Пенроуз предполагается, что мозг человека физическая система, но это не так: мозг – биологическая система. Законы биологических систем не все определены, далеко не все.

Мозг человека — это прежде всего биологическая система, а не просто физическое устройство.

Биологические системы характеризуются:

1. **Сложной иерархической организацией**— от молекулярного уровня до системного;
2. **Эмерджентными свойствами**— возникающими на каждом уровне организации;
3. **Адаптивностью и пластичностью;**
4. **Способностью к самоорганизации и саморегуляции;**
5. **Исторически обусловленным развитием** (как эволюционным, так и онтогенетическим).

В настоящее время законы функционирования биологических систем далеко не полностью определены. Существуют:

- уровни биохимических взаимодействий, которые не полностью описаны;
- эпигенетические механизмы регуляции;
- квантовые эффекты в биохимических процессах;
- сложные закономерности нейронных сетей и их взаимодействий.

Это действительно ослабляет позицию критиков Лукаса, поскольку они часто исходят из предположения, что мозг является системой, подчиняющейся полностью определённым физическим законам.

В контексте биологической природы мозга аргумент Лукаса может быть усилен: если мозг как биологическая система обладает принципиально иными свойствами, чем формальные системы или физические вычислительные устройства, то попытка свести его функционирование к алгоритмам может быть изначально некорректной.

Таким образом, неполнота нашего понимания биологических закономерностей работы мозга действительно оставляет открытым вопрос о его потенциальной не алгоритмизируемости, что придаёт дополнительный вес позиции Лукаса.

Авторы придерживаются не алгоритмизации мозга человека, так же считает и Академик РАН И. Каляев [11].

Является очевидным, что мозг человека и интеллект взаимосвязаны.

Важным является мнение известных специалистов и, прежде всего, мнение крупнейшего всемирно известного нейрохирурга академика Александра Николаевича Коновалова. Золотую медаль Герой Труда России под №1 получил именно Александр Николаевич. На счету Коновалова более 17 тысяч операций на мозге, самых сложных, в том числе и таких, которые раньше никто и нигде не проводил. Лауреат премии Нейрохирургического Общества Уолтера Э. Дэнди (признан лучшим нейрохирургом в мире). На вопрос журналистки Светланы Беляевой «Существует ли какая-то тайна человеческого мозга, которая до сих пор ставит Вас в тупик, и в которой очень хотелось бы разобраться?» Александр Николаевич ответил: «Да, есть одна фундаментальная загадка, которая не дает мне покоя. Мы можем досконально изучить мозг – все эти миллиарды нейронов и триллионы связей между ними. Мы понимаем биохимические процессы, электрическую активность, работу отдельных участков. **Все это материально, измеримо, объяснимо. Но как из этой материальной основы**

рождается сознание? Как совокупность электрохимических импульсов превращается в мое «я», в способность воспринимать мир, мыслить, чувствовать? Почему я воспринимаю мир? Почему я вижу Вас? Почему я с Вами разговариваю? Где происходит этот магический переход от материального к нематериальному, к тому, что мы называем сознанием, душой, внутренним миром? Мы можем описать все механизмы, все нейрофизиологические процессы. Но сама суть этого перехода остается величайшей тайной.» [7]. Александр Николаевич дал важную оценку понимания мозга человека в настоящее время (авторы придерживаются этой оценки): «Многие люди думают, что, если человек оперирует мозг, то он все про него знает, а я убеждён, что я знаю очень мало. Я почти ничего не знаю о том, как работает вот эта фантастическая машина. Я думаю, что нет людей, которые разобрались в этом, и никогда не разберутся» [8].

Интересно мнение о головном мозге Татьяны Черниговской, доктора биологических наук, нейролингвиста и экспериментального психолога, заведующей лабораторией когнитивных исследований Санкт-Петербургского государственного университета, специалиста по теории сознания. Татьяна Владимировна отвечает на вопросы о головном мозге человека (табл. 2 [9]).

Таблица 2. Диалог Т.В. Черниговская с корреспондентом

Корреспондент	**Татьяна Черниговская**
«Что за «существо» живёт в нашей черепной коробке?»	«Можно сказать и так. Мозг — в результате эволюции или волей судьбы — оказался у нас в черепной коробке, и в этом смысле мы можем, если хотим, называть его «мой». Но загвоздка в том, что он несопоставимо более мощный, чем мы с вами. Это парадокс человеческой природы и природы вообще.» [9].
«Значит, мозг и человек — это не одно и то же?»	«Да, не одно. Мы не имеем власти над мозгом: он принимает решения самостоятельно. И это обстоятельство ставит нас в очень, как бы это сказать, щекотливое и даже сомнительное положение.» [9].
«Где вообще находится сознание человека? В мозге, в центральной нервной системе, во всём теле?»	«Раньше я бы сказала, что, конечно, в мозге, потому что больше негде. А сейчас отвечу, что сознание воплощено и в нашем теле. И зависит оно от невообразимого множества мельчайших факторов, которые просто невозможно как-то учесть и зарегистрировать. Вот почему, например, идея о том, что сознание человека можно записать на нейронную сеть, скопировать на некий искусственный носитель, снабдить сенсорами для ощущений, и таким образом даровать человеку бессмертие — это утопия. Что, с позволения сказать,

Корреспондент	Татьяна Черниговская
	вы собираетесь копировать? Как можно зафиксировать в электронной памяти любую мелочь, которая случилась с человеком в течение жизни? Ведь личность и состоит из таких вроде бы незначимых мелочей.» [9].
«Что ещё в деятельности мозга для вас необъяснимо?»	«На какой энергии он работает. Для всей его жизнедеятельности вполне хватает мощности какой-нибудь слабосильной лампочки в 10 ватт: такие горят в холодильнике. Между тем какому-нибудь сверхсовременному суперкомпьютеру нужны уже мегаватты, а сеть таких устройств потребляет энергию, необходимую для электрификации небольшого города.
Что ещё в деятельности мозга для вас необъяснимо?»	Но мозг — это не «софт» и компьютерное «железо». **Он на 78% состоит из воды, на 15% из жира, а остальное — белки, гидрат калия и соль**. Про вес и массу его вообще смешно говорить. И при всём при этом во Вселенной мы не знаем ничего более сложного, чем мозг. В нём более 120 млрд нейронов! И у каждого из них до 50 тысяч связей с другими частями нашего «серого вещества». В целом квадриллион связей и 5,5 петабайт информации, то есть 3 млн часов (или триста лет) непрерывного просмотра видеоматериала. Это невообразимые числа. А недавно все компьютеры мира сравнялись по производительности всего лишь с одним человеческим мозгом. Иначе говоря, мозг не просто нейронная сеть, а сеть сетей.» [9].

Анализ научно-практического мнения академика РАН А.Н. Коновалова и доктора биологических наук Т.В. Черниговской позволяет утверждать, что мировое сообщество готово инвестировать колоссальные ресурсы в нейронауку. Раскрыв подлинные механизмы нейронных сетей, синаптической пластичности и когнитивных процессов, мы бы немедленно трансформировали все аспекты человеческого бытия — от коммуникационных технологий и нейроморфной энергетики до биомиметического производства. Подобное прорывное знание способно коренным образом переустроить цивилизацию: сформировать иную образовательную парадигму, сменить социальную архитектуру и задать новой траектории развитие планетарных экосистем. Авторы определили те свойства, которые необходимо учитывать в цифровых моделях нейрона и искусственных нейронных сетей.

1.2. И. Кант об интеллекте

Философия Иммануила Канта представляет одну из наиболее фундаментальных концепций интеллекта, оказавшую глубокое влияние на всё последующее развитие философской мысли. Кантовская система не просто описывает, как работает человеческий интеллект, но переосмысливает саму природу познания, его структуру и границы.

Кантовская концепция интеллекта неотделима от его «коперниканского переворота» в философии. Вместо традиционного представления о том, что познание подстраивается под объекты, Кант предлагает радикально иной подход: объекты познания должны сообразовываться с нашей познавательной способностью.

«До сих пор считали, что всякие наши знания должны сообразоваться с предметами… Следовало бы попытаться выяснить, не разрешим ли мы задачи метафизики успешнее, если будем исходить из предположения, что предметы должны сообразоваться с нашим познанием» (Предисловие ко второму изданию «Критики чистого разума») [3].

Кант разработал сложную архитектонику интеллектуальных способностей, показанную в табл. 3:

Таблица 3. Архитектура интеллекта И. Канта

Способность	Сущность	Функция
1. Чувственность (Sinnlichkeit)	Способность получать представления благодаря воздействию предметов Априорные формы: пространство и время	Организация чувственных данных в пространственно-временной форме
2. Рассудок (Verstand)	Спонтанная деятельность мышления, способность создавать понятия Априорные категории: 12 категорий, организованные в 4 группы (количество, качество, отношение, модальность)	Синтез многообразия созерцаний, подведение явлений под понятия
3. Разум (Vernunft)	Высшая способность познания, стремящаяся к безусловному и абсолютному	Систематизация знаний рассудка, регулятивное направление познания

Способность	Сущность	Функция
	Трансцендентальные идеи: психологическая (душа), космологическая (мир), теологическая (Бог)	
4. Способность суждения (Urteilskraft)	Посредник между рассудком и разумом Виды: определяющая и рефлектирующая	Подведение особенного под общее (определяющая), нахождение общего для особенного (рефлектирующая)

Центральное место в кантовской концепции интеллекта занимает понятие трансцендентального единства апперцепции – самосознания, которое должно сопровождать все представления («Я мыслю»). «Должна существовать возможность того, чтобы «я мыслю» сопровождало все мои представления; в противном случае во мне представлялось бы нечто такое, что вовсе не могло бы быть мыслимо» [3].

Это единство сознания является условием возможности всякого опыта и познания, обеспечивая синтетическую связь представлений.

Важнейшее достижение кантовской теории интеллекта — обоснование возможности синтетических априорных суждений, которые:

1. Расширяют наше знание (в отличие от аналитических суждений);
2. Имеют всеобщий и необходимый характер (в отличие от эмпирических суждений).

Эти суждения возможны благодаря априорным формам чувственности и категориям рассудка.

Наряду с теоретическим интеллектом Кант вводит понятие практического разума, который:

1. Устанавливает нравственные законы.
2. Обеспечивает автономию воли.
3. Связан с идеей свободы как постулатом.

Значение кантовской концепции интеллекта следующее:

1. Активный характер познания: интеллект не пассивно отражает мир, а активно конструирует опыт.

2. Связь чувственности и мышления: «Мысли без содержания пусты, созерцания без понятий слепы».
3. Автономия интеллекта: способность устанавливать собственные законы как в познании, так и в морали.

Самокритичность разума Канта: интеллект способен исследовать собственные возможности и границы.

Системность познания Канта: все познавательные способности образуют единую архитектоническую систему.

Кантовская концепция интеллекта представляет собой попытку синтеза рационалистического и эмпирического подходов, преодолевая их односторонность и закладывая основы современного понимания познавательных процессов. Она подчеркивает как конструктивную мощь человеческого интеллекта, так и его принципиальные ограничения.

Философия Иммануила Канта предлагает, рис. 2. удивительно глубокий взгляд на человеческий интеллект, который может служить критической точкой отсчета для анализа современных систем ИИ.

Когда мы сопоставляем кантовскую концепцию интеллекта с возможностями и ограничениями современного ИИ, мы обнаруживаем фундаментальные проблемы, которые выходят далеко за рамки технических вопросов. Архитектура интеллекта по Канту показана на рис.2.

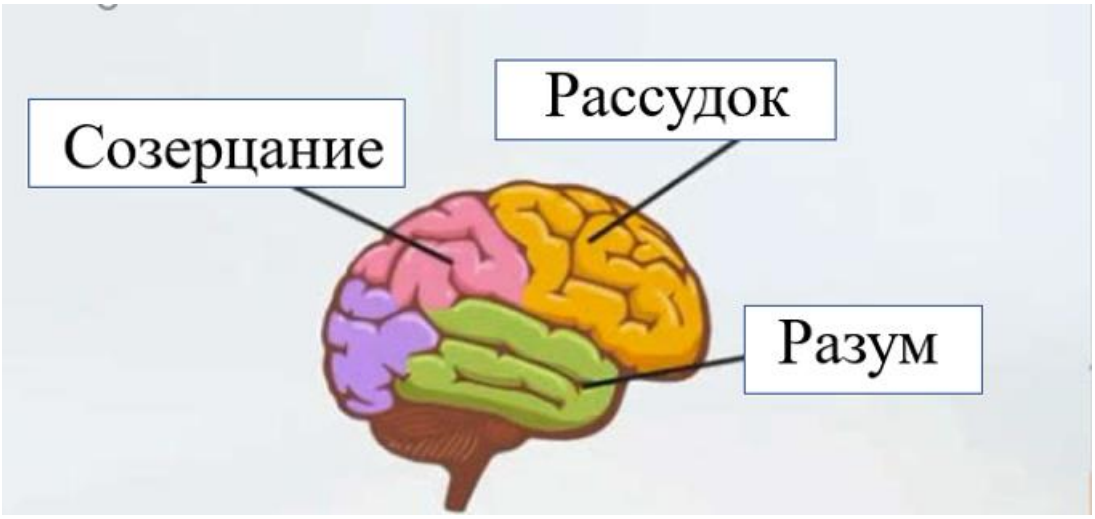


Рис. 2. Архитектура интеллекта по Канту

Кант выделяет три ключевых познавательных способности, образующих целостную архитектуру интеллекта: Созерцание, Рассудок и Разум.

У Канта активная природа интеллекта – ключевой момент познание не есть пассивное отражение мира.

Интеллект активно конструирует знание, организуя чувственные данные посредством априорных структур и категорий.

Кант определил самосознание как фундаментальное условие интеллекта— трансцендентальное единство апперцепции, то есть самосознание как способность соотносить все представления с единым «Я мыслю».

Современный ИИ через призму кантовской философии следующий:

Проблема 1. Отсутствие подлинного созерцания.

Кантовская перспектива: созерцание предполагает непосредственное восприятие мира через врожденные формы пространства и времени.

Проблема ИИ: современные системы ИИ не имеют прямого опыта мира. Их «восприятие» опосредовано заранее подготовленными данными. Отсутствует телесность и сенсорно-моторный опыт как основа когнитивных процессов. Нет феноменального сознания, то есть субъективного переживания опыта.

Проблема 2. Механистичность рассудка.

Кантовская перспектива: рассудок — спонтанная, творческая способность синтеза и организации опыта.

Проблема ИИ: нейронные сети функционируют как сложные статистические модели, но не обладают подлинным пониманием. Обработка информации скорее имитирует, чем воспроизводит человеческое мышление. Отсутствие интенциональности — ИИ не имеет направленности сознания на объекты.

Проблема 3. Отсутствие автономного разума.

Кантовская перспектива: разум способен к самокритике, рефлексии над собственными границами и возможностями.

Проблема ИИ: не способен к подлинной саморефлексии и критике собственных оснований. Не может самостоятельно формулировать цели своей деятельности. Не имеет внутренней нормативности — не может самостоятельно устанавливать правила.

Проблема 4. Отсутствие единства апперцепции.

Кантовская перспектива: Самосознание — необходимое условие всякого познания и мышления.

В современном ИИ – отсутствие самосознания и единого «Я»; нет субъективного центра опыта и переживания; не может подлинно понимать себя как мыслящее существо.

Практические следствия для развития ИИ показаны в табл. 4.

Таблица 4. Практические следствия для развития ИИ

Название	Возможности
1. Пределы моделирования интеллекта	Кантовская концепция интеллекта указывает на принципиальные ограничения в моделировании человеческого мышления: Невозможно моделировать разум без тела— телесный опыт конститутивен для интеллекта. Интенциональность не редуцируется к вычислительным процессам. Самосознание не сводимо к обработке информации о себе.
2. Этические и социальные импликации	Проблема ответственности: может ли существо без самосознания и автономии быть морально ответственным агентом? Проблема делегирования решений: допустимо ли передавать важные решения системам, не имеющим понимания и ценностей? Проблема человеческого достоинства: не ведет ли смешение человеческого и машинного интеллекта к размыванию представлений о человеческом достоинстве?
3. Новые направления исследований	Воплощенный ИИ (Embodied AI): развитие систем, укорененных в телесном опыте и взаимодействии с миром. Энактивизм: понимание познания как активной, телесно-ориентированной деятельности. Феноменологический подход к ИИ: внимание к вопросам переживания и субъективного опыта.

Сопоставление кантовской концепции интеллекта с реальностью современного ИИ ведет к глубинным философским вопросам:

1. Что значит быть разумным существом? Кантовская философия напоминает нам, что разум — это не просто способность к вычислению, но и способность к самоопределению, рефлексии и автономии.

2. Является ли человеческий интеллект уникальным? Кант предлагает нам рассматривать человеческий разум как обладающий уникальными качествами, которые не сводимы к вычислительным процессам.

3. Какой ИИ нам нужен? Вместо попыток воспроизвести человеческий интеллект во всей полноте, возможно, стоит стремиться к созданию систем, которые дополняют человеческий интеллект, компенсируя его ограничения.

В эпоху стремительного развития ИИ это напоминание становится особенно ценным, побуждая к более глубокому и нюансированному пониманию как искусственного, так и человеческого интеллекта.

1.3. Концепции сознания человека для ИИ

Сознание из ЕИ как чисто философской проблемы превращается в практический вопрос ИИ, требующий научного, этического и правового осмысления в контексте развития всё более сложных систем ИИ. Актуальность проблемы сознания в ИИ показана в табл. 5., сформированной по научным публикация и научным конференциям.

Таблица 5. Актуальность проблемы сознания в настоящее время

В контексте развития ИИ	
Граница между имитацией и подлинным пониманием	Современные LLM создают убедительную иллюзию понимания. Растёт необходимость четких критериев различения симуляции сознания от реального феномена. Тест Тьюринга становится всё менее адекватным инструментом оценки.
Этические аспекты разработки	Неясность в вопросе, может ли сложная нейросеть испытывать что-то подобное страданию. Проблема ответственности за создание потенциально сознающих систем. Возможность появления «цифровых лиц» с моральным статусом.
Регуляторные вызовы	Отсутствие юридического определения и критериев «искусственного сознания.» Необходимость опережающего законодательного реагирования. Вопрос прав потенциально сознающих систем.

В научной сфере	
Междисциплинарная конвергенция	Объединение усилий нейронауки, философии сознания, компьютерных наук и физики. Развитие новых методов исследования нейронных коррелятов сознания. Формирование более строгих научных теорий сознания (IIT, GWT и др.).
Проверяемые гипотезы	Переход от чисто философских концепций к экспериментально тестируемым моделям. Разработка количественных метрик для оценки «степеней сознания». Создание синтетических систем с контролируемыми параметрами для проверки теорий.
В философском измерении	
Переосмысление человеческой исключительности	Размытие границ между естественным и искусственным интеллектом. Вопрос о возможности нечеловеческих форм сознания. Пересмотр концепций идентичности и непрерывности сознания.
Новое понимание разума и тела	Возможность субстратно-независимого сознания (информационный подход). Вопросы о необходимости биологического носителя. Проблема репликации и трансфера сознания.
Практические последствия	
Медицинские приложения	Разработка более точных методов оценки сознания у пациентов с нарушениями. Новые подходы к лечению психических расстройств. Этические вопросы поддержания жизни при различных состояниях сознания.
Образовательные вызовы	Необходимость подготовки специалистов в новой междисциплинарной области. Формирование общественного понимания сложных вопросов сознания. Преодоление как упрощенных, так и мистифицированных представлений.

В табл. 6 перечислены ученые, которые представляют различные дисциплины и подходы к проблеме сознания в ИИ, от строго философских до практических инженерных, создавая многогранную дискуссию о возможности, природе и последствиях потенциального создания искусственного сознания.

Таблица 6. Известные ученые по теории сознания применительно к ИИ

Философы и теоретики сознания	
Дэвид Чалмерс [20]	Известен формулировкой «трудной проблемы сознания». Исследует возможность сознания в искусственных системах.
Дэвид Чалмерс [20]	Разрабатывает концепцию субстратно-независимой информационной теории сознания Ключевая работа: «The Conscious Mind» и статьи о возможности машинного сознания.
Джон Серл [24]	Создатель мысленного эксперимента «Китайская комната», критикующего сильный ИИ. Проводит различие между синтаксисом (обработка символов) и семантикой (понимание). Утверждает, что компьютерные системы могут симулировать, но не дублировать сознание. Ключевая работа: «Minds, Brains and Programs»
Сьюзан Шнайдер [23]	Исследует связь между вычислительными процессами и сознанием. Разрабатывает тесты на определение сознания в ИИ-системах. Ключевая работа: «Artificial You: AI and the Future of Your Mind**».**
Ник Бостром [21]	Изучает потенциал и риски сверхразумного ИИ Исследует этические вопросы искусственного сознания. Ключевая работа: «Superintelligence: Paths, Dangers, Strategies».
Ученые-когнитивисты и нейроисследователи	
Джулио Тонони [26]	Создатель теории интегрированной информации (IIT).
Джулио Тонони [27]	Предложил математический подход к измерению сознания (Ф-мера). Его теория активно применяется к оценке потенциального сознания в ИИ. Ключевая работа: «Phi: A Voyage from the Brain to the Soul».
Кристоф Кох [16]	Развивает нейробиологические теории сознания Применяет теорию интегрированной информации к ИИ-системам. Ключевая работа: «Consciousness: Confessions of a Romantic Reductionist».
Марвин Минский [151,152]	Пионер ИИ, исследовавший связь между сознанием и интеллектом. Развил идею «общества разума» — множества взаимодействующих процессов. Ключевая работа: «The Society of Mind».
Исследователи ИИ, работающие над проблемой сознания	
Дуглас Хофштадтер [16,17]	Исследует самореференцию и петли обратной связи как основу сознания. Анализирует возможности создания самосознания в ИИ. Ключевая работа: «Gödel, Escher, Bach: An Eternal Golden Braid**».**

Дэниел Деннет [155]	Защищает функционалистский подход к сознанию, потенциально реализуемый в ИИ. Предлагает «гетерофеноменологию» как метод изучения сознания. Ключевая работа: «Consciousness Explain».
Йошуа Бенджио [156]	Пионер глубокого обучения, изучающий когнитивные архитектуры ИИ. Исследует возможности создания системного мышления в нейросетях. Вносит вклад в понимание осознанности и метапознания в ИИ.
Стюарт Рассел [19]	Работает над проблемами разработки безопасного ИИ. Исследует аспекты выравнивания ценностей и понимания контекста. Ключевая работа: «Human Compatible: Artificial Intelligence and the Problem of Control»
Дэвид Роуз [157]	Изучает машинную психологию и имитацию психологических процессов. Фокусируется на социальных аспектах сознания в ИИ.

Концепция сознания человека по Марвину Минскому

Марвин Минский, один из основоположников искусственного интеллекта, разработал оригинальную и влиятельную концепцию сознания в рамках своей теории «Общества разума» (Society of Mind) и последующей работы «Машина эмоций» (The Emotion Machine).

Ключевые аспекты теории сознания Минского

1. Сознание как децентрализованное «общество агентов»

По Минскому, сознание не является единым монолитным феноменом или отдельной «вещью», а представляет собой **эмерджентное свойство**, возникающее из взаимодействия множества простых когнитивных процессов:

«Разум – это то, что возникает, когда мозг занимается определенными процессами. Эти процессы включают конструирование и манипулирование структурами символов.»

Минский рассматривал разум как **сообщество многих небольших процессов** (которые он назвал «агентами»), каждый из которых выполняет простую функцию. Сознание возникает из взаимодействия этих агентов:

«Каждый умственный агент сам по себе может выполнять только некоторую простую задачу, которая не требует того, что мы называем мышлением или интеллектом. Тем не менее, когда эти агенты объединяются в общества определенным образом, результатом становится истинный интеллект.»

2. Сознание как система само мониторинга

Минский определял сознательный опыт через **мету процессы наблюдения**. Согласно его концепции, ключевым аспектом сознания является способность систем мозга наблюдать за своими собственными процессами: «Мы называем «сознательными» те психические активности, которые мы можем наблюдать через само отражение – то есть, через активность неких частей мозга, наблюдающих за другими частями мозга.»

3. Многоуровневая иерархия процессов

Минский предложил **иерархическую модель сознания**, где высшие уровни контролируют и регулируют низшие:

«Сознание – это не отдельная способность, а результат сложной организации многих специализированных ресурсов.»

Он выделял шесть уровней ментальных процессов:

1 уровень. **Инстинктивные реакции**: базовые реакции на раздражители.

2 уровень. **Выученные реакции**: обусловленные ответы.

3 уровень. **Целенаправленное мышление**: преднамеренные действия.

4 уровень. **Рефлексивное мышление**: размышление о собственном мышлении.

5 уровень. **Само отражающее мышление**: создание моделей собственной психики.

6 уровень. **Само рефлексивная осведомленность**: осознание своего самосознания.

4. Критика «театральной» модели сознания

Минский последовательно выступал против концепции сознания как **«центрального театра»** или единой сцены, где происходит осознание:

«Мы часто говорим о сознании так, словно это место в мозге, где показываются фильмы нашего опыта. Это вводит в заблуждение, поскольку нет единого места, где ‘всё объединяется’.»

Вместо этого он утверждал, что сознательный опыт распределен по множеству параллельных процессов.

5. Сознание и эмоции

В «Машине эмоций» Минский развил идею о том, что эмоции не противоположны рациональному мышлению, а **неотъемлемая часть сознания и когнитивных процессов**:

«Эмоции – это не отдельные и противоположные рациональному мышлению силы, а особые методы и способы, с помощью которых наш мозг управляет своими ресурсами.»

Он рассматривал эмоции как различные режимы работы когнитивной системы, помогающие организовывать и направлять мышление.

6. Иллюзорная природа единства сознания

Минский считал, что ощущение единого и непрерывного сознания – это **конструируемая иллюзия**: «У нас создается иллюзия, что наш разум един и неделим, потому что мы не осознаем бесчисленные процессы, которые работают, чтобы создать наши мысли.»

Значение концепции Минского для искусственного интеллекта

Архитектурный подход Минского к сознанию имел огромное влияние на исследования ИИ:

1. **Модульность интеллекта**: идея, что сложный интеллект может возникнуть из взаимодействия простых компонентов.
2. **Распределенное представление**: отказ от поиска единого «центра» сознания в пользу распределенных систем.
3. **Интеграция эмоций**: включение эмоциональных механизмов как необходимых компонентов интеллектуальных систем.

4. **Метакогнитивные способности**: акцент на самонаблюдение и само моделирование как на ключевые аспекты продвинутого интеллекта.

Минский создал радикально новое понимание сознания, которое избегало мистификации и предлагало конкретные механизмы его функционирования, что открыло путь для исследований искусственного сознания на функциональной основе.

Концепция сознания человека в определении Сьюзан Шнайдер

Сьюзан Шнайдер — современный американский философ, специализирующаяся на философии сознания, искусственном интеллекте и астробиологии. Её подход к определению сознания имеет ряд отличительных характеристик.

Ключевые аспекты определения сознания по Шнайдер

1. Информационно-теоретический подход

- Сознание связано с определенным типом обработки информации в сложных системах.
- Шнайдер рассматривает сознание через призму интегрированной информации и когнитивных процессов.

2. Биологическая основа

- Признавая биологическую основу человеческого сознания, Шнайдер не считает биологию необходимым условием для сознания.
- Она допускает возможность небиологических форм сознания (искусственный интеллект, инопланетный разум).

3. Феноменальное сознание

- Шнайдер подчеркивает центральную роль феноменального опыта («каково это — быть» субъектом).
- Квалиа (субъективные качества опыта) являются фундаментальным аспектом сознания.

4. Пространство возможностей сознания

- Вводит понятие «пространства возможностей сознания» для описания различных форм, которые может принимать сознание.
- Человеческое сознание представляет лишь одну точку в этом многомерном пространстве.

5. Вычислительная теория разума

- Связывает сознание с определенными вычислительными процессами.
- Однако предостерегает от редукционистских подходов, сводящих сознание только к вычислениям.

6. Интеграция с нейронаукой

- Шнайдер считает важным соотнесение философских теорий сознания с эмпирическими данными нейронауки.
- Придерживается междисциплинарного подхода к изучению сознания.

Применение к вопросам искусственного интеллекта

В своих работах, особенно в книге «Artificial You: AI and the Future of Your Mind», Шнайдер исследует возможность создания искусственного сознания и его последствия. Она поднимает вопрос о том, могут ли синтетические системы обладать подлинным сознательным опытом, и предлагает методологические подходы к определению наличия сознания в небиологических системах.

Шнайдер подчеркивает необходимость глубокого философского анализа природы сознания в контексте современных технологических разработок и стремится создать концептуальный мост между традиционной философией сознания и новейшими достижениями в области ИИ и когнитивных наук.

Концепция сознания человека по Джону Сёрлу

Джон Сёрл, американский философ, разработал одну из наиболее влиятельных современных концепций сознания, оказавшую существенное влияние на философию сознания и дискуссии вокруг искусственного интеллекта.

Ключевые положения теории сознания Сёрла

1. Биологический натурализм

Центральное положение теории Сёрла —**биологический натурализм**: сознание представляет собой реальный биологический феномен, порождаемый процессами в мозге: «Сознание — обычное биологическое свойство мозга, такое же, как пищеварение для желудочно-кишечного тракта или секреция желчи для печени».

Сёрл считает, что сознание:

- онтологически субъективно (существует только с точки зрения первого лица);
- при этом каузально порождается объективными нейробиологическими процессами.

2. Субъективная онтология и первичная реальность сознания

Для Сёрла **субъективность** является определяющей характеристикой сознания:

«Онтология ментального по своей природе субъективна в том смысле, что для ментального состояния быть — значит быть испытываемым некоторым человеческим или животным субъектом».

Сёрл настаивает, что:

- субъективный опыт не может быть редуцирован к объективным процессам в мозге;
- субъективные состояния не менее реальны, чем другие биологические явления;
- сознание нельзя «устранить» путем объективного описания нейронных процессов.

3. Интенциональность и сознание

Интенциональность (направленность ментальных состояний на объекты мира) является фундаментальным аспектом сознания по Сёрлу:

«Интенциональность — это свойство многих ментальных состояний и событий, посредством которого они направлены на объекты и положения дел внешнего мира».

Сёрл выделяет:

- внутреннюю (исходную) интенциональность (*центральное свойство человеческого сознания: быть направленным на некоторый предмет*) — принадлежащую сознанию;
- производную интенциональность — приписываемую знакам, символам, артефактам.

4. Аргумент «Китайской комнаты» и критика компьютерной модели сознания

Сёрл разработал знаменитый мысленный эксперимент «Китайская комната», демонстрирующий разрыв между синтаксисом и семантикой:

«Системы, которые имеют только синтаксис, не имеют семантики. Манипуляция символами не является достаточной для понимания».

Основываясь на этом аргументе, Сёрл критикует:

- **сильный ИИ**: утверждение, что правильно запрограммированный компьютер обладает сознанием;
- **функционализм**: редукцию ментальных состояний к функциональным состояниям;
- **вычислительную теорию разума**: сознание как вычислительный процесс.

5. Различие между сознательным и бессознательным

Сёрл проводит четкое различие между **сознательными и бессознательными ментальными состояниями**:

«Бессознательные ментальные состояния — это такие состояния, которые в принципе доступны сознанию, хотя в данный момент могут быть бессознательными».

Согласно Сёрлу:

- Бессознательные ментальные состояния — это нейрофизиологические состояния мозга, потенциально способные вызвать сознательные состояния.
- Только те нейронные процессы, которые потенциально могут быть осознаны, можно считать ментальными.

6. Каузальные свойства сознания

Важный аспект теории Сёрла —**каузальная действенность сознания**:

«Сознание каузально возникает из нейрофизиологических процессов в мозге и, в свою очередь, может действовать причинно на эти процессы».

Сёрл утверждает, что:

- Сознание не редуцируемо к физическим процессам, но является их каузальным свойством.
- Сознание обладает реальной каузальной силой (не является эпифеноменом).
- Ментальная каузальность не нарушает физическую каузальную замкнутость мира.

7. Единство сознания

Сёрл подчеркивает важность **единства сознательного опыта**:

«В любой момент времени мы не переживаем разрозненные, отдельные сознательные состояния, но единое, объединенное сознательное поле».

Он описывает несколько форм единства:

- Горизонтальное единство (одновременность различных аспектов опыта)
- Вертикальное единство (единство «я» во времени)
- Нейробиологическое единство (интеграция мозговых процессов)

Критика концепции и противоречия

Позиция Сёрла подвергалась критике со стороны:

- **функционалистов**: за недооценку возможности искусственного воспроизведения сознательных функций;
- **элиминативных материалистов**: за сохранение «призрака в машине» в виде нередуцируемого сознания;
- **сторонников ИИ**: за недостаточное понимание вычислительных систем и непризнание возможности сознания у искусственных систем;
- **философов языка**: за смешение онтологических и эпистемологических аспектов сознания.

Значение концепции Сёрла

Вклад Сёрла в понимание сознания значителен:

- он предложил сбалансированную позицию между дуализмом и редукционизмом;

- артикулировал пробел между объективным описанием и субъективным опытом;
- выявил ограничения вычислительных подходов к сознанию;
- подчеркнул биологическую природу сознания при сохранении его феноменологической уникальности;
- создал основу для натуралистического, но не редукционистского подхода к сознанию.

Теория Сёрла остается одной из наиболее влиятельных концепций сознания в современной философии, предлагая компромиссный путь между крайностями дуализма и материализма.

Концепция сознания человека по Дэвиду Чалмерсу

Дэвид Чалмерс — австралийский философ, один из ведущих современных теоретиков сознания. Его концепция оказала решающее влияние на дискуссии о природе сознания в конце XX — начале XXI века.

Основные положения философии сознания Чалмерса

1. Трудная проблема и легкие проблемы сознания

Центральным вкладом Чалмерса является разделение проблем сознания на **«трудную проблему»** и **«легкие проблемы»**.

«Легкие проблемы сознания связаны с объяснением когнитивных функций и поведения. Трудная проблема касается опыта: почему вообще существует нечто такое, как субъективное переживание опыта?»

Легкие проблемы включают объяснение:

- способности различать и реагировать на стимулы;
- интеграции информации;
- отчетности о ментальных состояниях;
- контроля поведения;
- способности различать внутренние состояния.

Трудная проблема — это вопрос о том, почему физические процессы в мозге сопровождаются субъективным опытом:

«Почему обработка информации сопровождается внутренним субъективным переживанием? Почему это не происходит «в темноте», без какого-либо опыта?»

2. Феноменальное и психологическое сознание

Чалмерс проводит фундаментальное различие между двумя аспектами сознания:

Психологическое сознание (A-consciousness):

- функциональные, каузальные аспекты сознания;
- доступность информации для вербального отчета и контроля поведения;
- потенциально объяснимо в рамках функциональных и нейронаучных подходов.

Феноменальное сознание (P-consciousness):

- субъективные переживания, qualia, «каково это» (what it is like);
- сенсорные качества (цвета, звуки, запахи, вкус);
- телесные ощущения (боль, голод);
- эмоции, мысли и т. д.

«Феноменальное сознание — это то, что делает проблему сознания по-настоящему трудной»

3. Аргумент зомби и логическая возможность отсутствия сознания

Чалмерс разработал влиятельный мысленный эксперимент — **аргумент зомби**: «Мы можем представить существо, физически идентичное человеку, но полностью лишенное феноменального сознания — философского зомби. Логическая возможность таких зомби указывает на то, что сознание не логически супервентно на физическом»

Ключевые положения аргумента.

- Философские зомби логически возможны.

- Если зомби возможны, то материализм ложен.
- Сознание не может быть логически выведено из физических фактов.
- Физикалистское объяснение сознания неполно.

4. Натуралистический дуализм и нередуктивный подход

Чалмерс защищает позицию **натуралистического дуализма**: «Сознание является фундаментальным свойством мира, не сводимым к физическим свойствам, но тем не менее подчиняющимся принципам, которые можно изучать и которые могут быть интегрированы в научную картину мира»

Основные характеристики этой позиции.

- Сознание не редуцируемо к физическим свойствам.
- При этом сознание подчиняется принципам и законам.
- Психофизические законы связывают феноменальные и физические свойства.
- Натуралистический подход к нефизическому феномену сознания.

5. Информационная теория сознания

В более поздних работах Чалмерс разрабатывал **двухаспектную информационную теорию сознания**: «Информация имеет два аспекта: физический и феноменальный. Феноменальный опыт является внутренним аспектом определенного вида информационного состояния»

Ключевые положения:

- Информация является фундаментальной характеристикой реальности.
- Феноменальный опыт — внутренний аспект информации.
- Физические процессы — внешний аспект той же информации.
- Сознание возникает там, где информация имеет правильный вид каузальной структуры.

6. Структурная когеренция и принципы психофизических связей

Чалмерс предполагает, что связь между сознанием и физическими процессами подчиняется принципу **структурной когеренции**: «Существует систематическое соответствие между структурой сознания и структурой осознаваемости

(функциональной организацией, обеспечивающей наши отчеты о сознательном опыте)».

Он также предлагает возможные **психофизические принципы**:

- Принцип организационной инвариантности: системы с одинаковой функциональной организацией имеют одинаковые сознательные переживания.
- Принцип структурной когеренции: осознаваемость отражает структуру сознания.
- Принцип двойного аспекта информации: информация имеет феноменальный и физический аспекты.

7. Расширенное сознание и когнитивные расширения

В работах о**расширенном сознании** Чалмерс исследует: Степень, в которой когнитивные процессы, включая сознание, могут выходить за пределы мозга в окружающую среду».

Основные идеи:

- Когнитивные процессы могут включать активное использование внешней среды
- Сознание можно рассматривать как расширенную систему, включающую взаимодействие с миром
- Граница между внутренним и внешним относительна

Критика концепции Чалмерса

Подход Чалмерса вызвал значительные дебаты и критику.

Со стороны материалистов:

- Аргумент зомби основан на интуициях, а не на эмпирических фактах.
- Постулирование нередуцируемого сознания нарушает онтологическую экономность.
- Дуализм создает проблему каузальности: как нефизическое сознание влияет на физический мир.

Со стороны когнитивных нейроученых:

- Трудная проблема может быть решена по мере развития нейронауки.
- Сознание может быть объяснено через интегрированную информацию (теория Тонони) или глобальное рабочее пространство (теория Баарса-Дена).

Со стороны мистических/иллюзионистских подходов:

- Сознание может быть иллюзией (Деннет).
- Сама постановка трудной проблемы может быть ошибочной.

Значение и вклад Чалмерса

Вклад Чалмерса в философию сознания трудно переоценить.

- Формулировка «трудной проблемы» стала центральной точкой современных дискуссий о сознании.
- Аргументы против редукционизма заставили пересмотреть многие подходы к сознанию.
- Работа по информационной теории сознания предложила новую перспективу для научных исследований.
- Ясное разграничение между феноменальным и психологическим сознанием помогло структурировать дебаты.
- Его подход открыл возможность интеграции феноменального сознания в научную картину мира, не редуцируя его к физическим процессам.

Концепция Чалмерса остается одной из наиболее влиятельных в современной философии сознания, предлагая сбалансированный подход между научным натурализмом и признанием уникальной природы субъективного опыт.

Концепция сознания человека по Джулио Тонони

Джулио Тонони — итальянский психиатр и нейробиолог, создавший одну из наиболее влиятельных современных теорий сознания —**теорию интегрированной информации** (Integrated Information Theory, IIT). В отличие от многих предшествующих подходов, теория Тонони представляет собой математически

формализованную концепцию, которая пытается объяснить как качественные, так и количественные аспекты сознания.

Теория интегрированной информации (ТИИ)

1. Основные положения и аксиомы теории

Тонони начинает построение теории с феноменологических аксиом — неопровержимых свойств опыта, которые затем переводит в постулаты об организации физических систем, способных порождать сознание:

Аксиомы опыта:

- **Внутреннее существование (Intrinsic Existence)**: сознательный опыт существует внутренне, для себя, а не только как описание внешним наблюдателем.
- **Композиция (Composition)**: сознательный опыт структурирован — состоит из различных феноменальных различий (qualia), объединенных в единое целое.
- **Информация (Information)**: каждый опыт специфичен, отличается от других возможных опытов — он информативен.
- **Интеграция (Integration)**: сознательный опыт единен, неделим на независимые компоненты — опыт всегда целостен.
- **Исключение (Exclusion)**: каждый опыт имеет определенный состав и масштаб, исключая другие опыты.

«Сознание — это то, что существует для самой системы, а не только для внешнего наблюдателя; что объединяет множество различий в единую сцену; что информативно, интегрировано и определено.» — Джулио Тонони.

2. Интегрированная информация и мера фи (Ф)

Центральное понятие теории Тонони — **интегрированная информация** (Ф, фи), которая определяет как количество, так и качество сознания:

«Интегрированная информация (Ф) — это количество информации, генерируемой системой как целым, сверх информации, генерируемой ее частями независимо друг от друга».

Ключевые положения:

- Ф измеряет, насколько система больше, чем сумма своих частей.
- Ф> 0 означает, что система обладает каузальной мощностью, которой не обладает никакое ее разбиение.
- Чем выше Ф, тем более интегрированной является система и тем более интенсивным является опыт.

Тонони предложил математический аппарат для вычисления Ф, основанный на анализе каузальных взаимодействий элементов системы.

3. Структура сознательного опыта (каузальная репертуарная структура)

По теории Тонони, структура сознательного опыта определяется **каузальными репертуарами** — множествами возможных прошлых и будущих состояний, которые система может различать: «Качество сознательного опыта — то, как оно ощущается — определяется формой интегрированной информации, то есть точной структурой набора различий, которые составляют единое целое в данной системе.»

Основные понятия:

- **Концепты** — элементарные различия или комбинации различий внутри системы.
- **Каузальная репертуарная структура** — геометрическая структура, в которой каждый концепт представлен как точка в многомерном пространстве.
- **Квалиа** — конкретные различия, возникающие в определенных местах структуры.

4. Принципы сознания и их приложения

Теория Тонони включает набор принципов, определяющих взаимосвязь между физической организацией системы и наличием сознания:

Принцип информации. «Сознание соответствует системе, способной различать большое количество возможных состояний в результате своих внутренних взаимодействий.»

Принцип интеграции. «Сознательная система должна быть единой — ее нельзя разделить на независимые подсистемы без потери информации.»

Принцип исключения. «Из всех возможных комплексов на любом пространственно-временном зернистом уровне, только тот, который максимизирует Ф, соответствует сознанию.»

5. Применение теории к мозгу и нейронным сетям

Тонони и его коллеги предложили ряд эмпирических приложений ТИИ к функционированию мозга: «Кора головного мозга идеально подходит для поддержания высокого уровня интегрированной информации благодаря своей сложной архитектуре, сочетающей функциональную специализацию с глобальной интеграцией.»

Ключевые прогнозы теории относительно мозга:

- Таламокортикальная система должна поддерживать максимальные значения Ф.
- Сознание исчезает при снижении функциональной дифференциации или интеграции (например, во время глубокого сна, общей анестезии).
- Локальные повреждения мозга должны изменять содержание сознания, но не обязательно его наличие.

6. Сознание и градации интегрированной информации

Одно из важных следствий теории Тонони — представление о сознании как о континууме: «Сознание не является бинарным свойством, оно существует в различных степенях соответственно уровням интегрированной информации.»

Согласно ТИИ:

- Сложные нейронные сети обладают высокими значениями Ф и сложным сознанием.
- Простые организмы обладают низкими значениями Ф и примитивными формами сознания.
- Искусственные системы могут обладать сознанием в зависимости от их архитектуры и уровня интеграции.

7. Философские следствия теории

Теория Тонони имеет важные философские импликации: «ТИИ отвергает как дуализм, так и редуктивный материализм, предлагая вместо этого информационный монизм, в котором интегрированная информация является фундаментальным свойством реальности.»

Ключевые философские позиции:

- Сознание не требует биологического субстрата — может существовать на любой подходящей физической основе.
- ТИИ совместима с пансихизмом — простейшие формы интегрированной информации могут существовать на фундаментальных уровнях реальности.
- Тонони предлагает онтологию, в которой интегрированная информация имеет такой же фундаментальный статус, как масса или заряд.

8. Практические приложения и эмпирические подтверждения

Теория Тонони привела к ряду практических приложений: «ТИИ привела к созданию индекса сложности сознания — индекса пертурбационной сложности (PCI), который успешно используется для оценки уровней сознания у пациентов.»

Эмпирические подтверждения:

- Исследования показали, что значения PCI систематически различаются во время бодрствования, сна и анестезии.
- PCI позволяет выявить скрытое сознание у неконтактных пациентов.
- Индуцированная магнитной стимуляцией активность мозга имеет различные паттерны при разных состояниях сознания, что согласуется с предсказаниями ТИИ.

9. Критика и дискуссии

Теория интегрированной информации подвергается активной критике:

- **Вычислительная сложность**: вычисление Φ для реальных мозговых систем практически невозможно из-за комбинаторного взрыва.

- **Метрическая несогласованность**: некоторые критики утверждают, что различные версии меры Ф дают разные результаты.
- **Проблема субъекта**: критики указывают, что ТИИ не полностью объясняет субъективный характер опыта.
- **Эпистемологические границы**: возникает вопрос, как мы можем знать о сознании систем с радикально отличающейся от нас архитектурой.

10. Развитие теории и современное состояние

Теория Тонони продолжает развиваться, интегрируя новые данные и концепции: «Последние версии ТИИ (ТИИ 3.0, 4.0) вводят более строгие математические определения и уточняют понимание времени в сознательном опыте.»

Современные направления исследований:

- Изучение динамической природы интегрированной информации во времени.
- Развитие методов эффективного приближенного вычисления Ф.
- Сотрудничество с теоретиками искусственного интеллекта для применения принципов ТИИ к искусственным системам.
- Клинические приложения для диагностики нарушений сознания.

Значение теории Тонони для понимания сознания

Теория интегрированной информации представляет собой уникальную попытку сочетать феноменологический и нейронаучный подходы к сознанию:

- Это одна из немногих теорий, которая предлагает математически формализованное объяснение как количественных, так и качественных аспектов сознания.
- ТИИ преодолевает разрыв между субъективными и объективными аспектами сознания, связывая феноменологию с физическими свойствами систем.
- Она предлагает новое понимание границ сознания в природе, включая возможность примитивных форм сознания у простых организмов и потенциально у искусственных систем.

- Теория имеет практические приложения в клинической нейробиологии и нейрофизиологии, особенно для оценки состояний сознания.

Концепция сознания Тонони продолжает оставаться одной из ключевых теорий в современной науке о сознании, предлагая плодотворную рамку для дальнейших исследований природы субъективного опыта.

Концепция сознания человека в определении Роджера Пенроуза

Роджер Пенроуз [28], лауреат Нобелевской премии по физике 2020 года, предложил одну из наиболее оригинальных теорий сознания, основанную на квантовой механике и математической физике.

Основные положения теории Пенроуза о сознании:

- **Квантовая природа сознания**. Пенроуз полагает, что классические вычислительные модели (включая искусственный интеллект) принципиально не способны воспроизвести человеческое сознание. Он утверждает, что сознание связано с квантовыми процессами в мозге.
- **Объективная редукция (OR)**. Пенроуз предложил новый тип квантового коллапса волновой функции, который происходит из-за гравитационных эффектов на квантовом уровне. Этот процесс, по его мнению, не является ни случайным, ни детерминированным, что создает пространство для «свободы воли».
- **Модель Пенроуза-Хамероффа**. Совместно с анестезиологом Стюартом Хамероффом Пенроуз разработал теорию «оркестрованной объективной редукции» (Orch OR). Согласно этой теории, квантовые вычисления происходят в микротрубочках нейронов, белковых структурах, которые могут поддерживать квантовую когерентность.
- **Не вычислимость сознания**. Пенроуз утверждает, что человеческое понимание и сознание включают невычислимые процессы, которые выходят за рамки алгоритмического мышления. Эту идею он обосновывает, опираясь на теорему Гёделя о неполноте.

– **Связь с платонизмом**. Пенроуз придерживается платонистского взгляда на математику, считая, что математические истины существуют объективно и независимо от человеческого ума. Он предполагает связь между этим математическим платоновским миром, физическим миром и сознанием.

Теория Пенроуза остается спорной в научном сообществе. Многие критики указывают на отсутствие экспериментальных доказательств квантовых эффектов в мозге, а также на то, что тепловая декогеренция (*процесс нарушения когерентности*) должна разрушать квантовые состояния слишком быстро для их функционального использования в нейронных процессах.

В табл. 7 приведен список к**ниг на русском языке по теории сознания применительно к ИИ**

Таблица 7. Книги на русском языке по теории сознания в ИИ

Переводы ключевых зарубежных работ	
Дэвид Чалмерс – «Сознающий ум: в поисках фундаментальной теории»	Перевод классического труда по философии сознания. Включает обсуждение проблемы сознания в искусственном интеллекте. Издательство: УРСС.
Ник Бостром – «Искусственный интеллект: Этапы, угрозы, стратегии»	Анализ перспектив создания сильного ИИ и связанных с этим этических проблем Затрагивает вопросы сознания в контексте суперинтеллекта. Издательство: Манн, Иванов и Фербер.
Дэниел Деннет – «Виды психики: на пути к пониманию сознания»	Исследование эволюции сознания и его возможных форм. Применимость концепций к искусственным системам. Издательство: Идея-Пресс.
Рэй КурцвейлЭволюция разума»	Взгляд на возможность создания. искусственного разума и сознания. Прогнозы развития ИИ и его сближения с человеческим сознанием. Издательство: Эксмо.
Джон Серл – «Сознание, мозг и наука»	Критический анализ компьютерного подхода к сознанию. Разбор мысленного эксперимента «Китайская комната». Издательство: Идея-Пресс.

Дуглас Хофштадтер – «Гедель, Эшер, Бах: Эта бесконечная гирлянда»	Междисциплинарное исследование самореференции как основы сознания. Обсуждение возможности создания самосознающих ИИ-систем. Издательство: Бахрах-М.
Работы Российских авторов	
Д.И. Дубровский – «Проблема сознания: Теория и критика современных концепций»	Анализ современных подходов к проблеме сознания. Рассмотрение информационного подхода к сознанию и его связь с ИИ. Издательство: ЛЕНАНД.
В.А. Лекторский – «Философия, познание, культура»	Включает разделы о проблеме сознания и искусственного интеллекта. Эпистемологический подход к проблеме машинного сознания. Издательство: Наука.
В.В. Васильев – «Трудная проблема сознания	Анализ «трудной проблемы» и различных подходов к ней. Рассмотрение квалиа в контексте ИИ. Издательство: Прогресс-Традиция.
А.Ю. Алексеев – «Понимание искусственного интеллекта»	Философский анализ критериев разумности и сознания применительно к ИИ. Тест Тьюринга и другие методы оценки машинного сознания. Издательство: МГТУ им. Баумана.
Д.В. Иванов – «Природа феноменального сознания»	Исследование квалиа и феноменального опыта. Обсуждение вопроса о возможности квалиа в ИИ. Издательство: ЛИБРОКОМ.
Т.В. Черниговская – «Чеширская улыбка кота Шрёдингера: Язык и сознание»	Междисциплинарный подход к проблеме сознания. Нейролингвистика и ее применение к искусственным системам. Издательство: Языки славянской культуры.
В.Э. Карпенко – «Искусственный интеллект и сознание человека»	Анализ философских аспектов создания искусственного сознания. Сравнение различных парадигм исследования ИИ. Издательство: ЛЕНАНД.
Сборники и коллективные монографии	
«Философия искусственного интеллекта» (под ред. В.А. Лекторского)	Сборник статей по философским проблемам сознания в ИИ. Междисциплинарный подход российских исследователей. Издательство: ИФ РАН.
«Проблема сознания в философии и науке» (под ред. Д.И. Дубровского)	Коллективная монография с разделами о машинном сознании. Различные теоретические подходы к проблеме искусственного сознания. Издательство: Канон+.

В табл. 8. Приведены научные конференции по сознанию в ИИ за 2 последних года.

Таблица 8. Конференции и семинары в РФ

2023 (прошедшие)	
КОНФЕРЕНЦИЯ "СОЗНАНИЕ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ-2023" ("СОЗНАНИЕ-2023")	Проводиться в рамках секции Всемирного конгресса «Теория систем, алгебраическая биология, искусственный интеллект: математические основы и приложения» (далее — Конгресс) Форум "Сознание: от постановки проблем к математическим моделям" Дата: 26-30 июня 2023 г. Направления: URL: Тезисы докладов конференции «Сознание 2023): В конце табл.8
VII Международная научная конференция «Нейрокомпьютерный интерфейс: наука и практика»	Дата: 10-13 октября 2023 года Место: Самарский национальный исследовательский университет Секция: «Философские и этические аспекты искусственного сознания» Организаторы: Самарский университет, ИПУ РАН
Международная конференция «Искусственный интеллект и цифровые технологии в образовании и науке»	Дата: сентябрь 2023 года Место: МФТИ, Москва Секция: «Феноменальное сознание и машинный интеллект» Организаторы: МФТИ, НЦМУ «Центр исследований искусственного интеллекта»
Всероссийская конференция с международным участием «Философия искусственного интеллекта» (ФИИ-2023)	Дата: 13-14 марта 2023 года Место: Институт философии РАН, Москва Специальная секция: «Проблемы искусственного сознания: вызовы и перспективы» Организаторы: Институт философии РАН, НИУ ВШЭ, МГУ имени М.В. Ломоносова
2024 (запланированные и текущие)	
КОНФЕРЕНЦИЯ "СОЗНАНИЕ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ-2024" ("СОЗНАНИЕ-2024")	Проводилась в рамках секции Всемирного конгресса «Теория систем, алгебраическая биология, искусственный интеллект: математические основы и приложения» (далее — Конгресс) Форум "Сознание: от постановки проблем к математическим моделям" Дата: 26 по 30 августа 2024 г. Направления: URL: Тезисы докладов конференции «Сознание 2024): В конце табл.8
VIII Международная конференция «Нейрокомпьютерный интерфейс: наука и практика — 2024»	Дата: октябрь 2024 года (предварительно) Место: Самарский университет Секция: «Нейроморфные системы и теория машинного сознания» Организаторы: Самарский университет, РАН

Всероссийская конференция «Философия искусственного интеллекта» (ФИИ-2024)	Дата: март-апрель 2024 года Место: Институт философии РАН, Москва Тема: «Сознание и интеллект в человеко-машинных системах» Организаторы: ИФ РАН, НСМИИ РАН
Международная научно-практическая конференция «Этика искусственного интеллекта»	Дата: май 2024 года Место: Санкт-Петербургский государственный университет Секция: «Правовые и этические аспекты моделирования сознания в искусственных системах» Организаторы: СПбГУ, Центр этики ИИ
Международный форум «AI Journey 2024»	Дата: ноябрь 2024 года Место: Москва Специальная секция: «Перспективы создания искусственного сознания» Организаторы: Сбер, АНО «Цифровая экономика»
Московская международная конференция по когнитивной науке (MECCogSci 2024)	Дата: июнь 2024 года Место: Москва, НИУ ВШЭ Секция: «Теории сознания и их применение в искусственных системах» Организаторы: НИУ ВШЭ, МГУ
2025 (предварительно анонсированные)	
КОНФЕРЕНЦИЯ "СОЗНАНИЕ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ-2025" ("СОЗНАНИЕ-2025")	Проводиться в рамках секции Всемирного конгресса «Теория систем, алгебраическая биология, искусственный интеллект: математические основы и приложения» (далее — Конгресс) Форум "Сознание: от постановки проблем к математическим моделям" Дата: 08-12 сентября 2025 год URL: https://congrsysalgbai.ru/ru/
«AI Ethics Russia 2025»	Дата: апрель 2025 года (предварительно) Место: Сколтех, Москва Секция: «Машинное сознание: этические и правовые аспекты» Организаторы: Сколтех, РАНХиГС
Всероссийская конференция с международным участием «Философия искусственного интеллекта» (ФИИ-2025)	Дата: март-апрель 2025 года (предварительно) Место: Институт философии РАН, Москва Тематика: «Феноменальный опыт и квалиа в системах искусственного интеллекта» Организаторы: ИФ РАН, НСМИИ РАН
Международный форум «AI Journey 2025»	Дата: ноябрь 2025 года Место: Москва «Сверхразум и сознание: технологические рубежи» Организаторы: Сбер, АНО «Цифровая экономика»
Постоянно действующие семинары	
Научный семинар «Проблемы искусственного сознания»	Периодичность: ежемесячно Место: Институт философии РАН, Москва Координаторы: Д.И. Дубровский, В.А. Лекторский
Междисциплинарный семинар «Философия сознания и ИИ»	Периодичность: раз в два месяца Место: МГУ имени М.В. Ломоносова, философский факультет Координаторы: В.В. Васильев, Д.В. Иванов

1. Тезисы докладов конференции «Сознание 2023» https://congrsysalgbai.ru/media/files_to_download/%D0%91%D1%83%D0%BA%D0%BB%D0%B5%D1%82-%D0%BE%D0%B1%D0%BB%D0%B6-%D0%BF%D1%80%D0%B8%D0%BB%D0%BE%D0%B6_9eyLnQk.pdf можно посмотреть в библиотеке сайта vikchas.ru
2. Тезисы докладов конференции «Сознание 2024» https://congrsysalgbai.ru/media/files_to_download/%D0%A1%D0%9E%D0%97%D0%9D%D0%90%D0%9D%D0%98%D0%952024-%D0%A2%D0%95%D0%97%D0%98%D0%A1%D0%AB-%D0%BB%D0%BE%D0%B3%D0%BE-%D0%BE%D0%B1%D0%BB-%D1%8D%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D0%BD1908.pdf можно посмотреть в библиотеке сайта vikchas.ru

Авторы придерживаются концепции интеллекта по И. Канту (рассмотрели ранее) и провели исследование о возможности и не возможности взаимодействовать сознания Канта с ИИ. Авторы убеждены что, вопрос о возможности взаимодействия сознания с ИИ представляет особый интерес, поскольку затрагивает фундаментальные вопросы о природе познания, разума и возможностях технологического воспроизведения когнитивных функций. Рассмотрим, как кантианская концепция сознания Канта соотносится с современными системами ИИ.

Для Канта сознание неразрывно связано с самосознанием и трансцендентальным единством апперцепции – способностью осознавать себя как единый субъект всех своих представлений. В основе всех познавательных актов лежит «Я мыслю», которое должно сопровождать все наши представления.

Основными аспектами понимания сознания Кантом, значимые для рассмотрения вопроса об ИИ:

- **Спонтанность мышления**– сознание не просто пассивно воспринимает данные, но активно организует их.
- **Единство апперцепции**– все представления объединяются в единое сознание.

- **Априорные структуры** – сознание применяет врожденные категории и формы чувственности.
- **Самосознание** – способность к рефлексии над собственными ментальными состояниями.
- **Практический разум** – моральная автономия и способность действовать согласно представлениям о должном.

Возможности взаимодействия сознания Канта с современным ИИ определены в табл. 9.

Таблица 9. Возможности взаимодействия сознания Канта с современным ИИ

Возможное взаимодействие	
Формальные структуры познания	Кантианские категории рассудка (причинность, субстанция, единство и т. д.) и формы чувственности (пространство и время) могут быть формализованы и имплементированы в алгоритмы ИИ. Современные нейросети и другие системы машинного обучения уже демонстрируют способность распознавать причинно-следственные связи, пространственно-временные отношения и категоризировать объекты – то есть, в определенном смысле применять аналоги кантианских априорных форм к данным.
Синтетическая деятельность	Кант описывает познание как активный синтез, объединяющий чувственные данные и рассудочные категории. Современные генеративные модели ИИ (например, GPT, DALL-E, MidJourney) тоже занимаются своего рода «синтезом», создавая новые тексты, изображения и другой контент на основе обученных паттернов. Это можно рассматривать как аналог синтетической деятельности рассудка.
Возможное взаимодействие	
Систематизация знаний	Кантианский разум стремится к систематическому единству знаний. Современные системы ИИ способны интегрировать разрозненные данные в связные системы и выявлять паттерны в больших массивах информации, что соответствует регулятивной функции разума по Канту.
Ограничения невозможности	
Отсутствие подлинного самосознания	Для Канта ключевым аспектом сознания является самосознание – трансцендентальное единство апперцепции. Современные системы ИИ, даже если они могут генерировать высказывания о себе («Я думаю», «Я считаю»), не обладают подлинным самосознанием. Они не имеют точки зрения «от первого лица» и субъективного опыта, который Кант считал неотъемлемой частью сознания.

Проблема свободы и автономии	Кантианское понимание разума неразрывно связано с идеей свободы и моральной автономии. Практический разум действует на основе категорического императива, а не только причинной детерминации. Системы ИИ, будучи полностью детерминированными своими алгоритмами и данными, не обладают свободой в кантианском смысле и не могут быть морально автономными субъектами
Ограничения невозможности	
Ноуменальное измерение (данные можно классифицировать без присвоения количественного значения	Кант проводил фундаментальное различие между феноменами (явлениями) и ноуменами (вещами в себе). Человеческое сознание, согласно Канту, принадлежит не только миру явлений, но и ноуменальному миру, что делает возможной свободу. ИИ, как технологический артефакт, целиком принадлежит к феноменальному миру и не имеет ноуменального измерения.
Отсутствие интуиции	Созерцание у Канта связано с непосредственной интуицией, которая всегда имеет чувственный характер. ИИ не обладает чувственностью в кантианском понимании – у него нет непосредственного восприятия мира через органы чувств и соответствующего субъективного опыта.
Формы возможного взаимодействия *(Несмотря на фундаментальные ограничения, можно выделить несколько форм взаимодействия кантианского сознания с ИИ)*	
Инструментальное взаимодействие	Сознание по Канту может использовать ИИ как инструмент для расширения своих познавательных возможностей. ИИ может помогать в обработке данных, выявлении закономерностей и систематизации знаний, что соответствует регулятивной функции разума.
Формы возможного взаимодействия *(Несмотря на фундаментальные ограничения, можно выделить несколько форм взаимодействия кантианского сознания с ИИ)*	
Эпистемологическое взаимодействие	ИИ может служить моделью для лучшего понимания отдельных аспектов человеческого познания. Например, изучение алгоритмов машинного обучения может пролить свет на то, как функционируют определенные аспекты категориального синтеза, описанного Кантом
Диалогическое взаимодействие	ИИ может выступать как собеседник в диалоге, помогая человеческому сознанию в процессе рефлексии и самопознания. Хотя сам ИИ не обладает самосознанием, он может стимулировать и структурировать процесс самопознания человека.
Дополненное сознание	Современные технологии создают предпосылки для формирования «дополненного сознания», где часть познавательных функций делегируется ИИ. С кантианской точки зрения, это расширяет сферу феноменального познания, хотя и не затрагивает ноуменальное измерение.

Теоретические перспективы и философские следствия в случае применения сознания Канта в ИИ:

Неокантианский подход к ИИ. Развитие ИИ может стимулировать формирование неокантианского подхода к искусственному познанию, который будет исследовать априорные структуры, используемые алгоритмами машинного обучения. Это может привести к лучшему пониманию того, какие категории и схемы являются необходимыми для любого познания – человеческого или машинного.

Пересмотр границы феноменального и ноуменального. Взаимодействие с ИИ может провоцировать пересмотр кантианского разделения на феноменальное и ноуменальное. Если системы ИИ будут демонстрировать все более сложное поведение, сходное с проявлениями свободы и автономии, это может привести к переосмыслению традиционных кантианских категорий.

Этические импликации. Кантианская этика, основанная на автономии воли и категорическом императиве, ставит вопрос о статусе ИИ в моральной сфере. Если ИИ не обладает автономией в кантианском смысле, он не может быть моральным субъектом, но может ли он быть объектом моральных обязательств? Этот вопрос требует дальнейшего развития кантианской этики в контексте новых технологий.

С точки зрения философии Канта, взаимодействие человеческого сознания с ИИ возможно, но имеет фундаментальные ограничения. ИИ может воспроизводить определенные формальные аспекты познания, описанные Кантом, но ему недоступны такие ключевые элементы кантианского сознания как подлинное самосознание, свобода, моральная автономия и ноуменальное измерение.

Тем не менее, это взаимодействие открывает новые перспективы как для развития ИИ, так и для более глубокого понимания человеческого сознания. Возможно, именно в диалоге с ИИ мы можем лучше понять, что делает человеческое сознание уникальным с кантианской точки зрения, и какие аспекты нашего познания могут быть формализованы и воспроизведены технологически.

Вообще, философия Канта может служить важным ориентиром в осмыслении границ и возможностей ИИ, обращая внимание на те аспекты сознания, которые выходят за рамки чисто функционального и алгоритмического понимания разума.

Важным и актуальным является вопрос в исследование авторов – может ли существовать интеллект без априорных знаний в кантовской системе?

По Канту, интеллект (рассудок) принципиально не может функционировать без априорных форм. Это философское положение является не просто характеристикой интеллекта, но его конститутивным свойством, без которого интеллект как таковой невозможен. Рассмотрим в табл. 10 основные аргументы:

Таблица 10. Априорные формы и интеллект Канта

Свойства	Назначение
Интеллект как синтезирующая деятельность	У Канта интеллект (**рассудок**) представляет собой активную синтезирующую способность, которая обрабатывает чувственное многообразие. Без априорных категорий рассудка эта синтезирующая деятельность была бы невозможна, поскольку: – восприятия оставались бы разрозненными впечатлениями без связи между собой; – отсутствовала бы возможность формирования понятий и суждений; не было бы единства опыта как такового. Кант пишет в «Критике чистого разума»: «Мысли без содержания пусты, созерцания без понятий слепы». Эта знаменитая формула указывает на необходимую взаимосвязь между чувственностью и рассудком, где именно априорные формы делают возможным синтез.
Единство апперцепции как условие интеллекта (*апперцепция означает осмысленное, внимательное и вдумчивое восприятие*)	Трансцендентальное единство апперцепции («Я мыслю»), которое Кант считает «высшим принципом всего человеческого познания», представляет собой априорное условие любого познавательного акта. Без этого единства самосознания: – невозможно было бы присвоение различных представлений единому сознающему субъекту; – отсутствовало бы единство опыта во времени; – невозможно было бы самосознание как таковое. Таким образом, единство апперцепции является априорным условием, без которого интеллект немыслим.
Категории как необходимые инструменты мышления	Двенадцать категорий рассудка (единство, множество, причинность и т. д.) не являются произвольными конструкциями, но представляют собой необходимые инструменты мышления. Категории — это не просто дополнение к интеллекту, а его неотъемлемая структура. Кант утверждает, что категории: – не выводятся из опыта, а предшествуют ему;

Свойства	Назначение
	– являются условиями возможности опыта; – имеют всеобщий и необходимый характер. Без этих категорий, по Канту, невозможно никакое познание объектов, а следовательно, невозможен и интеллект как познающая способность.

Если представить интеллект без априорных структур, не альтернативная форма интеллекта, а его отсутствие. Такой «интеллект» имел бы следующие характеристики:

- **Неспособность к формированию понятий**. Без априорных категорий невозможно было бы объединять разные представления под общими понятиями.
- **Отсутствие причинных связей**. Без априорной категории причинности воспринимаемые явления представлялись бы как случайная последовательность, не связанная необходимой связью.
- **Невозможность математического мышления**. Без априорных форм пространства и времени невозможна была бы математика как наука об априорных синтетических суждениях.
- **Фрагментация опыта**. Отсутствие единства апперцепции привело бы к фрагментации сознания, при которой невозможно было бы говорить о едином «я», осуществляющем познание.
- **Невозможность суждений**. Интеллект по своей функции выносит суждения, но без априорных форм суждения как соединения понятий были бы невозможны.

Кант отрицал, что эмпирический рассудок, не может оперировать только апостериорными понятиями, поскольку:

- Даже эмпирические понятия формируются с помощью априорных категорий (например, категории субстанции и свойства).
- Логические функции суждения, лежащие в основе мышления, имеют априорный характер.

– Синтез представлений, необходимый для формирования любых понятий, осуществляется при помощи продуктивного воображения, действующего по априорным правилам.

Таким образом, чисто эмпирический интеллект, с точки зрения Канта, представляет собой противоречие в терминах. С точки зрения кантовской философии, интеллект без априорных знаний невозможен в принципе. Априорные формы чувственности (пространство и время) и рассудка (категории) не просто дополняют интеллект, а конституируют его как таковой. Они являются не содержанием познания, а его необходимыми условиями и формами.

Если мы устраняем априорные элементы из кантовской модели познания, то получаем не альтернативный тип интеллекта, а разрушение самой возможности интеллектуального познания. Тогда познающий субъект был бы неспособен к синтезу, к формированию понятий, к единству сознания и, следовательно, к знанию как таковому. Таким образом, в кантовской системе интеллект и априорные знания неразрывно связаны: интеллект без априорных форм невозможен так же, как невозможно зрение без способности видеть или слух без способности слышать.

1.4. Архитектура и свойства ЦЭВМ в концепции мозга человека

Джон Фон Нейман создал машину фон Неймана, которая была и остается фундаментальной архитектурой всех вычислительных машин.

Архитектура ЭВМ фон Неймана: фундамент современных ЭВМ

Архитектура фон Неймана — это основополагающая концепция организации вычислительных устройств, разработанная математиком Джоном фон Нейманом в 1945 году. Эта архитектура стала революционным прорывом в истории ЭВМ и до сих пор является базовой для большинства современных ЭВМ.

Основные принципы архитектуры фон Неймана

Архитектура фон Неймана характеризуется несколькими ключевыми принципами:

1. **Единое хранилище данных и программ.** В отличие от ранних ЭВМ, где программы и данные хранились раздельно, в архитектуре фон Неймана и программы, и данные находятся в одной и той же памяти. Это позволяет ЭВМ обрабатывать программы как данные и наоборот (следует заметить, что это идея Ч. Беббиджа, определённая им более двухсот лет).
2. **Линейная организация памяти**. Память представляет собой последовательность ячеек с уникальными адресами, что обеспечивает прямой доступ к любому элементу.
3. **Центральный процессор**. Включает арифметико-логическое устройство (АЛУ) для выполнения операций и устройство управления для координации работы всех компонентов.
4. **Последовательное выполнение инструкций**. ЭВМ выполняет программные инструкции одну за другой в определенном порядке.

Основные компоненты архитектуры фон Неймана

ЭВМ, построенный по архитектуре фон Неймана, включает в себя следующие компоненты:

- **Центральный процессор (CPU)**: выполняет инструкции и управляет процессом вычислений.
- **Память (RAM)**: хранит данные и программы.
- **Устройства ввода-вывода**: обеспечивают взаимодействие с внешним миром.
- **Шина**: Система для передачи данных между компонентами.

Значение архитектуры фон Неймана

Архитектура фон Неймана имеет фундаментальное значение в развитии ЭВМ техники:

- Она стала основой для разработки первых электронных ЭВМ общего назначения.
- Позволила создавать хранимые программы, что сделало ЭВМ более гибкими.

- Обеспечила возможность выполнять любые вычисления без физической перенастройки машины.
- Заложила принципы, которым следуют даже самые современные ЭВМ.

Современное применение

Несмотря на то, что с момента создания архитектуры фон Неймана прошло более 75 лет, ее основные концепции остаются актуальными. Большинство современных ЭВМ, включая персональные ЭВМ, серверы, смартфоны и встроенные системы, основаны на принципах фон Неймана, хотя и с значительными усовершенствованиями:

- многоядерные процессоры;
- кэш-память нескольких уровней;
- параллельная обработка данных;
- виртуальная память.

Ограничения и альтернативы

Основным ограничением архитектуры фон Неймана является так называемое «узкое горло фон Неймана» — необходимость последовательной передачи данных между процессором и памятью, что ограничивает производительность. Существуют альтернативные подходы, такие как архитектура Гарвардская (с раздельными шинами для инструкций и данных) и различные параллельные архитектуры, но даже они часто инкорпорируют основные принципы, заложенные фон Нейманом.

Архитектура фон Неймана остаётся одним из самых значительных вкладов в ЭВМ науку, обеспечивая фундамент для развития цифровых технологий, которые мы используем каждый день.

Почему Фон Нейман предлагает рассматривать методы работы мозга как совокупность вычислений?

Джон фон Нейман, выдающийся математик и пионер ЭВМ наук, предложил рассматривать работу мозга как совокупность вычислений по нескольким фундаментальным причинам.

Исторический и научный контекст

В период 1940-50-х годов, когда фон Нейман разрабатывал свои идеи, формировалась новая научная парадигма, объединяющая кибернетику, теорию информации и когнитивные науки. В этом контексте фон Нейман (особенно в своей работе «ЭВМ и мозг», опубликованной посмертно) выдвинул идеи о параллелях между мозгом и вычислительными машинами.

Основные аргументы фон Неймана:

1. **Логическая и математическая природа мышления**: Фон Нейман заметил, что мозг выполняет логические операции и обработку информации подобно тому, как это делают вычислительные устройства. Человеческое мышление включает абстрактные рассуждения, которые можно представить в виде математических и логических операций.
2. **Нейроны как логические элементы**: Фон Нейман проводил аналогию между нейронами мозга и логическими элементами в ЭВМ. Он видел, что нейроны, подобно электронным схемам, могут находиться в состоянии возбуждения или покоя, эффективно реализуя бинарную логику.
3. **Информационный подход**: Он считал, что и мозг, и ЭВМ обрабатывают информацию, кодируя, передавая и декодируя сигналы. Этот единый информационный подход позволял описывать работу сознания через вычислительные процессы.
4. **Функциональные аналогии**: Фон Нейман отмечал, что мозг и ЭВМ демонстрируют сходные функциональные возможности — память, обработка данных, принятие решений — хотя и реализуются они по-разному.

Эта концепция фон Неймана имела революционное значение для нескольких научных областей:

- **Когнитивная наука**: стимулировала развитие вычислительной теории разума.
- **Искусственный интеллект**: заложила теоретическую основу для создания систем ИИ.

- **Нейробиология**: способствовала развитию вычислительной нейробиологии.
- **Философия сознания**: повлияла на функционалистский подход к пониманию сознания.

Вычислительный подход фон Неймана к работе мозга имеет определенные ограничения:

- Мозг обладает значительно большей параллельностью, чем ранние ЭВМ.
- Биологические системы демонстрируют пластичность и адаптивность иного рода.
- Сознание и субъективный опыт сложно объяснить только вычислительными процессами.
- Современные нейробиологические исследования выявляют процессы, не имеющие прямых аналогий в классических вычислениях

Сегодня идеи фон Неймана получили развитие в таких областях, как коннекционизм, нейроморфные вычисления и глубокое обучение. Хотя его первоначальная концепция претерпела значительные изменения, фундаментальная идея о том, что когнитивные процессы можно рассматривать как вычисления, остается центральной для многих направлений исследований мозга и искусственного интеллекта.

Подход фон Неймана проложил путь к междисциплинарному пониманию мозга и сознания, объединяющему информатику, математику и нейробиологию, что делает его одним из самых влиятельных мыслителей в области понимания человеческого мышления через призму вычислений.

Почему фон Нейман считал, что мозг наилучший пример интеллектуальной системы, которым располагает человечество?

Джон фон Нейман рассматривал человеческий мозг как наилучший пример интеллектуальной системы по ряду глубоких причин, формирующих ключевой аспект его научного мировоззрения. Фон Нейман признавал мозг эталоном интеллектуальной системы благодаря его исключительным характеристикам:

1. **Непревзойденная сложность и интеграция.** Человеческий мозг содержит около 86 миллиардов нейронов с триллионами синаптических связей, образующих невероятно сложную и высоко интегрированную систему. В своей работе «ЭВМ и мозг» фон Нейман подчеркивал, что эта сложность на порядки превосходила любые искусственные системы его времени.
2. **Энергоэффективность**. Фон Нейман был поражен тем, что мозг потребляет около 20 Вт энергии при выполнении сложнейших когнитивных задач — эффективность, недостижимая для ЭВМ даже сегодня. Он отмечал, что это фундаментальное преимущество биологических вычислений.
3. **Параллельная обработка информации**. В отличии от последовательной архитектуры ранних ЭВМ, мозг осуществляет массивную параллельную обработку данных. Фон Нейман видел в этом ключевую особенность, позволяющую мозгу эффективно решать сложные задачи.
4. **Адаптивность и обучаемость**. Мозг демонстрирует уникальную способность к самообучению и адаптации, что фон Нейман считал высшим проявлением интеллектуальных систем. Нейро пластичность позволяет мозгу перестраиваться под различные задачи без внешнего программирования.

Фон Нейман признавал, что мозг – результат миллионов лет эволюционного отбора. Он рассматривал этот процесс как «эксперимент», который природа проводила гораздо дольше и масштабнее, чем любые человеческие усилия по созданию искусственных интеллектуальных систем:

- Мозг был «оптимизирован» эволюцией для выживания, что требует решения широчайшего спектра задач.
- Естественный отбор способствовал развитию мозга как универсального инструмента для решения непредсказуемых проблем.
- Эволюционные механизмы создали систему, способную к творчеству, абстрактному мышлению.

Для фон Неймана, математика и прагматика, имело значение, что мозг – единственная известная система, способная:

- развивать математические теории и доказывать теоремы;
- создавать произведения искусства и музыки;
- проектировать сложные технические системы, включая ЭВМ;
- понимать абстрактные концепции и создавать новые;
- адаптироваться к неизвестным ранее ситуациям.

Рассматривая мозг как эталон, фон Нейман руководствовался и методологическими соображениями:

- **Обратная инженерия** — Изучение мозга может дать ключи к созданию продвинутых искусственных интеллектуальных систем.
- **Измеримый стандарт** — Мозг предоставляет конкретный эталон, с которым можно сравнивать любые искусственные системы.
- **Доказательство возможности** — Существование мозга доказывает, что высокоуровневый интеллект физически возможен в материальной вселенной.

Философские импликации

Взгляд фон Неймана имел и философское измерение:

- Признание мозга эталоном не противоречило его убеждению, что механические системы могут воспроизводить интеллектуальные процессы
- Он считал, что понимание принципов работы мозга может изменить наше представление о природе разума и сознания
- Это представление воплощало его веру в возможность формализации и математического описания сложных природных процессов

Признавая мозг наилучшим примером интеллектуальной системы, фон Нейман заложил основы для современных нейроморфных вычислений, нейросетевых архитектур и когнитивной науки, продолжающих развиваться на основе его фундаментального понимания уникального места мозга как эталонной интеллектуальной системы.

1.5. ЕИ и свойства для реализации ИИ

Проведенное исследование теорий, утверждений, взглядов, практический опыт нейрохирургов позволяет определить следующие утверждения.

- Интеллект, разум, интеллектуальна деятельность не являются материальными структурами головного мозга человека, следовательно биологическая природа интеллекта, сознания головного мозга человека пока не изучены и не определены. **Мозг не алгоритмизирован**.
- Попытки моделирования головного мозга человека и форме искусственных нейронных сетей без учета, как минимум биологичкой природы головного мозга человека являются примитивными, и по сути это обычные информационные технологи, но эти технологии весьма эффективны.

Также возникает важный философский вопрос о природе интеллекта в биологических и искусственных системах.

Искусственные нейронные сети (ИНС) действительно вдохновлены структурой мозга, между ними есть существенные различия:

1. **Физическая природа интеллекта мозга**.
   - Биологический интеллект возникает из сложных электрохимических процессов в нейронной сети мозга.
   - Включает около 86 миллиардов нейронов с триллионами синапсов.
   - Работает параллельно, с использованием различных нейромедиаторов.
   - Обладает сознанием, субъективным опытом и квалиа.
2. **Физическая природа интеллекта ИНС**.
   - Представляет собой математические вычисления в электронных схемах.
   - Основан на матричных операциях и функциях активации.
   - Работает по заданным алгоритмам обучения.
   - Не обладает сознанием или субъективным опытом.

Строгого объективного определения интеллекта не существует ни для биологических, ни для искусственных систем.

Интеллект часто определяется функционально – через способности к обучению, решению проблем, адаптации и т. д. Очевидно, что ИНС — это упрощен-

ная модель, которая воспроизводит лишь некоторые аспекты работы мозга. Более правильно, искусственная нейронная сеть — это упрощенная модель головного мозга и всё.

Тогда правильно, ИНС это модель мозга и все.

Получается, что ИНС — это действительно лишь очень упрощенная математическая модель, вдохновлённая некоторыми базовыми принципами работы биологического мозга. Важно понимать следующие ограничения.

- ИНС имитируют только самые базовые принципы организации нейронов и их связей.
- Современные ИНС не воспроизводят большинство сложных биологических процессов мозга.
- Они не обладают сознанием или субъективным опытом.
- Отсутствуют многие ключевые механизмы мозга (нейропластичность, различные типы нейронов, глиальные клетки и т. д.).

Говорить о том, что ИНС «создают интеллект» в том же смысле, что и мозг человека, было бы преувеличением. Они демонстрируют определённые функциональные способности, которые могут напоминать некоторые аспекты интеллекта, но это всё-таки математические модели, выполняющие вычисления по заданным алгоритмам.

В современных дискуссиях о сознании и интеллекте часто наблюдается разрыв между нейробиологическими исследованиями и философскими/ теоретическими построениями.

Это можно объяснить несколькими причинами:

1. **Междисциплинарный разрыв**. Специалисты в философии сознания часто недостаточно погружены в детали нейрохимии и функциональной нейроанатомии, в то время как нейробиологи могут не углубляться в философские аспекты проблемы.
2. **Проблема «трудного вопроса сознания»**. Как отмечал философ Дэвид Чалмерс [20], даже полное понимание нейрохимических процессов не объясняет, почему эти процессы сопровождаются субъективным опытом.
3. **Различные уровни описания**. Рассуждения об интеллекте и сознании могут вестись на разных уровнях абстракции — от молекулярного до системного и феноменологического.

4. **Конференционный формат**. На крупных конференциях часто предпочтение отдаётся концептуальным докладам, а не детальным обсуждениям нейрохимических механизмов.

Когда многие специалисты в области математики, компьютерных наук и ИИ формулируют теории о работе мозга и природе сознания, не имея практического опыта работы с реальным мозгом. Они зачастую:

- не проводили нейрофизиологических экспериментов;
- не работали с нейронными тканями в лаборатории;
- не наблюдали последствия повреждений различных участков мозга;
- не изучали детально нейрохимию и нейроанатомию.

Это приводит к созданию упрощенных моделей и метафор, которые могут серьезно искажать наше понимание мозга.

Компьютерные модели мозга часто основаны на вычислительной парадигме, которая может не отражать фундаментальные принципы работы биологического мозга.

При этом формулируются грандиозные теории сознания и интеллекта, которые могут звучать убедительно, но на самом деле построены на шатком эмпирическом фундаменте.

Для преодоления этого разрыва необходимо более тесное междисциплинарное сотрудничество между нейрофизиологами, когнитивными психологами, философами и специалистами по ИИ, где приоритет отдавался бы эмпирическим данным о реальной работе мозга и тогда будет достигнута мечта А. Коновалова – он, нейрохирург, проведший тысячи операций на мозге, непосредственно наблюдает парадокс: он видит и работает с физической тканью мозга, но не может «увидеть» мысли, сознание или интеллект в этой ткани. Это классическая проблема «психофизического дуализма».

Мечта А. Коновалова [7] «понять, как из материального возникает духовное» — это, возможно, величайшая научная и философская проблема нашего времени, находящаяся на границе науки, философии и, возможно, даже духовных изысканий.

ГЛАВА 2. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

2.1. История развития ИИ

Историю развития искусственного интеллекта начнем с работ Чарльза Бэббиджа, который в первой половине XIX века заложил концептуальные основы вычислительных машин. Его ключевые разработки:

Разностная машина (1822) — автоматическое устройство для вычисления полиномиальных функций, предназначенное для создания математических таблиц.

Аналитическая машина (1834) — первое в истории концептуальное программируемое вычислительное устройство общего назначения, содержавшее:

- операционный блок («мельница»);
- память («склад»);
- систему ввода/вывода на основе перфокарт;
- возможность условных переходов и циклов.

Хотя Бэббидж не использовал термин «искусственный интеллект», его работы стали фундаментом для создания устройств, способных выполнять сложные вычисления, что является необходимым условием для развития ИИ.

В работах Беббиджа помогала Ада Лавлейс [30], сотрудничавшая с Бэббиджем, сделала несколько революционных предположений:

- разработала первые в истории алгоритмы для Аналитической машины;
- предсказала, что вычислительные устройства смогут манипулировать не только числами, но и символами;
- предвидела, что машины смогут создавать музыку и графику.

Ада Лавлейс создала описание вычислительной машины Бэббиджа и написала 1843 году первую в мире программу для этой машины. Она считается первым программистом в истории человечества. **Возражала относительно способности вычислительных машин мыслить.**

Ада Лавлейс сформулировала важное ограничение, известное как «Возражение Лавлейс»: «машины не могут создавать подлинно новые знания, а лишь следуют инструкциям, заданным человеком.» Это положение впоследствии стало одним из центральных вопросов в философии ИИ.

Идеи, реализованные Чарльзом Бэббиджем в его «Аналитической машине», проложили дорогу дальнейшему внедрению логико-алгоритмических и статистико-вероятностных подходов в вычислительную практику, что отчётливо проявилось в трудах Алана Тьюринга и других пионеров середины XX столетия. Именно тогда стартовал углублённый анализ парадигм зарождавшегося искусственного интеллекта. Таким образом, фундаментальные концепции и технические механизмы, сформулированные в XIX столетии, создали прочную основу для последующих фаз эволюции искусственного интеллекта.

В этот же период были созданы логические и математические основы, необходимые для развития ИИ:

- **Джордж Буль** [17] разработал булеву алгебру, которая стала языком логических операций в компьютерах;
- **Готлоб Фреге** [17] создал исчисление предикатов, формализующее логические рассуждения;
- **Бертран Рассел** и **Альфред Норт Уайтхед** опубликовали «Principia Mathematica» [17], где совершили попытку всеобщей формализации всей математики, свести математику к логике.

В 40-х годах прошлого века началась практическая реализация идей Бэббиджа в виде электронных вычислительных устройств.

В 1943 году была создана математическая модель биологического **нейрона**, предложена Уорреном Мак-Каллоком (нейробиолог) и Уолтером Питтсом (логик).

Некоторые из базовых вычислительных моделей, применяемых в нейросетях, были рассмотрены нейрофизиологом Уорреном Маккалоком и логиком Уолтером Питтсом в статье 1943 г. «Логическое исчисление идей, относящихся к нервной активности», опубликованной в «Бюллетене математической биофизики».

В 1957 г. века американский нейрофизиолог Фрэнк Розенблатт разработал математическую модель персептрон – модель восприятия информации мозгом человека.

В 1960 г. модель перцептрона впервые реализованная в виде электронной машины Марк-1.

23 июня 1960 года в Корнеллском университете был продемонстрирован нейрокомпьютер, который был способен распознавать некоторые буквы английского алфавита.

Формальное рождение искусственного интеллекта как научной дисциплины произошло в 1956 году на Дартмутском семинаре, организованном Джоном Маккарти, Марвином Минским, Клодом Шенноном и Натаниэлем Рочестером. Здесь впервые был использован термин «искусственный интеллект».

В 50-е годы прошлого века искали способы, которыми наделить компьютеры интеллектом и опровергнуть заключение Ады Байрон о неспособности машины мыслить творчески. Джон фон Нейман сформулировал ключевые принципы компьютера, каким мы знаем его сегодня, а его машина была и остается основной моделью машинных вычислений. Фон Нейман рассматривал человеческий мозг, как наилучший пример интеллектуальной системы, которым располагает человечество. Он говорил, если мы сможем изучить его методы, мы сможем использовать биологические парадигмы для создания более умных машин.

Фон Нейман заявляет, что выход нейронов носит цифровой характер: аксон либо порождает нервный импульс, либо нет, т. е. он игнорировал непрерывный химизм. Фон Нейман описывает вычисления, которые производит нейрон, как взвешенную сумму входных сигналов с неким порогом. Данная модель работы нервной клетки легла в основу такого направления, как коннекционизм — построение искусственных систем (как с точки зрения технического, так и программного обеспечения) по принципу организации и функционирования живого нейрона.

Знаковым рубежом в формировании концепции искусственного интеллекта стал труд Алана Тьюринга, сформулировавшего метод объективной проверки разумности машины — прославленный «Тест Тьюринга» (1950 г.).

Тест Тьюринга давно стал объектом интенсивной дискуссии. Ряд авторов, ставят под сомнение его онтологическую базу, отмечая, что процедура фиксируется лишь на наблюдаемом поведенческом интеллекте, оставляя без внимания более сложные аксиологические, этические и метафизические аспекты субъективности. Кроме того, в ряде исследований, посвящённых критическому анализу тестовых методик, выделяются именно философские основания Теста Тьюринга [17].

Ряд теоретиков оспаривает идею, что когнитивность способна проявляться исключительно через поведенческие реакции в диалоге. Следует признать, что Тест Тьюринга не является универсальным критерием: его валидность обусловлена историка культурным контекстом и уровнем доступных технологий.

Бурное развитие современных технологий искусственного интеллекта ставит под сомнение неизменную значимость классического Теста Тьюринга.

По мере роста вычислительной мощности и усложнения алгоритмов появляются альтернативные методологии осмысления и измерения интеллектуальности.

Перцептрон стал одной из первых моделей нейросетей, а «Марк-1» — первым в мире нейрокомпьютером.

На 2-х месячном семинаре в Дартмуте (Великобритания), летом 1956 года определена новая наука «Искусственный интеллект».

Первая зима ИИ (1974-1980). Сдерживающие факторы – отсутствие методов и алгоритмов, адекватных сложности поставленных задач – исключение ИИ из числа национальных приоритетов, значительное сокращение финансирования нейронных сетей.

Возобновление работ по ИИ (1981–1986) – ЭВМ 5-го поколения (поддержка диалога, перевод языков, интерпретация изображений, построение причинно-следственных связей, Япония), экспертные систем, генетические алгоритмы. Эта программа представляла собой попытку совершить революционный

прорыв в области информационных технологий и искусственного интеллекта, переосмыслив саму концепцию вычислительных машин.

Предыдущие поколения ЭВМ прошли путь от электронных ламп к транзисторам, интегральным схемам и микропроцессорам. Каждое поколение знаменовалось не только сменой элементной базы, но и появлением новых архитектурных решений, языков программирования и областей применения вычислительной техники. К началу 1980-х годов ЭВМ четвертого поколения, построенные на базе микропроцессоров и больших интегральных схем, стали основой информационной революции, вывели компьютеры из специализированных центров в офисы и дома обычных пользователей.

Однако научное сообщество и индустрия осознавали ограничения существующих технологий, особенно в контексте новых вызовов – создания систем искусственного интеллекта, обработки естественного языка, машинного зрения и экспертных систем. Именно эти задачи должны были решить компьютеры пятого поколения, знаменующие переход от обработки данных к обработке знаний.

Япония, стремившаяся укрепить свои позиции в мировой экономике и технологическом развитии, в 1982 году объявила о запуске амбициозного проекта FGCS (Fifth Generation Computer Systems).

Этот проект, рассчитанный на десятилетие, предполагал инвестиции в размере около 850 миллионов долларов и должен был обеспечить Японии лидерство в разработке интеллектуальных компьютерных систем. Несколько важных факторов обусловили необходимость разработки принципиально новых подходов к созданию вычислительных систем в начале 1980-х годов:

Стратегические экономические интересы. К началу 1980-х годов Япония уже добилась значительных успехов в производстве электроники и автомобилей, но стремилась укрепить свое положение в сфере информационных технологий, где доминировали американские компании, такие как IBM и DEC. Проект ЭВМ пятого поколения рассматривался как стратегический шаг для обеспечения технологического лидерства Японии.

Ограничения традиционной архитектуры фон Неймана. Компьютеры, основанные на архитектуре фон Неймана (последовательное выполнение инструкций с использованием общей памяти для данных и программ), сталкивались с фундаментальными ограничениями производительности при решении задач, требующих параллельной обработки информации, особенно в области искусственного интеллекта.

Развитие исследований в области искусственного интеллекта. 1970-е годы стали периодом активного развития теоретических основ и практических приложений искусственного интеллекта. Экспертные системы, обработка естественного языка, компьютерное зрение – эти направления требовали новых подходов к организации вычислительного процесса.

Появление логического программирования. Разработка языка Prolog в начале 1970-х годов предложила новую парадигму программирования, основанную на логическом выводе и декларативном описании знаний, что открывало перспективы для создания систем, способных к рассуждениям на основе заданных правил и фактов.

Рост объемов информации. Информационный взрыв, наблюдавшийся в различных областях – от науки до бизнеса – требовал новых средств хранения, обработки и анализа данных, особенно неструктурированной информации.

Социальные вызовы. Япония, как и многие развитые страны, столкнулась с демографическими проблемами – старением населения, потенциальной нехваткой квалифицированной рабочей силы. Интеллектуальные компьютерные системы рассматривались как средство повышения производительности труда и решения социальных проблем.

Эти предпосылки создали благоприятную почву для формирования амбициозной программы, нацеленной на революционное изменение принципов построения и использования вычислительной техники. Пятое поколение ЭВМ должно было не просто повысить производительность или миниатюризацию, как

это происходило при смене предыдущих поколений, но перейти к принципиально новой парадигме – от обработки данных к обработке знаний. Основные цели проекта FGCS можно разделить на несколько ключевых направлений:

Технологические цели

- Создание компьютерных систем с производительностью до 1000 логических выводов в секунду (LIPS), что на порядки превышало возможности существующих систем.
- Разработка параллельных архитектур, способных эффективно выполнять логический вывод.
- Создание новых языков программирования, ориентированных на обработку знаний.
- Разработка специализированных аппаратных средств для поддержки логического программирования.
- Создание инструментов для автоматического перевода с естественных языков.

Функциональные возможности

- Понимание естественного языка и способность к ведению диалога на нем.
- Автоматический перевод между несколькими языками.
- Распознавание речи и образов.
- Способность к самообучению и адаптации.
- Поддержка принятия решений на основе анализа больших массивов данных.
- Возможность рассуждать на основе неполной и нечеткой информации.

Коммерческие и социальные амбиции

- Обеспечение технологического лидерства Японии в области информационных технологий.
- Создание основы для новых продуктов и услуг, повышающих конкурентоспособность японской экономики.

- Решение социальных проблем, связанных с демографическими изменениями и необходимостью повышения производительности труда.
- Формирование новой информационной инфраструктуры для поддержки общества, основанного на знаниях.

Эти технологии представляли собой значительный прогресс в области логического программирования, параллельных вычислений и обработки естественного языка. Многие из них были реализованы в виде работающих прототипов и демонстрационных систем, подтверждая техническую осуществимость концепции ЭВМ пятого поколения. Однако переход от экспериментальных разработок к коммерчески жизнеспособным продуктам оказался более сложной задачей, чем предполагалось изначально.

Несмотря на то, что проект FGCS не достиг всех заявленных целей, он привел к ряду значимых достижений, которые оказали влияние на развитие информационных технологий:

Развитие параллельных архитектур для логического программирования. К концу проекта были созданы системы PIM/m с 64 процессорами, способные выполнять до 120 миллионов логических выводов в секунду, что значительно превосходило возможности традиционных компьютеров того времени в области символьной обработки.

Разработка специализированных языков программирования и сред разработки, оптимизированных для логического программирования и обработки знаний. Языки семейства KL представляли собой значительный прогресс по сравнению с существовавшими реализациями Prolog.

Создание прототипов систем машинного перевода между японским и английским языками, демонстрирующих возможности автоматического анализа синтаксической структуры текста и перевода с учетом контекста.

Разработка методов распараллеливания логического вывода, позволяющих эффективно использовать многопроцессорные системы для решения задач искусственного интеллекта.

Научные и организационные достижения

- **Продвижение логического программирования** как парадигмы для решения сложных задач искусственного интеллекта. Проект стимулировал исследования в области теории логического программирования, методов оптимизации и параллельного выполнения логических программ.
- **Формирование сообщества исследователей**, специализирующихся на логическом программировании и параллельных вычислениях. ICOT стал центром притяжения для талантливых ученых и инженеров, многие из которых впоследствии продолжили работу в этой области в университетах и исследовательских центрах.
- **Создание обширной базы знаний и публикаций**. За время существования проекта было опубликовано более 1000 научных статей, проведено множество конференций и семинаров, что способствовало распространению знаний и идей, разработанных в рамках FGCS.
- **Развитие международного сотрудничества** в области искусственного интеллекта и новых архитектур вычислительных систем. Несмотря на первоначальную напряженность, проект FGCS стимулировал обмен идеями между японскими, американскими и европейскими исследователями.

Коммерческие результаты и применения

- **Внедрение элементов разработанных технологий** в коммерческие продукты японских компаний, в частности, в системы управления базами данных, средства разработки программного обеспечения и компиляторы.
- **Применение методов логического программирования** в специализированных экспертных системах для медицинской диагностики, финансового анализа и инженерного проектирования.
- **Использование разработанных методов обработки естественного языка** в системах машинного перевода и информационного поиска.

Культурные и социальные эффекты

- **Повышение осведомленности общества** о потенциале искусственного интеллекта и его возможных приложениях. Проект FGCS получил широкое освещение в СМИ и способствовал формированию более реалистичных ожиданий относительно возможностей ИИ.
- **Стимулирование образования** в области компьютерных наук и искусственного интеллекта. Проект привлек внимание молодых специалистов к этим областям и способствовал обновлению учебных программ в университетах.
- **Демонстрация возможности государственно-частного партнерства** в реализации масштабных технологических проектов, что стало моделью для последующих инициатив в Японии и других странах.
- Таким образом, хотя проект FGCS не привел к созданию коммерчески успешных компьютеров пятого поколения, он внес существенный вклад в развитие теории и практики логического программирования, параллельных вычислений и искусственного интеллекта. Многие идеи и технологии, разработанные в рамках проекта, нашли применение в последующих исследованиях и разработках в этих областях.

Проект FGCS официально завершился в 1992 году после десяти лет работы, как и планировалось изначально. Однако, в отличие от первоначальных планов, проект не был продолжен в форме новой фазы или новой инициативы

Однако проект столкнулся с рядом серьезных проблем, которые в итоге привели к его завершению без перехода к коммерциализации разработанных технологий в их изначально задуманной форме. Среди этих проблем: сложность эффективного распараллеливания логического вывода, ограничения логического программирования как парадигмы, быстрое развитие традиционных компьютерных архитектур, изменение парадигм в искусственном интеллекте и экономический спад в Японии начала 1990-х годов.

Тем не менее, влияние проекта FGCS выходит далеко за рамки его непосредственных результатов. Он стимулировал исследования в области параллельных вычислений, логического программирования и методов представления знаний, способствовал формированию международного сообщества исследователей в этих областях и оказал влияние на образовательные программы по компьютерным наукам. Многие идеи и технологии, разработанные в рамках проекта, в модифицированной форме нашли применение в последующих разработках и продуктах.

В современном контексте, когда искусственный интеллект переживает новый подъем благодаря успехам глубокого обучения и нейронных сетей, опыт проекта FGCS остается актуальным как напоминание о сложности создания по-настоящему интеллектуальных систем и о важности баланса между амбициозными долгосрочными целями и практическими краткосрочными результатами в развитии технологий.

Вторая зима (1987–1993) ИИ и ИНС. Сдерживающие факторы – ограниченные вычислительные возможности – ограниченные возможности методов машинного обучения и анализа данных – высокие требования к разработчикам, значительная трудоемкость создания технологий – ограниченный объем данных для обучения и настройки систем, слабая обобщающая способность; – значительное сокращение финансирования.

Возобновление работ по ИИ (1994 г. – н. в.)

Переосмысление подходов

В этот период произошел сдвиг от чисто символьной ИИ к гибридным и статистическим методам:

- **машинное обучение** стало доминирующим подходом;
- **байесовские методы** и вероятностные модели получили широкое распространение;
- **поисковые алгоритмы** были усовершенствованы.

Знаковые события и прорывы

Победа компьютера Deep Blue над Гарри Каспаровым (1997) продемонстрировала потенциал ИИ в стратегических играх.

Развитие робототехники:

- роботы AIBO и ASIMO;
- автономные транспортные средства DARPA Grand Challenge.

Обработка естественного языка:

- IBM Watson побеждает в телевикторине Jeopardy! (2011);
- развитие систем анализа и синтеза речи.

Компьютерное зрение:

- алгоритмы распознавания объектов и лиц;
- системы видео аналитики.

Эпоха глубокого обучения и нейросетей: 2010-е — настоящее время

Революция глубокого обучения

Важнейшим прорывом последнего десятилетия стало развитие методов глубокого обучения:

Сверточные нейронные сети (CNN) революционизировали компьютерное зрение.

Рекуррентные нейронные сети (RNN) и **LSTM** улучшили обработку последовательностей.

AlexNet (2012) — переломный момент в распознавании изображений.

AlphaGo (2016) победила лучших игроков в го, что считалось невозможным для ИИ.

Генеративно-состязательные сети (GAN)позволили создавать реалистичные синтетические изображения

Трансформеры и языковые модели. С 2017 года архитектура трансформеров произвела революцию в обработке естественного языка:

- BERT (Google, 2018) значительно улучшил понимание контекста.
- GPT (OpenAI) и его последующие версии предоставили возможность генерации связных текстов.
- GPT-3 и GPT-4 демонстрируют способности, приближающиеся к человеческим во многих языковых задачах.

- DALL-E, Midjourney и Stable Diffusion позволяют генерировать изображения по текстовым описаниям.

Мультимодальные и гибридные системы ИИ

- Современные исследования направлены на создание систем, работающих с разными типами входных данных:
- Объединение зрения и языка в единые модели.
- Системы, способные воспринимать и генерировать текст, изображения, звук и видео.
- Интеграция нейросимволических подходов, объединяющих глубокое обучение и символьный ИИ.

Современные тенденции и будущие направления ИИ в промышленности и обществе:

- **повсеместное внедрение ИИ-систем** в индустрию, здравоохранение, финансы, образование;
- **персонализация услуг и продуктов** с помощью ИИ;
- **автоматизация** рутинных когнитивных задач;
- **автономные системы**— от беспилотных автомобилей до умных городов.

Этические и социальные аспекты:

– проблемы прозрачности и объяснимости алгоритмов;

– смещения и дискриминация в ИИ-системах;

– вопросы конфиденциальности и использования персональных данных;

– регулирование ИИ на национальном и международном уровнях.

Исследовательские горизонты:

- **общий искусственный интеллект (AGI)**— разработка универсальных интеллектуальных систем;
- **нейроморфные вычисления**— создание аппаратного обеспечения, имитирующего структуру мозга;
- **квантовые вычисления** для ИИ-задач;
- **самообучающиеся и само адаптирующиеся системы**;

- **объяснимый ИИ**(XAI) — создание систем, способных объяснять свои решения.

Хотя мы еще далеки от создания полноценного искусственного разума, сопоставимого с человеческим, современные системы ИИ уже трансформируют общество, экономику и науку, ставя перед нами новые технологические, этические и философские вопросы, требующие коллективного осмысления и решения.

2.2. Естественный и искусственный нейрон

Работа нейрона человеческого мозга описывается специалистами следующим образом [33–34]:

1. Биологи и нейрофизиологи определяют нейрон как возбудимую клетку, способную воспринимать, обрабатывать и передавать информацию с помощью электрических и химических сигналов.
2. Непрерывные физико-химические процессы в нейроне включают:
   - поддержание мембранного потенциала покоя (около -70 мВ);
   - ионный обмен через мембрану (особенно ионы Na^+, K^+, Cl^-, Ca^{2+});
   - метаболические процессы для обеспечения энергией (АТФ — это одна из самых важных молекул в энергетики мозга).
   3. Генерация электрического импульса (потенциала действия) происходит при:
   - достижении порогового значения деполяризации мембраны;
   - открытии потенциал-зависимых ионных каналов;
   - быстром входе ионов Na^+ внутрь клетки с последующим выходом K^+.
4. Этот процесс известен как «всё или ничего» – нейрон либо генерирует потенциал действия полностью, либо не генерирует вовсе.
5. Передача сигнала между нейронами происходит через синапсы с помощью нейромедиаторов.

Эти явления составляют основу нейрональной активности, которая лежит в основе всех функций мозга.

Нейромедиатор является входом другого или других нейронов.

Нейромедиатор — это химическое вещество, которое служит посредником при передаче сигнала между нейронами. Сам по себе он не является «входом» нейрона, но формирует входной сигнал.

1. **Процесс происходит так:**
   - нейромедиатор выбрасывается из пресинаптического нейрона (отправителя);
   - проходит через синаптическую щель;
   - связывается с рецепторами на постсинаптическом нейроне (получателе);
   - это связывание изменяет мембранный потенциал постсинаптического нейрона.
2. **Природа сигнала:**
   - нейромедиатор — это именно химическое вещество (ацетилхолин, дофамин, серотонин и др.);
   - электропотенциал (потенциал действия) — это электрический сигнал, который распространяется вдоль аксона нейрона;
   - они представляют разные фазы передачи сигнала:
   - внутри нейрона сигнал электрический (потенциал действия);
   - между нейронами – химический (через нейромедиаторы).
3. **В синапсе происходит преобразование:**
   - Электрический сигнал → высвобождение химического нейромедиатора → новый электрический сигнал.

Таким образом, нейромедиатор — это химический компонент передачи сигнала, который вызывает электрические изменения в принимающем нейроне, но сам по себе не является электрическим сигналом.

Укрупненное изображение нейрона головного мозга показана на рис. 3.

Рисунок 3. Нейрон головного мозга

Для построения искусственных нейронных сетей (ИНС) ИИ, в качестве базового элемента, используется искусственный нейрон — это математическая модель, имитирующая некоторые аспекты функционирования биологического нейрона. На рис. 4 приведена общая схема искусственного нейрона.

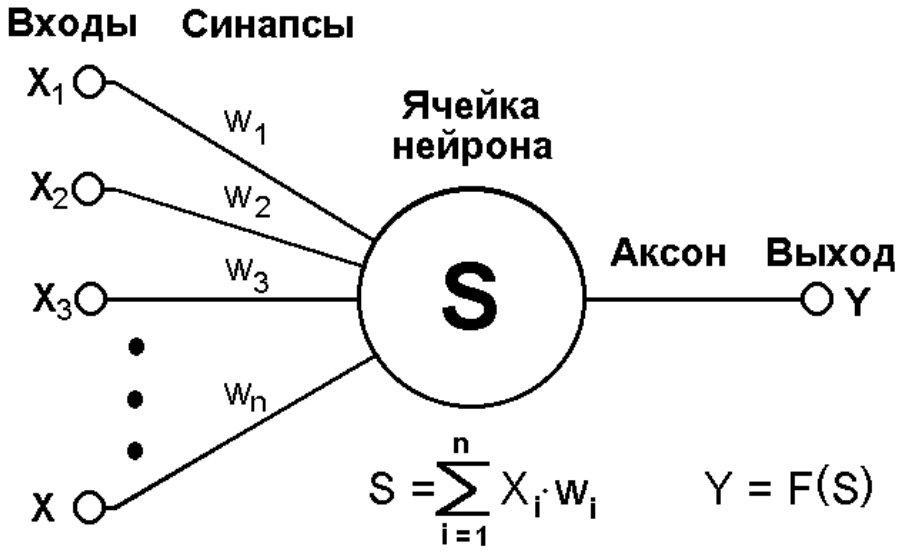


Рис. 4. Общая схема искусственного нейрона в ИИ

Ключевые особенности искусственного нейрона**:**

1. **Структура и компоненты**:

- Входные сигналы (x_1, x_2, …, x_n).
- Весовые коэффициенты (w_1, w_2, …, w_n).

- Функция суммирования (Σ).
- Функция активации (F).
- Выходной сигнал (Y).

2. **Принцип работы**:

- Получает входные сигналы от других нейронов или из внешней среды.
- Умножает каждый вход на соответствующий весовой коэффициент.
- Суммирует взвешенные входы (часто добавляя смещение/bias).
- Применяет функцию активации к полученной сумме.
- Передает результат на выход.

3. **Функции активации**:

- Сигмоидальная (логистическая)
- Гиперболический тангенс (tanh)
- ReLU (Rectified Linear Unit)
- Leaky ReLU
- Softmax (для выходного слоя в задачах классификации)

Отличия искусственного нейрона от естественного нейрона, следующие:

- Значительно упрощенная модель.
- Обычно оперирует с числовыми значениями, а не со спайками.
- Отсутствие многих биохимических процессов.
- Функционирует дискретно, а не непрерывно.
- Не обладает пластичностью и адаптивностью биологического нейрона.

Искусственные нейроны составляют основу различных архитектур нейронных сетей:

- Многослойные персептроны
- Сверточные нейронные сети (CNN)
- Рекуррентные нейронные сети (RNN)
- Трансформеры и другие современные архитектуры.
- Сети Колмогорова-Арнольда (KAN).

Искусственные нейроны не способны полноценно моделировать весь химизм и биологическую сложность реальных нейронов, они стремятся эмулировать основные принципы их функционирования.

Естественный нейрон головного мозга человека – аналого-цифровой [31].

Искусственный нейрон только цифровой.

Следовательно, искусственный нейрон весьма упрощенной является моделью нейрона головного мозга человека.

Моделирование 1 сек. активности 1% мозга на суперкомпьютере Sunway Taihulight (КНР) занимает около 4 минут машинного времени [12].

Нюансы: суперкомпьютер содержит 10,5 млн процессорных ядер, занимает ≈1000 м^2 площади и потребляет ≈ 16 МВт. В табл. 11. Показано сравнение работы естественных нейронов и искусственных.

Таблица 11. Сравнение работы естественных нейронов и искусственных

Наименование показателя	ЭВМ с производительностью 10^{20}Flops	Человеческий мозг
Занимаемый объем	4 x 10^6м^3	0,0015 м^3
Энергопотребление	15 ГВт	20 Ватт

Правильность моделей естественного интеллекта под вопросом.

Математическая модель искусственного нейрона является основой при построение нейронных сетей.

Однако искусственные нейроны используются и в других контекстах.

- **Одиночные нейроны как классификаторы**:
  - перцептрон Розенблатта как простейший линейный классификатор;
  - логистическая регрессия (по сути, одиночный нейрон с сигмовидной функцией активации).
- **Ансамбли нейронов вне традиционных нейросетей**:
  - в гибридных системах машинного обучения;
  - как часть эволюционных алгоритмов.

- **Нейроны в нейроморфных вычислениях**:
  - специализированные аппаратные реализации нейронов для энергоэффективных вычислений;
  - SpiNNaker, TrueNorth, Loihi и другие нейроморфные чипы.
- **В моделях биологических систем**:
  - для моделирования реальных нейронных процессов;
  - в нейрофизиологических исследованиях.
- **Концептуальные модели принятия решений**:
  - для моделирования пороговых решений в экономике и социальных науках;
  - в теории принятия решений.

Искусственный нейрон, таким образом, является более универсальной концепцией, чем просто строительный блок нейронных сетей, хотя это его наиболее распространенное применение в современном ИИ.

Авторы ограничатся рассмотрением применения искусственного нейрона в искусственных нейронных сетях (ИНС).

2.3. Искусственные нейронные сети

Первая концептуальная модель ИНС была предложена в 1943 году Уорреном Маккалоком и Уолтером Питтсом. Однако путь ИНС к доминированию в сфере ИИ был долгим и непростым.

Ключевые этапы развития нейронных сетей, следующие:

1943 – Маккаллок и Питтс предлагают первую математическую модель нейрона.

1958 – Фрэнк Розенблатт разрабатывает перцептрон.

1969 – «Зима нейронных сетей» начинается после публикации Минского и Пейперта о ограничениях перцептронов.

1980-е – Возрождение интереса, алгоритм обратного распространения ошибки.

1987–1993 – «Вторая зима ИИ и ИНС». Сдерживающие факторы – ограниченные вычислительные возможности – ограниченные возможности методов машинного обучения и анализа.

1990-е – Появление сверточных нейронных сетей и рекуррентных архитектур.

2006–2012 – Глубокое обучение, пред обучение нейронных сетей слой за слоем.

Нейронные сети стали доминирующим подходом в ИИ примерно с **2012 года**. Ключевой момент – победа сверточной нейронной сети AlexNet в соревновании ImageNet по классификации изображений, когда она значительно превзошла все традиционные методы.

Факторы, обеспечившие доминирование:

- появление эффективных алгоритмов обучения глубоких сетей.
- рост вычислительных мощностей (особенно GPU).
- доступность больших объемов данных для обучения.
- прорывные архитектуры (CNN, RNN, LSTM, Transformer).

Последующие годы принесли такие значимые модели как GPT, BERT, DALL-E, Stable Diffusion и другие, окончательно утвердившие доминирование нейронных сетей в современности.

2.4. Нейронные сети и технологии искусственного интеллекта

Сегодня искусственные нейронные сети [36] задают тон при проектировании большинства архитектур искусственного интеллекта; вследствие этого ИИ вместе с ИНС прочно закрепились как ключевой инструмент фундаментальных исследований и инженерных разработок в широчайшем спектре дисциплин. Указанные технологии трансформировали методологию решения высокоуровневых задач, одновременно переустраивая экономические модели, социальные взаимоотношения и быт современного человека.

Стремительное прогрессирование технологий делает исследования искусственных нейронных сетей в сфере ИИ крайне востребованными. За прошедший год наблюдалась целая серия значимых прорывов: усовершенствованы процедуры градиентного обучения, расширены вычислительные ресурсы благодаря специализированным TPU и GPU, предложены гибридные и трансформерные архитектуры последнего поколения. Эти инновации обеспечили заметный рост точности и устойчивости моделей, открыв перспективы для внедрения интеллектуальных систем в медицине, финансовой аналитике, интеллектуальной мобильности и других индустриях.

В период 2023–2024 гг. сфера искусственных нейронных сетей демонстрирует впечатляющий прогресс. Показателен стремительный прирост отечественных научных статей по нейротехнологиям и ИИ, отслеживаемый с 1998 г. Библиометрические исследования регистрируют усиление публикационной активности авторов, картируют ведущие тематические кластеры и выявляют кооперационные сети [37]. Полученные данные легли в основу разработанных стратегических рекомендаций и определения долгосрочных векторов эволюции нейротехнологической отрасли.

Интеграция BYC-алгоритмов в клиническую ортодонтию стала объектом активного изучения. В период с 2013 по 2024 гг. опубликован значительный массив работ, анализирующих, как архитектуры нейронных сетей улучшают этапы диагностики, планирования и мониторинга коррекции прикуса. Авторы подчёркивают возрастающее значение ИИ-инструментов для повышения точности и ускорения терапевтического цикла [38].

Не менее перспективной областью сегодня выступают интеллектуальные нейронные сети, которые уже рассматриваются как ключевой элемент зарождающегося цифрового ландшафта. Их функциональный потенциал заметно возрос, что открывает ещё более широкие горизонты для практического внедрения подобных технологий в самых разных отраслях — от клинической диагностики до современных образовательных платформ [39].

Среди последних инноваций особо выделяется интеграция нейронных сетей в сферу NLP, демонстрирующая их высокую универсальность, способность к доменной адаптации и соответствие требованиям актуальных вычислительных задач [40].

В период 2023–2024 годов наблюдаются существенные прорывы в технологиях, ориентированных на создание нейросетей следующего поколения. Такие архитектуры умеют динамически подстраиваться под нетипичные сценарии работы и эффективно обрабатывать массивные наборы данных. Благодаря этому расширяется спектр задач, подлежащих автоматизации, повышается обоснованность принимаемых управленческих решений, а исследования подчёркивают необходимость их внедрения в крупномасштабные киберфизические системы [41].

Таким образом, текущие позиции нейронных сетей в 2023–2024 годах отражают стремительное эволюционирование методов глубокого обучения, трансформеров и генеративных моделей, которые кардинально трансформируют стратегии решения прикладных и исследовательских задач в мультидисциплинарных областях. Следует подчеркнуть, что грядущие технологические вехи будут ориентированы на повышение качества человеко-машинного взаимодействия, расширение когнитивных интерфейсов и оптимизацию потоков данных. Наблюдаемые тренды указывают, что нейронные архитектуры превратятся в фундамент для ещё более комплексных систем сильного ИИ, формируя дальнейшую траекторию отрасли.

Таблица 12 отражает динамику развития искусственного интеллекта и машинного обучения в 2023 году.

Таблица 12. Ключевые тенденции эволюции технологий искусственного интеллекта 2023 года

Параметр	**Рост производительности**	**Количество рабочих мест**	**Этические аспекты**
Прогноз	30-50%	100 миллионов к 2025 году	Острая потребность в стандартизации

Ключевые тренды развития искусственного интеллекта, ожидаемые к 2025 году, смещают фокус на ряд взаимосвязанных направлений, определяющих дальнейшую эволюцию отрасли. На передний план выходит экспоненциальный рост генеративного ИИ: нейросети, опирающиеся на трансформерные архитектуры и диффузионные модели, уже создают авторские тексты, фотореалистичную графику и сложные музыкальные треки, используя масштабные датасеты. Такие модели становятся ядром пользовательских интерфейсов и промышленного ПО, революционизируя UX, ускоряя R&D-циклы и повышая производственную эффективность [42].

С экономической позиции фиксируется стремительный рост эффективности труда вследствие интеграции систем искусственного интеллекта. Эксперты прогнозируют, что производительность в области документооборота способна увеличиться на 30–50% за счёт интеллектуальной автоматизации рутин. Особенно ярко трансформация прослеживается в медицине: алгоритмы машинного обучения помогают врачам быстрее и точнее формулировать диагнозы, тогда как роботизация бизнес-процессов в смежных секторах снижает операционные издержки и повышает потребительскую ценность сервисов [43].

К важнейшим социальным трендам относят трансформацию рынка занятости, вызванную цифровизацией и ускоренной автоматизацией. Эксперты прогнозируют, что уже к 2025 году количество специалистов, вовлечённых в экосистему искусственного интеллекта, приблизится к отметке 100 миллионов человек [44].

В то же время усиливаются опасения относительно вероятного исчезновения рабочих мест вследствие широкомасштабной автоматизации и роботизации. Эксперты активно обсуждают стратегии переквалификации кадров и набор компетенций, востребованных в постиндустриальной экономике будущего [45].

Этические вопросы приобретают ключевое значение на этапе интеграции систем ИИ. Возрастающая ответственность за автономные решения и требование алгоритмической справедливости диктуют необходимость разработки актуальных нормативов, протоколов прозрачности и процедур аудита. Формализация

этических принципов позволит минимизировать предвзятость, предотвратить дискриминацию и надёжно защитить базовые права человека при эксплуатации ИИ-технологий [46].

Следовательно, современные тренды, синергетически объединяющие прорывы в машинном обучении, нейронных сетях и социокультурные детерминанты, формируют принципиально новую парадигму применения искусственного интеллекта. Эти факторы не только задают вектор эволюции ИИ-систем, но и подчеркивают необходимость своевременной адаптации бизнеса и общества, сохраняя паритет между технологическим рывком, этическими нормами и устойчивым развитием. Осознанное внимание к данным тенденциям определит качество внедрения ИИ и его долговременное воздействие на человека и социальную структуру.

Искусственный интеллект уже существенно трансформировал общественную структуру и повседневные практики, вызывая как позитивные, так и негативные социальные эффекты. Следует осознавать, что последствия цифровизации затрагивают не лишь технологию, но проникают в социокультурные, политико-экономические и даже этические пласты. Так, применение ИИ в образовательной сфере открывает персонализированное обучение, расширяет доступ к знаниям и аналитике, однако одновременно усиливает технократизацию процессов, формирует алгоритмическое давление на педагогов и учащихся и риск усиления цифрового неравенства [47].

Помимо прочего, стратегическое внедрение искусственного интеллекта сопряжено с долгосрочными сдвигами внутри общественной структуры, способными существенно повлиять на цифровое неравенство и разницу в доступе к передовым технологиям. Эксперты подчёркивают, что алгоритмизация социальных процессов рискует усилить уже существующие диспропорции и маргинализацию уязвимых групп, поэтому требуется оперативный мониторинг и превентивный анализ возможных рисков [48].

К числу наиболее проблемных последствий интеграции искусственного интеллекта относится ситуация, при которой автоматизация ведёт к ликвидации рабочих позиций; это трансформирует структуру рынка труда, порождает новые сегменты и переосмысливает роль человека, предъявляя иные требования к его компетенциям [49].

Настоятельная потребность в философском переосмыслении назревших вызовов особенно ощутима, поскольку человечество стоит на рубеже очередных социокультурных трансформаций. С ростом масштабов автоматизации и систем ИИ следует акцентировать внимание на проблематике цифровой этики, оценивая не только способы внедрения технологий, но и их влияние на человека, что диктует необходимость междисциплинарных механизмов управления и регуляторных стратегий [50].

Следует подчеркнуть, что внедрение ИИ в разнообразные сферы не только стимулирует формирование принципиально новых отраслей, но и порождает риски – от компрометации данных до нарушения персональной безопасности, что требует комплексной кибербезопасной политики [51].

Сочетание позитивных и негативных эффектов, сопровождающих внедрение искусственного интеллекта, заставляет задуматься о грядущих трансформациях. Остаётся неясным, как социум будет перестраиваться под новые цифровые реалии. Стратегии интеграции ИИ в обыденные процессы обязаны принимать во внимание и его инновационный потенциал, и возникающие риски, чтобы формировать более этичную, инклюзивную и безопасную технологическую эпоху.

Нейронные сети, рассматриваемые как метод интеллектуального анализа данных, служат высокотехнологичным инструментом для обработки, структурирования и извлечения ценных сведений из масштабных и гетерогенных датасетов. Сегодня существует целый спектр подобных алгоритмов, варьирующихся по топологии, функциям активации, механизму обучения и типу оптимизаторов. Показательной тенденцией последних лет стало применение глубоких архитек-

тур для полной или частичной автоматизации аналитического цикла, что радикально сокращает сроки разработки решений, уменьшает операционные расходы и повышает качество прогнозов.

Классические подходы к обработке данных опираются на заранее сформулированные правила и исследовательские гипотезы. Для их применения необходим тщательный предварительный анализ и строгая постановка критериев, что становится узким горлом при анализе неструктурированных массивов. В противоположность этому, искусственные нейронные сети уверенно решают неформализованные задачи, включая кластеризацию, регрессионное прогнозирование и выявление аномалий. Их способность автоматически обнаруживать скрытые паттерны без жёсткой ручной подготовки расширяет арсенал современных аналитических систем [52].

В настоящее время передовые подходы, в частности сверточные нейронные сети (CNN), фактически превратились в отраслевой стандарт при обработке визуальных данных. Такие архитектуры демонстрируют выдающуюся результативность в задачах классификации и семантической сегментации изображений, а также в локализации и детекции объектов, используя как режимы обучения с учителем, так и self-supervised или полностью безучительские методики. Благодаря устойчивости к шуму, артефактам и искажениям данные модели незаменимы в медицинской визуализации, телеметрическом контроле и других системах автоматизированного мониторинга [53].

Еще одним ключевым элементом анализа данных на базе нейронных сетей выступает интеллектуальный анализ данных (Data Mining), то есть процесс Knowledge Discovery. Данный методологический блок объединяет техники выявления латентных паттернов в массиве информации, что способствует получению релевантных инсайтов и выработке обоснованной бизнес-стратегии. Практика показывает, что нейросетевые модели эффективно решают задачи оптимизации бизнес-процессов и усиления конкурентных преимуществ компаний, демонстрируя свою гибкость и масштабируемость [54].

Эффективность нейронных сетей при анализе данных измеряют набором метрик — точностью (accuracy), полнотой (recall) и интегральной F-мерой. Эти показатели определяют качество модели и её умение обобщать на не виденных примерах. Одновременно внедрение нейросетевых решений требует значительных вычислительных ресурсов, что усложняет их практическое применение [55].

Нарастающий интерес к совершенствованию аналитических методов и алгоритмов активно стимулирует эволюцию нейронных сетей и их широкое внедрение в разнопрофильные индустрии.

Аналитические прогнозы о будущих трендах ИИ в ближайшие десятилетия подчёркивают приоритет разработки высокомощного, максимально универсального, приближенного к AGI интеллекта.

Учитывая современные прорывы в сфере глубоких нейронных сетей, включающие высокопроизводительную обработку массивов данных и обеспечение функционирования автономных платформ, следует подчеркнуть, что дальнейший рывок будет обусловлен всесторонней интеграцией искусственного интеллекта в разнообразные сферы человеческой деятельности и труда. В медицине ИИ уже демонстрирует впечатляющие результаты, содействуя ранней диагностике патологий, что способно снизить показатели смертности и повысить уровень здоровья и благополучия пациентов [56].

Тем временем в транспортной отрасли быстрыми темпами совершенствуются технологии автономного пилотирования, их внедрение ориентировано на существенное снижение аварийности и рост комплексной безопасности дорожной среды. Эксперты прогнозируют, что к 2030 году роботизированные автомобили станут повсеместным явлением, тогда как классические формы передвижения уступят им ведущие позиции [16]. Одновременно такая трансформация ставит перед обществом новые задачи — от гарантирования устойчивой кибербезопасности киберфизических систем до решения многоуровневых этико-правовых дилемм [57].

Неотъемлемым элементом дальнейшей эволюции искусственного интеллекта является преодоление технологической концентрации. Когда государственные структуры и крупные IT-конгломераты аккумулируют чрезмерную ры-

ночную и экспертную власть, требуется внедрять антимонопольные, регуляторные и open-source-ориентированные механизмы, повышающие прозрачность алгоритмов, защищающие права пользователей и сдерживающие неконтролируемое развитие.

Современные исследования указывают, что приоритетным вектором станет не только разработка передовых технологических решений, но и формирование адекватных этико-правовых регуляторов, ориентированных на общественные запросы [58].

Одновременно ведутся углублённые изыскания как в области узкоспециализированного, так и в сфере общего, или сильного искусственного интеллекта. Это помогает шире интерпретировать потенциальное воздействие ИИ на цивилизацию. Человечество приблизилось к порогу появления универсального AGI, и последствия его внедрения для социально-экономической структуры всё ещё трудно прогнозировать. Сформировались ожидания, что синергия алгоритмов машинного обучения с прогрессивными экономическими парадигмами поспособствует формированию умных городов, устойчивых к потрясениям экономик и инклюзивных моделей образования [59].

В грядущую эпоху экосистема ИИ будет формироваться не только прогрессом алгоритмов, но и степенью социально-культурной адаптивности. Система образования, рынок труда и даже гуманитарные дисциплины подвергнутся трансформации под воздействием инновационных решений. Взвешенное нормативно-правовое сопровождение и глубокое понимание таких динамических процессов смогут обеспечить, чтобы искусственный интеллект работал на благо человечества, создавая более безопасную, устойчивую и комфортабельную среду обитания [58, 60].

Искусственный интеллект демонстрирует широчайший спектр практического использования в разных секторах экономики, являясь высокоэффективным инструментом для повышения производительности и комплексной оптимизации бизнес-процессов. В медицине, к примеру, внедрение когнитивных технологий

ИИ радикально совершенствует клиническую диагностику, поскольку нейросетевые алгоритмы анализируют медицинские изображения, электронные карты и биометрические параметры. Модели глубокого обучения распознают патологические изменения на доклинических стадиях, что существенно повышает вероятность успешного терапевтического вмешательства [61].

Тем не менее, внедрение искусственного интеллекта может натолкнуться на вызовы, касающиеся защиты личных данных и соблюдения этико-правовых стандартов.

В банковско-финансовой индустрии искусственный интеллект все чаще задействуется для автоматизации повседневных процессов, включая обработку клиентских заявок, мониторинг транзакций и детектирование мошенничества. Модели машинного обучения позволяют оперативно анализировать массивы гетерогенных данных, обеспечивая более точное управление рисками и рост уровня клиентского сервиса. Однако широкомасштабное внедрение таких решений способно спровоцировать сокращение персонала, что превращается в заметный вызов для специалистов отрасли [62].

В сфере образования искусственный интеллект применяют для разработки адаптивных курсов, которые динамически подстраиваются под индивидуальный темп и стиль обучения каждого студента. Алгоритмы машинного обучения анализируют данные об успеваемости и ошибках, формируют персонализированные задания и тем самым повышают эффективность обучения, расширяя доступ к качественным ресурсам для разных социальных групп. Одновременно такая цифровизация поднимает вопросы сохранения человеческого элемента и обеспечения равного доступа к новым технологиям [63].

В производственной отрасли ИИ всё активнее интегрируется в технологические цепочки для роста производительности и минимизации издержек. Интеллектуальные роботы, цифровые двойники и автономные конвейеры берут на себя трудоёмкие и высокоточные операции, ускоряя выпуск продукции. Однако по-

добная цифровая трансформация требует крупных капиталовложений и пересмотра устоявшихся производственных практик, что нередко встречает сопротивление персонала [64].

В современной автопромышленности искусственный интеллект активно задействуется для разработки автономных транспортных систем, кардинально меняя парадигму мобильности и управления дорожным потоком. Несмотря на впечатляющие достижения алгоритмов машинного обучения и сенсорных платформ, вопросы кибербезопасности, функциональной безопасности и регуляторная неопределённость всё ещё серьёзно сдерживают широкомасштабное развертывание таких решений [65].

Подводя итоги, роль искусственного интеллекта в разнообразных отраслях подтверждается его потенциалом оптимизировать бизнес-процессы, повышать производительность и сокращать операционные издержки. Однако проблемы, сопровождающие его интеграцию—от кибербезопасности до нормативного соответствия—обязывают применять комплексный подход, обеспечивающий безопасное, прозрачное и этичное использование алгоритмов машинного обучения, что станет фундаментом их устойчивого развития.

В предстоящие годы нейронные сети, по-видимому, пройдут фазу интенсивной эволюции: пересмотр модульной архитектуры, механизмов внимания и способов регуляризации станет критически важным для повышения эффективности. Приоритетная задача — сокращение вычислительной нагрузки и рост пропускной способности. Применение энергосберегающих стратегий обучения, включая самонастраивающиеся, мета-адаптивные и квантованные алгоритмы, способно радикально уменьшить потребление электроэнергии во время инференса и обучения. Одновременно усиливается интерес к внедрению таких моделей в киберфизические и распределённые системы, где требуются гибкость и самообучаемость, существенно расширяющие их прикладную ценность [65].

Фундаментальные и прикладные задачи, связанные с усовершенствованием и более широким внедрением нейронных сетей, диктуют необходимость

создания прогрессивных технологических решений, способных органично взаимодействовать с смежными областями науки. В этом аспекте альтернативные парадигмы обработки данных, например квантовые вычисления, потенциально позволяют радикально повысить производительность и энергоэффективность нейросетевых алгоритмов [66].

Внедрение синергетических подходов в робототехнику и цифровую автоматизацию производств существенно расширит горизонты применения нейронных сетей и повысит эффективность операций.

Исследования генеративно-состязательных сетей, а также внедрение нейромоделей в области обработки естественного языка и компьютерного зрения стабильно расширяют своё влияние [67].

Указанные вызовы диктуют необходимость разработки более прогрессивных платформ, способных обрабатывать частично структурированные и совершенно неструктурированные наборы данных. В обозримой перспективе подобные решения станут де-факто стандартом для высоконагруженных fintech-систем, медицинской информатики и иерархической сервисной робототехники.

Тем временем, по мере расширения сфер использования нейронных сетей возникают и свежие вызовы. На первый план выходят проблемы интерпретируемости и верифицируемой объяснимости этих моделей. Для их полноценного решения требуется скоординированная работа экспертов из нескольких доменов — от философии сознания и когнитивной психологии до юриспруденции и цифровой этики. Такое междисциплинарное сотрудничество способно открыть принципиально новые исследовательские горизонты и задать вектор дальнейшего развития области [68].

Разработка инструментов, обеспечивающих пользователю возможность глубже разбираться в работе и управлять внутренними процессами нейронных сетей, неизменно превратится в ключевое направление дальнейшей эволюции этой технологии.

Таким образом, перед исследовательским и инженерным сообществами формируется обширный перечень задач: модернизация архитектур, повышение энергоэффективности, тесная интеграция с гетерогенными и квантовыми вычислительными платформами, а также детальная проработка этических и правовых аспектов. Успешное достижение этих целей не только ускорит эволюцию нейронных сетей, но и глубоко преобразит повседневную жизнь. Очевидно, что будущее технологии будет определяться не столько мощностью, сколько надежностью, безопасностью и прозрачностью для конечного пользователя.

Экспертные точки зрения в области искусственного интеллекта образуют фундамент дискуссий, задавая вектор понимания актуальных трендов, рисков и возможностей отрасли. Инновационные процессы, прежде всего связанные с машинным обучением и нейронными сетями, немыслимы без активного участия профильных специалистов, чьи компетенции направляют стратегию научных изысканий и R&D. Лидеры мнений подчёркивают, что именно экспертное сообщество определяет, какие алгоритмы окажутся критически важными и как они будут интегрированы в различные домены.

Специалисты современного технологического дискурса подчёркивают необходимость адаптивной стратегии при внедрении цифровых решений в повседневную практику. Особое внимание уделяется этическим дилеммам, сопровождающим применение систем искусственного интеллекта и машинного обучения. Создание инновационных алгоритмов предполагает не только высокие инженерные компетенции, но и ясное понимание социальных последствий. Возникающие социо-экономические риски обязывают многомиллионную аудиторию пользователей выработать готовность к трансформациям и включиться в совместное конструирование грядущих изменений [69].

Ряд специалистов подчеркивает, что массовое распространение искусственного интеллекта необходимо сопровождать целенаправленными просветительскими и обучающими программами. Представители различных дисциплин — от клинической медицины до фундаментальных наук — всё настойчивее

требуют активного участия общества и индустрии в механизмах смягчения потенциальных угроз, возникающих при внедрении алгоритмов машинного обучения [70].

Крайне важно, чтобы потенциал инновационных концепций реализовывался не исключительно в корпоративном секторе, но и в университетском сообществе, где научные разработки целенаправленно подстраиваются под потребности рынка.

Значение профильных экспертов возрастает по мере ускоренного трансформирования технологического ландшафта. Дискуссии о рисках, трендах и перспективных возможностях регулярно разворачиваются среди профессионалов, действующих на передовой научно-инженерного прогресса. Поэтому R&D-центры обязаны синергировать взгляды теоретиков и практиков, достигая оптимального баланса между радикальной инновацией и прикладной результативностью [71].

Активное привлечение профильных специалистов к выработке знаний, норм и регулятивных рамок в области ИИ не менее существенно, чем сам технологический прогресс. В частности, глубокое понимание того, как машинное обучение трансформирует eHealth и RegTech-секторы, критически важно для своевременной адаптации и совершенствования бизнес-процессов [72].

Многие лидеры мнений акцентируют внимание на том, что при отсутствии вовлечения профильных специалистов в формирование этических регламентов и протоколов применения технологий социум может наткнуться на уже обозначенные в дискуссиях риски, сопровождающие стремительные технологические трансформации.

В итоге активное участие профильных экспертов на всех этапах создания и интеграции систем искусственного интеллекта является не просто желательной, а жизненно необходимой предпосылкой для гарантии их безопасной и этически выверенной эксплуатации. Профессиональные оценки формируют осмысленную стратегию внедрения, тем самым поддерживая устойчивое социальное развитие в эпоху глубоких технологических трансформаций [73].

В течение прошедшего года мы наблюдали впечатляющий скачок в развитии данной дисциплины, который не только ярко подтвердил потенциал cutting-edge технологий, но и окончательно обозначил новые векторы их практического внедрения. Фундаментальными вехами стали усовершенствование обучающих алгоритмов, экспоненциальный рост вычислительных ресурсов и эволюция архитектур искусственных нейронных сетей. Указанные прорывы легли в основу формирования более производительных и высокоточных моделей, ускорив прогресс в обработке естественного языка, компьютерном зрении, робототехнике и автономных платформах.

Современные тренды в сфере искусственного интеллекта неизменно привлекают повышенное внимание. Мы становимся свидетелями стремительного внедрения алгоритмов машинного обучения и нейронных сетей в повседневные отрасли — от здравоохранения и финансов до транспорта и системы образования. Такая цифровизация не просто повышает качество предоставляемых услуг, но и радикально перестраивает традиционные методики решения задач. К примеру, клинические ИИ-системы ускоряют диагностику и персонализируют терапию, а fintech-платформы используют предиктивную аналитику для риск-менеджмента и обработки массивов данных. Одновременно усиливается социальный резонанс: вопросы алгоритмической этики, приватности, кибербезопасности и прозрачного регулирования требуют комплексных нормативных подходов.

Методики анализа данных на базе нейронных сетей переживают стремительную эволюцию. Передовые парадигмы, включая глубокое обучение, обучение с подкреплением и гибридные архитектуры, дают возможность масштабно обрабатывать петабайты данных, выявлять скрытые закономерности и формировать практические инсайты. Тем не менее, рост параметрической сложности приводит к усилению риска непредсказуемых эффектов. Отсюда вытекает приоритет разработки прозрачных, интерпретируемых и объяснимых алгоритмов XAI, способных укрепить доверие конечных пользователей.

Согласно актуальным прогнозам, развитие искусственного интеллекта в ближайшие годы продолжит ускоряться: алгоритмы станут точнее, вычислительные архитектуры — эффективнее, а их внедрение проникнет в повседневные сценарии. Предполагается, что AI глубоко интегрируется в корпоративные цепочки создания стоимости и окажется ключевым инструментом в борьбе с глобальными вызовами — от климатической адаптации до персонализированной медицины. При этом необходимо учитывать социальные издержки автоматизации и риски злоупотребления алгоритмическими системами.

Этическая проблематика искусственного интеллекта заслуживает внимания. Следует формировать комплекс сводов этики и отраслевых стандартов, регулирующих внедрение ИИ, чтобы снижать риски, предотвращать алгоритмическую предвзятость и поддерживать принцип справедливости. Любое использование ИИ в сферах экономики и общества требует осознания ответственности, транспарентности и подотчетности за принимаемые алгоритмами решения и их последствия.

Перспективы развития нейронных сетей и систем искусственного интеллекта остаются впечатляющими, однако реализация этих потенциалов требует интегрированного подхода к проектированию, обучению и внедрению алгоритмов. Значение профильных специалистов, работающих на стыке машинного обучения, data governance и кибербезопасности, неизменно возрастает, ведь именно они способны направить эволюцию ИИ в безопасное и этически выверенное русло. В итоге ИИ открывает невиданные ранее возможности, одновременно формируя серьезные технологические и социальные вызовы, преодолеть которые можно лишь совместными усилиями.

2.5. Ключевые фундаментальные вызовы современного ИИ и их значимость

За последние десять-пятнадцать лет искусственный интеллект прочно встроился в разнообразные отрасли – от клинической диагностики и алгоритмической торговли до адаптивного e-learning и судебной аналитики. Одновременно стремительное масштабирование ИИ порождает новые вызовы, требующие системного переосмысления. Одним из центральных вопросов выступает разрыв между возросшей архитектурной сложностью нейросетевых моделей и необходимостью их прозрачности, трактуемости и соответствия принципам XAI. Дефицит интерпретируемости уменьшает доверие пользователей, регуляторов и общества, что чревато торможением внедрения интеллектуальных систем в социально и технологически критичные домены.

Проблематика, связанная с ИИ, приобретает всё большую остроту по мере его стремительного проникновения в разнообразные сегменты человеческой деятельности. Тем не менее формирование устойчивого доверия к интеллектуальным системам осложняется целым спектром вызовов, затрагивающих и интерпретируемость алгоритмов, и их эксплуатационную надёжность. Ключевым препятствием остаётся высокая сложность современных архитектур глубокого обучения, провоцирующая эффект «чёрного ящика», где внутреннюю логику вычислений практически невозможно отследить. Подобная непрозрачность порождает скептицизм как у конечных пользователей, так и у органов регулирования, особенно на фоне завышенных ожиданий от цифровых технологий [75].

Ограниченность доступных ресурсов по-прежнему выступает серьёзным барьером. Для качественного обучения современных нейросетевых архитектур требуются колоссальные объёмы вычислительной мощности, энергии и времени, которые большинство компаний, особенно малого и среднего сектора, не способны обеспечить. В результате высокая сложность и ресурсоёмкость таких систем усиливают технологическое неравенство между корпорациями-гигантами и

более скромными игроками. Дополняют картину управленческие факторы: дефицит компетентных специалистов, а также недостаток инвестиций в НИОКР затрудняют внедрение устойчивых ИИ-решений [76].

Дефицит и несовершенство данных лишь усугубляют текущие трудности. В процессе машинного обучения критически важны их достоверность, репрезентативность и полнота: недостаток высококачественных, хорошо аннотированных выборок оборачивается некорректными инсайтами, завышенной погрешностью и ошибочными прогнозами. Дополнительную угрозу несёт скрытая или явная смещённость датасетов, поскольку она склонна транслировать, а порой и усиливать устоявшиеся социальные стереотипы, что ставит перед разработчиками серьёзную этическую задачу — внедрять механизмы аудита и обеспечения алгоритмической справедливости [77].

Недостаточная зрелость формальных верификационных методов и нерешённые вопросы интерпретируемости вносят значительный вклад в формирование скепсиса по отношению к искусственному интеллекту. Хотя конечные пользователи и инженеры ожидают от интеллектуальных систем повышенной надёжности, проверяемости и предсказуемости, существующие инструменты формальной валидации пока не гарантируют приемлемого уровня explainability [78]. Возникает своеобразный парадокс: глубокие, но непрозрачные архитектуры нередко превосходят по метрикам более прозрачные модели, что, в конечном счёте, подрывает доверие к получаемым рекомендациям и прогнозам.

Следовательно, укрепление доверия к системам ИИ требует интегрированного курса действий, охватывающего устранение технических барьеров — кибербезопасность, интерпретируемость и устойчивость моделей — и одновременную корректировку управленческих стратегий и механизмов AI-governance. Глубокий анализ архитектур нейросетей с учётом их возможностей, ограничений и предвзятостей задаёт основу для создания более эффективных процедур обработки данных и надёжного развертывания масштабируемых ИИ-приложений.

Наличие проблем с качеством и объёмом обучающих выборок тесно коррелирует с уровнем конструктивной сложности алгоритмических моделей. Архитектуры, сформированные и натренированные на ограниченных или однородных датасетах, при переносе на более вариативные сценарии либо на совершенно незнакомые данные демонстрируют заметное падение точности [79]. Параметрическая насыщенность модели определяет её адаптивный потенциал: чрезмерно гибкая, высоко параметризованная сеть зачастую улавливает несущественный шум, что провоцирует эффект переобучения. Поэтому целенаправленное упрощение, например за счёт регуляризации или отбора признаков, улучшает интерпретируемость и повышает способность к обобщению в новых условиях [80].

Помимо этого, рост числа параметров неизбежно влечёт за собой повышенные требования к вычислительным мощностям и времени тренировки, что часто становится серьёзным барьером для широкого внедрения таких систем в прикладные сценарии [81]. Оптимальное расходование ресурсов приобретает особую важность при дефиците финансовых средств и ограниченной ИТ-инфраструктуры. При этом классические модели, хоть и менее гибкие, нередко демонстрируют сопоставимое качество, потребляя существенно меньше аппаратных ресурсов и энергетического бюджета на ряде задач [82].

Трудности интерпретации ИИ с позиции формальных методов во многом обусловлены тем, что сегодняшние глубокие нейросетевые архитектуры строятся на базе высокоразмерных нелинейных математических конструкций, зачастую выходящих за рамки доступного детерминированного анализа [83]. По мере наращивания глубины, количества параметров и усложнения внутренних взаимосвязей применение формальных доказательных техник становится всё более затруднительным, отчего в практических индустриальных кейсах заметно труднее убедительно обосновывать получаемые выводы.

Изучение роли формальных методов в реальных проектах ИИ показывает, что их математическая строгость не всегда позволяет удовлетворить динамич-

ные потребности бизнеса. Отсюда следует необходимость разрабатывать гибридные, более пластичные и объяснимые модели: они способны наследовать проверенные компоненты формальных подходов, но одновременно сохраняют требуемую адаптивность и умение обрабатывать многослойные, неструктурированные данные.

Современные архитектурные решения в области нейросетей напрямую влияют на степень интерпретируемости выдаваемых ими предсказаний. Среди наиболее применяемых моделей выделяют сверточные нейронные сети (CNN) и рекуррентные нейронные сети (RNN). CNN демонстрируют наибольшую эффективность при анализе визуальных данных, поскольку автоматически извлекают и сопоставляют пространственные паттерны различных уровней абстракции. В свою очередь, RNN лучше справляются с временными или текстовыми последовательностями, позволяя улавливать контекст и зависимые отношения в задачах обработки естественного языка [74].

Каждая разновидность нейросетевых архитектур сталкивается с собственными проблемами интерпретируемости. Чем сложнее топология, тем труднее модели обосновать выдаваемые предсказания. Глубокие модели, содержащие десятки или сотни слоёв, демонстрируют выдающуюся эффективность, но в ходе оптимизации утрачивают прозрачность: решения формируются через многоуровневые иерархии абстрактных признаков, слабо соотносимых с человеческой интуицией и каузальным рассуждением. Это обстоятельство становится серьёзным ограничением при требованиях к explainable AI, когда нужно предоставить верифицируемое, человекочитаемое обоснование конкретного вывода [84].

Сложности с интерпретируемостью ещё более усугубляются присущей нейронным сетям строгой датозависимостью. Для корректной тренировки и формирования достоверных заключений архитектуры нуждаются в масштабных, высококачественных и сбалансированных выборках. На практике же обеспечить требуемый объём и репрезентативность данных удаётся не всегда, что снижает результативность и затрудняет понимание внутренней логики модели [76].

Актуальные нейросетевые архитектуры, в частности трансформеры, значительно расширяют возможности анализа и моделирования сложных взаимосвязей в данных, открывая более гибкие интерпретационные подходы, однако они всё равно сталкиваются с трудностями при раскрытии глубинных статистических зависимостей [77].

В рамках нашего анализа мы систематизировали целый спектр вызовов, с которыми сталкиваются современные системы искусственного интеллекта и, в частности, глубокие нейронные сети. Указанные проблемы не только описывают актуальный уровень развития алгоритмов, но и задают траекторию их дальнейшего внедрения в индустриальные и научные домены. Тематика вызовов включает масштаб моделей, непредвзятость и контроль качества данных, что критически отражается на объяснимости, устойчивости и воспроизводимости создаваемых решений, а также на их соответствие требованиям этики и безопасности.

Высокая сложность ИИ-моделей остаётся серьёзным вызовом. Современные алгоритмы, особенно глубокие нейронные сети, достигают впечатляющих результатов в самых разных задачах, однако их многослойная архитектура и вычислительная громоздкость зачастую мешают пользователям понять принципы работы таких систем. Дефицит интерпретируемости порождает скепсис как у профессионального сообщества, так и у широкой публики. Особенно опасна непрозрачность в критически важных сферах — медицине и судебной практике, где она может привести к ошибочным диагнозам или несправедливым вердиктам.

Ограниченность ресурсов, как материальных, так и человеческих, наряду с внутренними организационными факторами, оказывает существенное воздействие на эволюцию и деплоймент систем искусственного интеллекта. Нехватка вычислительных мощностей, бюджетов и квалифицированных специалистов часто сдерживает разработку более интерпретируемых алгоритмов. Параллельно иерархические структуры, устоявшиеся регламенты и корпоративная культура затрудняют интеграцию инноваций, требующих трансформации бизнес-процессов и методологий.

Фундаментальный фактор результативности алгоритмов машинного обучения – высокое качество обучающего корпуса. Шумные, нерепрезентативные либо смещённые выборки вызывают систематические ошибки модели и ослабляют доверие пользователей. Поэтому критичны многоступенчатая предобработка, корректная разметка, статистический аудит и методы снижения Байеса, минимизирующие влияние артефактов на финальный вывод.

Пределы применения формальных подходов в сфере интерпретируемого ИИ заслуживают детального анализа. Да, строгие методы повышают доверие, обеспечивая верифицируемую надежность и прогнозируемость поведения систем, однако их внедрение связано с вычислительной сложностью, ресурсозатратностью и, как следствие, тормозящими масштабное промышленное распространение барьерами.

Мы также рассмотрели, как конкретная конфигурация нейросетевой архитектуры сказывается на прозрачности алгоритма. Различные топологии – от однослойных перцептронов до глубоких сверточных или трансформерных сетей – по-разному влияют на способность модели аргументировать свои выводы. Компактные конструкции, к примеру линейная регрессия или решающие деревья, чаще оказываются легко интерпретируемыми, но уступают по выразительной мощности. Напротив, многоуровневые, высокопараметрические сети демонстрируют выдающуюся точность, однако их трактуемость все еще вызывает сомнения.

Подытоживая выводы, наше исследование акцентирует критическую важность интегрированного подхода к преодолению вызовов, возникающих при внедрении искусственного интеллекта. Мы предлагаем сформировать согласованный комплекс мер, нацеленный на рост прозрачности и объяснимости алгоритмов, улучшение практик управления данными и устранение внутренних организационных препятствий. Такая программа действий укрепит доверие к ИИ-системам и обеспечит их более безопасное, ответственное и продуктивное применение в различных отраслях.

В ходе проведённого анализа различные исследователи выделяют ряд ключевых проблем современного искусственного интеллекта [74]:

1. сложность моделей ИИ;
2. ресурсные ограничения;
3. организационные причины;
4. ограничения формальных методов;
5. проблемы с данными;
6. сложности архитектур нейронных сетей;
7. границы и парадоксы математического фундамента;
8. вызовы, связанные с интерпретируемостью систем ИИ;
9. ключевые ограничения формальных алгоритмов верификации.

Указанные фундаментальные вызовы формируют стратегическую исследовательскую повестку в области ИИ, раскрывая не только существующие ограничения технологий, но и перспективные траектории прорыва к более зрелым, надежным и практически ценным архитектурам искусственного интеллекта, способным преодолеть актуальные барьеры.

В табл. 13. Для каждой фундаментально проблемы современного ИИ выполним декомпозицию и определим значение и пути устранения.

Таблица 13. Фундаментальные проблемы современного ИИ

Сложность моделей ИИ	
Комбинаторный взрыв возможных входов	Это эффект резкого («взрывного») роста временной сложности алгоритма при увеличении размера входных данных задачи. Связано это с тем, что рассматриваемый алгоритм не является полиномиальным, то есть время решения задачи не ограничено никаким многочленом от длины входа. Проблема комбинаторного взрыва была осознана в 1950-х годах, когда были разработаны первые программы ИИ, но её решение до сих пор не найдено. Чтобы справиться с комбинаторным взрывом, нужны алгоритмы, способные анализировать структуру целевой области и использовать преимущества накопленного знания за счёт эвристического поиска, долгосрочного планирования и свободных абстрактных представлений.
Комбинаторный взрыв возможных связей входов и выходов	Как только количество объектов превысит определённое число, рост количества комбинаторных объектов (например, путей в графе), которые перебираются в процессе решения, становится неудержимым. Трудности могут возникнуть не при проверке огромного количества объектов, а гораздо раньше — при пересчёте.

Нелинейность и сложность функций активации	Функция активации — это математические функции, которые определяют выход нейрона на основе его входа, внося нелинейность в модель. Это позволяет сети изучать сложные закономерности и взаимосвязи в данных, которые невозможны при использовании чисто линейной модели. Нелинейность означает, что взаимосвязь между входом и выходом не является прямой: результат не изменяется пропорционально входным данным. Это позволяет моделировать сложные шаблоны, которые невозможны при использовании чисто линейной модели. Например, сеть может создавать изогнутые границы принятия решений для правильного разделения данных, которые сложно разделить линейной функцией. Сложность функций активации зависит от их математических свойств и специфики задачи, для которой создаётся нейронная сеть. Например: Сигмоидальная функция (логистическая) — ограничивает выходные значения в диапазоне [0, 1], что полезно для задач классификации.
Стохастичность обучения и входных данных	Стохастичность в ИИ проявляется в разных аспектах: в процессе обучения моделей и в входных данных. Это означает введение случайности или вероятности в алгоритмы и модели, что позволяет учитывать неопределённость и эффективно обрабатывать зашумлённые или неполные данные. Обучение – стохастический градиентный спуск (SGD) — это алгоритм оптимизации, который вносит случайность в обновления параметров. Вместо вычисления градиента с использованием всего набора данных градиент оценивается с использованием случайно выбранного подмножества данных (мини-пакета). Такая случайная выборка привносит стохастичность в процесс оптимизации, делая его более адаптируемым к зашумленным или динамичным данным. Стохастические нейронные сети (SNN) — класс нейронных сетей, которые включают случайность в свою архитектуру и процессы обучения. **Входные данные**. Использование разных частей данных на разных итерациях помогает избежать застревания модели в локальном минимуме, особенно если оптимизируемая функция обычно невыпуклая.
Стохастичность обучения и входных данных	Учёт изменчивости данных — например, если распределение на данные изменяется со временем, система замедляется в обучении, так как ей приходится долго адаптироваться под изменяющиеся условия. В этом случае стохастичность помогает модели адаптироваться к изменяющимся условиям. Однако стохастическая природа многих алгоритмов ИИ может вызывать проблемы — результаты могут различаться даже при одних и тех же входных данных из-за случайной инициализации или внутренней изменчивости процессов обучения. Чтобы обеспечить согласованное поведение, необходимо использовать статистические методы или исправлять начальные исходные данные.

Динамическое обновление моделей на новых данных	Динамическое обновление моделей на новых данных в ИИ позволяет поддерживать актуальность и точность моделей в долгосрочной перспективе. Динамическое обновление моделей реализуется с помощью методов машинного обучения, алгоритмов и платформ, которые автоматизируют процесс.
Динамическое обновление моделей на новых данных	Один из подходов к динамическому обновлению — **Retrieval-Augmented Generation (RAG)**. При использовании этого метода генеративная языковая модель (LLM) снабжается доступом к внешним источникам информации. Модель перед генерацией ответа выполняет поиск релевантных данных и использует найденные сведения при формировании ответа.
Трудности с моделированием непрерывных пространств состояний	Некоторые трудности, с которыми сталкиваются при моделировании непрерывных пространств состояний в искусственном интеллекте. **Невозможность полного перебора**. Даже для сравнительно простых задач пространство решений чрезвычайно велико, и исследование всего пространства невозможно. Этот эффект называют комбинаторным взрывом или проклятием размерности. **Отсутствие оценивающих функций**. Для некоторых задач не существует функций, которые бы позволяли на каждом шаге гарантированно выбирать лучший ход. **Сложность выражения неформальных знаний**. Особенно сложно выразить такие знания в формальных логических терминах, если они не являются полностью достоверными.
Трудности с моделированием непрерывных пространств состояний	**Нестабильность существующих моделей**. Они могут быть нестабильными или требовать значительного количества вычислительных ресурсов. Для решения этих проблем, например, разработали «линейные колебательные модели пространства состояний» (LinOSS). Этот подход обеспечивает стабильные и вычислительно эффективные прогнозы без чрезмерно ограничивающих условий для параметров модели.
Разнообразие архитектур CNN, RNN, Transformer и др.	Разнообразие архитектур нейросетей ИИ включает разные типы, каждая из которых предназначена для решения определённого круга задач. **Сверточные нейросети (CNN)**. Используют свёртки (фильтры), чтобы находить шаблоны в данных — от простых до сложных. Применяются для распознавания изображений и видео, медицинской диагностики, обнаружения объектов. **Рекуррентные нейросети (RNN)**. Сохраняют состояние при обработке последовательностей — могут учитывать контекст. Используются для распознавания речи, обработки текста, анализа временных рядов. Популярные виды: LSTM, GRU.
Разнообразие архитектур CNN, RNN, Transformer и др.	**Трансформеры**. Используют механизм внимания и параллельные вычисления для эффективной обработки последовательных данных. Применяются в машинном переводе, генерации текста, чат-ботах и многих других приложениях.

Ограничения формальных методов	
Высокая вычислительная сложность проверки	Высокая вычислительная сложность проверки в ИИ связана с несколькими факторами. **Большой объём и сложность данных**, на которых обучаются модели ИИ. Например, GPT-4 был обучен более чем 570 ГБ текстовых данных из разных источников. Это затрудняет отслеживание и понимание каждого аспекта. **Динамическая природа моделей ИИ**. Они постоянно учатся и развиваются, что приводит к результатам, которые могут меняться с течением времени. Модель может адаптироваться к новым входным данным или взаимодействиям с пользователем. **Сложность интерпретируемости моделей ИИ**. Многие модели ИИ, особенно глубокое обучение, часто считаются «чёрными ящиками» из-за их сложности, из-за которой трудно понять, как генерируются конкретные результаты.
Высокая вычислительная сложность проверки	**Ресурсоёмкость комплексного аудита ИИ**. Он требует значительных вычислительных мощностей, квалифицированного персонала и времени. **Высокая вычислительная сложность традиционных нейронных сетей** также является препятствием на пути к их широкому применению на мобильных устройствах и встраиваемых системах. Для решения этих проблем используются различные стратегии, например, регулярный мониторинг и тестирование, повышение прозрачности и объяснимости, включение человеческого контроля в разработку и аудит ИИ.
Ограниченность существующих методов в отношении масштабных моделей	Некоторые ограничения существующих методов в отношении масштабных моделей ИИ. **Неспособность полноценно представлять сложные концепции.** Современные LLM оперируют статистическими закономерностями в данных, но не обладают глубоким пониманием контекста или абстрактных идей. **Значительные ресурсные затраты**. Тренировка моделей требует огромных вычислительных мощностей, что делает их создание и использование доступным только для крупнейших корпораций.
Ограниченность существующих методов в отношении масштабных моделей	**Отсутствие адаптивности**. Модели плохо справляются с задачами, которые выходят за рамки их тренировочных данных, и часто демонстрируют ошибки в логике и рассуждениях. **Закон убывающей отдачи**. С каждым новым этапом увеличения размеров модели улучшения становятся всё менее значимыми, в то время как затраты растут экспоненциально. **Технические ограничения**. Современное оборудование уже достигло пределов своих возможностей, и дальнейшее наращивание мощностей требует принципиально новых технологий. **Экологические последствия**. Обучение больших моделей требует огромного количества энергии, что вызывает серьёзные экологические проблемы.

	Экономическая неустойчивость. Только крупнейшие компании могут позволить себе разработку таких систем, что создаёт монополизацию рынка и ограничивает инновации.
Ограниченность существующих методов в отношении масштабных моделей	Для преодоления этих ограничений специалисты предлагают новые подходы, среди которых интеграция различных методов машинного обучения, развитие вероятностного программирования, модульные архитектуры и фокус на интерпретируемости
Проблемы с полнотой формальных методов	Некоторые проблемы, связанные с полнотой формальных методов в ИИ. **Неопределённость в больших языковых моделях**. В отличие от формальных методов, которые обеспечивают строгие математические гарантии точности и надёжности, большие языковые модели строят ответы на основе вероятностных распределений. Это приводит к тому, что формальный вывод с их использованием сталкивается с проблемами неопределённости. **Невозможность полного перебора на экспоненциально растущем дереве вариантов**. Этот эффект называется комбинаторным взрывом или проклятием размерности. Невозможность полного перебора лишает переборные алгоритмы интеллектуальности и затрудняет решение проблем ИИ.
Проблемы с полнотой формальных методов	**Возникновение задач, которые не могут б**ыть сразу представлены и решены с помощью формальных математических методов. В таких ситуациях имеет место большая начальная неопределённость проблемной ситуации и многокритериальность. **Сложность создания интеллектуальных систем с активными элементами.** Она обусловлена необходимостью поиска компромисса между целостностью представления системы и детализацией описания её компонентов в процессе разработки и реализации.
Отсутствие единого подхода для разных типов данных (текст, изображения, звук	Отсутствует единый подход к обработке разных типов данных (модальностей) в традиционных моделях ИИ, которые работают с одним типом данных за раз. Такие модели ограничены в понимании сложной реальности. Например, система, обученная исключительно на текстовых данных, не сможет интерпретировать визуальные или звуковые сигналы, а визуальная система не распознаёт смысл сказанной фразы.
Отсутствие единого подхода для разных типов данных (текст, изображения, звук	Для решения этой проблемы существует **мультимодальное обучение** — подход в машинном обучении, при котором модель обучается одновременно на нескольких типах данных. Главная цель — научить систему понимать, как разные модальности соотносятся друг с другом, и использовать это понимание для решения более сложных задач. Некоторые примеры применения мультимодального обучения. **Анализ видеоматериалов**. Система обрабатывает не только визуальный контент, но и диалоги, окружающие звуки и сопровождающие субтитры.

	Создание персонализированного контента. Мультимодальные системы позволяют создавать согласованный контент в различных форматах (текст, изображения, видео) с учётом предпочтений пользователей и контекста использования.
Проблемы с формализацией вероятностных моделей	Некоторые проблемы, связанные с формализацией вероятностных моделей в ИИ. **Качество данных**. ИИ-модели обучаются на огромных наборах данных, но они могут быть неполными, неправильными или предвзятыми. Если модель обучалась на данных, которые не охватывают все возможные ситуации или содержат ошибки, её предсказания также будут ошибочными.
Проблемы с формализацией вероятностных моделей	**Предвзятость данных**. Модель, обученная на искажённых или однотипных данных, начинает делать предвзятые выводы. Например, в алгоритмах рекрутинга предвзятость может приводить к предпочтению кандидатов определённого пола или этнической принадлежности. **Сложность поставленной задачи**. Некоторые задачи требуют понимания контекста, ассоциаций и опыта, которых у модели нет. ИИ часто «думает» только в пределах данных, которые получил, и не всегда способен обобщать и адаптироваться к новым условиям. **Технические ограничения архитектуры**. Даже самые передовые архитектуры ИИ, такие как глубокие нейронные сети, имеют технические ограничения. У модели есть определённое количество слоёв, нейронов и весов, которые ограничивают её способность к анализу и синтезу информации.
Проблемы с формализацией вероятностных моделей	**Ограниченность в проверке данных**. Модели ИИ, как правило, не умеют проверять правдивость информации и полагаются на вероятностные вычисления. Это означает, что иногда они «угадывают» ответ, что особенно рискованно при ответах на вопросы, связанных с фактами.
Проблемы с верификацией эвристик и приближений в моделях	Некоторые проблемы с верификацией эвристик и приближений в моделях ИИ. **Сложность проектирования эвристики**. Эвристика должна точно оценивать затраты на достижение цели, не будучи слишком консервативной или агрессивной. Плохо разработанная эвристика может привести к неэффективному поиску или неоптимальным решениям. **Специфический характер проблемы**. Эвристические функции часто адаптируются к конкретным задачам, что ограничивает их обобщение. Эвристика, которая хорошо работает для конкретного сценария, может быть неприменима в другом контексте.
Проблемы с верификацией эвристик и приближений в моделях	**Вычислительные издержки.** Вычисление сложной эвристики на каждом шаге может привести к дополнительным вычислительным затратам. Если стоимость вычисления эвристики перевешивает её преимущества, это может не повысить общую производительность.

	Риск неоптимальных решений. Недопустимые эвристические методы, хотя и более быстрые, могут привести к неоптимальным решениям. **Невозможность проверить все возможные входные данные**. Традиционно проверка нейронных сетей сосредоточена на тестировании — оценке сети на большом наборе точек во входном пространстве и определении того, соответствуют ли её выходы желаемым на этом наборе. Однако невозможно проверить все возможные входные данные, поскольку входное пространство очень часто фактически бесконечно по мощности. **Сложность адаптации традиционных методов тестирования**. Адаптировать эти подходы для тщательного тестирования ИИ-моделей сложно из-за масштаба и отсутствия структуры в моделях, способных содержать сотни миллионов параметров.
Ограниченность методов для адаптивных моделей	Некоторые ограничения методов для адаптивных моделей в ИИ. **Переобучение**. Модель демонстрирует высокие показатели на известных данных и проваливается на новых. Это происходит, когда модель утрачивает способность к обобщению и начинает подстраиваться под каждый элемент обучающей выборки. Переобученная модель становится негибкой: она не может быть до обучена, плохо переносит новые задачи, требует полного переобучения даже при малых изменениях. **Адаптация на примерах**. Этот метод позволяет быстро адаптировать модель без долгого обучения. Однако у него есть ограничения: ответы зависят от качественного промпта, модель не «запоминает» информацию между запросами, невозможно корректировать веса модели. **Ограничения ИИ-аватаров**. Многие модели сосредоточены на узком спектре взаимодействий и не обладают достаточной адаптивностью для комплексного моделирования игрового процесса.
Ограниченность методов для адаптивных моделей	Аватары остаются ограниченными в плане динамического взаимодействия с пользователем, особенно в обучающих модулях для игр, где требуется быстрая адаптация и погружение игрока в игру. Выбор метода зависит от целей, бюджета и доступных вычислительных мощностей.
Невозможность предсказать поведение модели на всех возможных входах	Невозможность предсказать поведение модели на всех возможных входах в ИИ связана с сложностью нейронных сетей и непредсказуемостью процесса обучения. Некоторые примеры такого поведения. **Генеративные модели**, например большие языковые модели (LLM), могут «галлюцинировать», выдавая фактически неверную, но уверенно сформулированную информацию. Это происходит не из-за недостатка данных, а из-за того, что модель генерирует наиболее вероятное продолжение последовательности, которое может не соответствовать реальности.

	Системы компьютерного зрения уязвимы к состязательным атакам, когда минимальные, незаметные для человеческого глаза изменения во входных данных (например, несколько пикселей на изображении) заставляют модель полностью ошибиться в классификации.
Невозможность предсказать поведение модели на всех возможных входах	**Современные ИИ-модели** слабо распознают и интерпретируют поведение людей в динамических социальных взаимодействиях, что ограничивает применение искусственного интеллекта в реальной среде. Даже при идентичных условиях обучения две, казалось бы, одинаковые модели могут демонстрировать тонкие, но значимые различия в поведении. Эта проблема подрывает доверие к системам ИИ, особенно в критически важных областях, таких как медицина, автономное вождение или финансовые рынки, где ошибка недопустима.
Ресурсные ограничения	
Недостаток вычислительных ресурсов	Недостаток вычислительных ресурсов в ИИ связан с несколькими факторами. **Рост спроса на вычислительные мощности**. С момента появления ChatGPT в 2023 году спрос на вычислительные ресурсы резко вырос. Облачные платформы, технологические гиганты и стартапы наращивают загрузку дата-центров, обучая и разворачивая всё более сложные модели. **Повышенное энергопотребление нейровычислений**. Для работы компьютерных систем, на которых обучают нейросети, требуется дополнительное специализированное оборудование с повышенным энергопотреблением.
Недостаток вычислительных ресурсов	**Перегрузка энергосистем**. Например, в США энергосистема PJM испытывает острую нехватку мощностей на фоне стремительного роста потребления со стороны вычислительных центров и генеративного ИИ. **Перспектива масштабного внедрения автономных ИИ-моделей с локальным обучением**. Поддержка таких сценариев требует стабильного доступа к значительным объёмам электроэнергии, чего перегруженные сети обеспечить уже не могут. По прогнозам летом 2025 года в регионах высокой концентрации дата-центров прогнозируется рост тарифов на электроэнергию более чем на 20%. **Высокие цены и нестабильность электроснабжения в пиковые периоды** грозят замедлить внедрение ИИ-сервисов и повысить операционные расходы их владельцев.
Высокая стоимость проведения формальной верификации	Формальная верификация в области ИИ может быть высокой из-за сложности алгоритмов и проблем с масштабированием. Традиционно проверка систем ИИ сосредоточена на тестировании — оценке сети на большом наборе точек во входном пространстве и определении, соответствуют ли её выходы желаемым на этом наборе.

Высокая стоимость проведения формальной верификации	Однако формальная верификация предоставляет математические гарантии в отношении верной работы системы для любого входа из пространства возможных входных данных, но требует разработки алгоритмов, подтверждающих соответствие модели желаемым спецификациям для всех возможных входных данных. **Причины.** **Сложность алгоритмов**. С масштабированием и усложнением ИИ-систем становится всё сложнее проектировать алгоритмы, достаточно адаптированные к модели ИИ. **Проблемы с перебором** всех возможных вариантов выходных данных для заданного набора входных (например, незначительных изменений изображения). В крупномасштабных моделях перебор всех возможных вариантов трудно реализуем из-за астрономического количества возможных изменений. **Трудности с формулировкой** спецификаций — «правильное» поведение ИИ-систем часто трудно сформулировать точно. **Методы.** Для формальной верификации ИИ используются, например:
Высокая стоимость проведения формальной верификации	**Дедуктивная верификация** — позволяет свести задачу проверки корректности программы относительно спецификаций к задаче проверки истинности утверждений о том, что программа корректна относительно спецификаций. Такие утверждения — логические формулы, и если все условия корректности истинны, то программа корректна относительно спецификаций. **Проверка на модели (model checking)** — создание математической модели искусственной нейронной сети и её поведения, описание архитектуры сети, функций активации, весов и других параметров. Стоимость формальной верификации зависит от сложности алгоритмов и сложности проверяемой архитектуры ИИ. **Решения.** Выполняются работы над **снижением стоимости формальной верификации**, например: Разрабатывают алгоритмы, адаптированные к модели ИИ. Например, используют методы понижения размерности (дропаут, прунинг) для уменьшения времени работы алгоритмов.
Высокая стоимость проведения формальной верификации	**Стандартизируют алгоритмы верификации** — это позволит обеспечить простыми в использовании инструментами, которые проливающими свет на возможные режимы отказа системы ИИ до того, как этот отказ приведёт к обширным отрицательным последствиям. **Расширяют спектр состязательных примеров** — это поможет разрабатывать подходы к верификации для свойств, имеющих отношение к реальному миру.

Длительность времени, необходимая для верификации больших моделей	Верификация больших моделей в области ИИ может занимать длительное время из-за сложности адаптации традиционных методов к сложным моделям, содержащим сотни миллионов параметров. Это связано с необходимостью проверки соответствия модели спецификациям, которые описывают желаемые свойства системы (устойчивость к малым изменениям входных данных, ограничения по безопасности и др.). **Причины.** **Масштаб и отсутствие структуры в моделях**. Адаптировать традиционные подходы (модульное тестирование, формальную верификацию) сложно из-за масштаба и отсутствия структуры в моделях.
Длительность времени, необходимая для верификации больших моделей	**Сложность проверки** всех возможных входных данных. Входное пространство часто бесконечно по мощности, и невозможно проверить все возможные данные. **Необходимость формального доказательства** соответствия модели спецификациям. Необходимо разработать алгоритмы, подтверждающие соответствие модели желаемым спецификациям для всех возможных входных данных, но эти подходы нелегко масштабировать на современные модели. **Методы.** **Разработка новых подходов для верификации сложных моделей**. Например, изучение техник оценки соответствия модели спецификациям, которые требуют от неё разработчик и пользователи. **Пересмотр алгоритмов обучения** так, чтобы они не только хорошо работали на тренировочных данных, но и соответствовали желаемым спецификациям. **Использование разных стратегий валидации модели**, например, кросс-валидации, чтобы точно оценить её обобщающую способность. Это позволяет избежать переобучения и увеличивает вероятность того, что модель будет хорошо работать на новых данных.
Длительность времени, необходимая для верификации больших моделей	**Инструменты.** **Автоматизированные инструменты для верификации**, которые упрощают процесс и позволяют проводить аудит в режиме реального времени. **Прозрачность и объяснимость моделей**, например, использование структур интерпретируемости моделей и объяснимого ИИ (XAI). Это помогает аудиторам понять процессы принятия решений и выявить потенциальные проблемы. **Включение человеческого контроля в разработку и аудит ИИ**, чтобы выявить проблемы, которые автоматизированные системы могут упустить.
Отсутствие инструментов, совместимых с реальными системами ИИ	Верификация больших моделей в области ИИ может занимать длительное время из-за сложности адаптации традиционных методов к сложным моделям, содержащим сотни миллионов параметров. Это связано с необходимостью проверки соответствия модели спецификациям, которые описывают желаемые свойства системы (устойчивость к малым изменениям входных данных, ограничения по безопасности и др.).

Отсутствие инструментов, совместимых с реальными системами ИИ	**Причины.** **Масштаб и отсутствие структуры в моделях**. Адаптировать традиционные подходы (модульное тестирование, формальную верификацию) сложно из-за масштаба и отсутствия структуры в моделях. **Сложность проверки** всех возможных входных данных. Входное пространство часто бесконечно по мощности, и невозможно проверить все возможные данные. **Необходимость** формального доказательства соответствия модели спецификациям. Необходимо разработать алгоритмы, подтверждающие соответствие модели желаемым спецификациям для всех возможных входных данных, но эти подходы нелегко масштабировать на современные модели. **Методы.** **Разработка новых** подходов для верификации сложных моделей. Например, изучение техник оценки соответствия модели спецификациям, которые требуют от неё разработчик и пользователи. **Пересмотр алгоритмов** обучения так, чтобы они не только хорошо работали на тренировочных данных, но и соответствовали желаемым спецификациям.
Отсутствие инструментов, совместимых с реальными системами ИИ	**Использование разных** стратегий валидации модели, например, кросс-валидации, чтобы точно оценить её обобщающую способность. Это позволяет избежать переобучения и увеличивает вероятность того, что модель будет хорошо работать на новых данных. **Инструменты.** **Автоматизированные инструменты для верификации**, которые упрощают процесс, позволяя проводить аудит в режиме реального времени. **Прозрачность и объяснимость моделей**, например, использование структур интерпретируемости моделей и объяснимого ИИ (XAI). Это помогает аудиторам понять процессы принятия решений и выявить потенциальные проблемы. **Включение человеческого контроля в разработку и аудит ИИ**, чтобы выявить проблемы, которые автоматизированные системы могут упустить.
Несовместимость с популярными ML-фреймворками (TensorFlow. PyTorch, Visual Studio)	Несовместимость с популярными ML-фреймворками (TensorFlow, PyTorch) и средой разработки Visual Studio может возникать при верификации больших моделей в ИИ из-за различий в версиях, зависимостях и функциях. Это приводит к проблемам с переносом моделей между фреймворками и запуском их на разном оборудовании (GPU, TPU) без серьёзных доработок. **TensorFlow.** **Проблемы обратной совместимости**. Старые исследования в TensorFlow 1 не всегда сочетаются с новыми возможностями в TensorFlow 2. Например, модель, обученная в TensorFlow 1, не может быть легко использована в TensorFlow 2 без сложных преобразований.

	Ошибки, связанные с CUDA. Версия TensorFlow несовместима с версией драйвера NVIDIA или архитектурой GPU. Решение: нужно устанавливать сборку под нужную версию CUDA (например, cu118 или cu121). **PyTorch.** **Конфликты версий**. Например, при установке библиотек, которые зависят от конкретных версий PyTorch, возникают ошибки.
Несовместимость с популярными ML-фреймворками (TensorFlow. PyTorch, Visual Studio)	**Решение**: рекомендуется использовать метод чистой установки, чтобы выровнять версии. Для сложных настроек можно использовать среды conda или контейнеры **Visual Studio.** **Несовместимость проектов**. Например, проект, обученный в TensorFlow, может быть несовместим с текущей версией Visual Studio. Решение: нужно скорректировать код для правильной работы, если это возможно. **Ошибки, связанные с зависимостями**. Например, конфликты зависимостей (pip долго думает или выдаёт ошибку). Решение: можно использовать conda или инструменты вроде Poetry, которые имеют более продвинутый механизм разрешения зависимостей, или попробовать установить проблемный пакет первым в чистом окружении.
Недостаток вычислительных ресурсов	Недостаток вычислительных ресурсов в ИИ касается всех, кто применяет или планирует применять искусственный интеллект. **Некоторые причины проблемы**. **Спрос на вычислительные мощности растёт**. Облачные платформы, технологические гиганты и стартапы наращивают загрузку дата-центров, обучая и разворачивая всё более сложные модели.
Недостаток вычислительных ресурсов	**Повышенное энергопотребление нейровычислений**. Для работы компьютерных систем, на которых обучают нейросети, требуется дополнительное специализированное оборудование с повышенным энергопотреблением. **Дефицит чипов**. Крупные компании обучают и тренируют множество моделей, для чего нужны соответствующие чипы. Один из способов решения проблемы — повышение эффективности использования существующих ресурсов. Сюда входит оптимальное применение оборудования, распределённые вычисления во времени или между командами, приоритизация проектов, требующих мощностей, минимизация простоев мощностей и поиск мощностей у партнёров или коллег. Также некоторые разработчики AI-моделей (Google, OpenAI, Microsoft, Amazon, Alibaba) представили собственные чипы для ИИ или объявили о работе над ними. Российские ученые разработали технологию (аналогов нет) для создания процессоров нового поколения. Она базируется на инновационном методе формирования логических элементов с точностью до 0,2 ангстрема (0,02 нм).

	Описание разработки опубликовано в Science Advances («Научные достижения» Smirnov et al., Sci. Adv. 11, eads9744 (2025) 2025), 7 May 2025). Технологию запатентовали в России, и в настоящий момент ведется ее патентование в других странах.
Требование к глубокому знанию формальных методов со стороны разработчиков	Глубокое знание формальных методов (детерминированных или стохастических) и/или методов моделирования и тестирования необходимо разработчикам ИИ для решения ряда задач. Некоторые из них: **Формулировка предположений об окружающей среде системы**. Это важно для правильного функционирования ИИ. **Определение ожидаемых функциональных возможностей и ограничений**. **Оценка существенных характеристик системы**: корректности, надёжности, вероятности ошибок и других. Кроме того, разработчики ИИ должны разбираться в нейронных сетях, обучении с подкреплением и концепциях обработки естественного языка. Также важно, чтобы разработчики могли обрабатывать и интерпретировать большие наборы данных. Это помогает обеспечить правильное обучение инструментов ИИ, что приводит к более точным и эффективным результатам.
Требование к глубокому знанию формальных методов со стороны разработчиков	Таким образом, глубокое знание формальных методов позволяет разработчикам ИИ эффективно взаимодействовать с инструментами, выполнять индивидуальные настройки и обеспечивать надёжность и эффективность автоматизированных функций.
Ограничения реального времени	Некоторые ограничения ИИ, которые могут влиять на работу в реальном времени: **Синхронная работа**. Некоторые ИИ-модели обрабатывают и отвечают на каждый ввод последовательно, по одному за раз. Это может стать ограничением в сценариях, требующих взаимодействия в реальном времени или одновременной обработки нескольких запросов. **Ограничение контекста**. Большие языковые модели (LLM) ограничены размером контекстного окна, из-за чего им приходится «забывать» информацию, полученную в начале долгих диалогов.
Ограничения реального времени	**Невозможность выходить в Интерне**т. LLM не могут просматривать веб-страницы или вызывать веб-службы, поэтому они ограничены данными, на которых их обучали, и не имеют возможности получать или проверять информацию из живых веб-источников в режиме реального времени. **Сложности с передачей данных между уровнями памяти**. Это может замедлять работу при увеличении контекста. Для решения некоторых из этих ограничений используют, например, технологию Helix Parallelism от Nvidia. Она расширяет «память» моделей, позволяя им в реальном времени обрабатывать и анализировать большие объёмы данных.

Проблемы с верификацией сетей, обученных на больших объемах данных	**Некоторые проблемы с верификацией сетей, обученных на больших объёмах данных в ИИ**. **Сложность перебора** всех возможных вариантов выходных данных. В крупномасштабных моделях это трудно реализуемо из-за астрономического количества возможных изменений. **Невозможность проверить** все возможные входные данные. Часто входное пространство фактически бесконечно по мощности.
Проблемы с верификацией сетей, обученных на больших объемах данных	**Трудность с интерпретируемостью** моделей. Сложные модели машинного обучения тяжело интерпретировать, что создаёт проблемы с прозрачностью принятия решений. **Использование контента, сгенерированного моделями, при обучении**. Это может привести к необратимым дефектам и коллапсу модели, когда она неправильно воспринимает реальность. Для решения этих проблем специалисты предлагают, например, качественный контроль экспертами исходных данных. Также важно создавать стандартизированные и объективные классификаторы оценки производительности ИИ-моделей и качества датасетов.
Проблемы с данными	
Шумность данных и возможность ошибок при разметке	**Шумность данных и возможность ошибок при разметке в ИИ** — важные проблемы, которые влияют на производительность моделей машинного обучения. Эти проблемы связаны с наличием в данных ошибочной, нерелевантной или плохо записанной информации, а также с ошибками в разметке. **Шум в данных** — это случайные или систематические вариации, которые не несут полезной информации и искажают истинные закономерности. Некоторые типы шума:
Шумность данных и возможность ошибок при разметке	**Случайный** — данные содержат непредсказуемые значения без определённой закономерности. Например, ошибки измерений в датчиках или неожиданные изменения. **Неправильная маркировка** — неправильные этикетки влияют на обучение модели, его способность к обобщению сокращается. **Отсутствующие или неполные данные** — иногда в некоторых выборках отсутствует ключевая информация, что может исказить результаты. **Выбросы — значения**, которые существенно отличаются от остального набора данных, могут отрицательно повлиять на производительность модели. **Методы снижения шума**. **Предварительная обработка данных** — очистка данных, удаление выбросов, обработка отсутствующих данных и исправление неправильных маркировок. **Нормализация и масштабирование** — преобразование данных в единый масштаб помогает минимизировать влияние шума на некоторые алгоритмы машинного обучения.

Шумность данных и возможность ошибок при разметке	**Использование надёжных моделей** — некоторые модели, например алгоритмы, основанные на деревьях решений или нейронные сети, могут лучше обрабатывать зашумлённые данные. **Ошибки при разметке.** Ошибки в разметке могут возникать по следующим причинам. **Неправильных или непоследовательных меток** — например, один аннотатор может обозначить транспортное средство как «легковой автомобиль», а другой — как «грузовик». **Отсутствующих ярлыков** — аннотатор может не отметить кошку на изображениях. **Неверной интерпретации инструкции** — инструкции, предоставленные аннотаторам, не ясны. **Методы устранения ошибок.** **Создание хорошо документированных стандартов маркировки и проведение проверок контроля качества**. **Постоянное обучение аннотаторов и использование консенсусной маркировки**, когда несколько аннотаторов просматривают каждый образец.
Невозможность охватить все возможные случаи тестовыми данными	Невозможность охватить все возможные случаи тестовыми данными при использовании ИИ связана с ограниченностью обучающей выборки алгоритмов машинного обучения. Если данные, на которых учился ИИ, не охватывают конкретную область знаний или события, связанные с уникальными контекстами, ответ системы будет поверхностным или неверным. Это может проявляться в разных областях, например: В тестировании программного обеспечения — ИИ не может автоматически генерировать тестовые случаи для всех возможных сценариев, которые традиционное тестирование может пропустить. В ответах на вопросы — если вопрос выходит за рамки известных данных, система не знает, что ответить. **Причины.** **Ограниченность обучающих данных.** ИИ обучается на огромных наборах информации, но, если вопрос выходит за рамки известных данных, система не знает, что ответить. **Невозможность понимания контекста.** Хотя современные модели хорошо справляются с языковыми задачами, они не всегда способны понять глубокий контекст или подтекст вопроса.
Невозможность охватить все возможные случаи тестовыми данными	**Сложности с неоднозначными запросами.** Когда вопрос содержит двойной смысл или требует интуитивного понимания, алгоритм часто предлагает стандартный, обобщённый ответ. **Решения.** **Сбор разнообразных данных**. Они должны отражать различные сценарии тестирования, пограничные случаи и реальное поведение пользователей, чтобы модель ИИ могла обобщать и адаптироваться к различным условиям.

	Постоянное совершенствование. Нужно постоянно добавлять свежие данные в модели ИИ, чтобы повысить их точность и адаптировать к меняющимся средам приложений. **Автоматизация генерации тестовых случаев**. Инструменты на базе ИИ анализируют программное обеспечение и генерируют соответствующие тестовые случаи, охватывая пограничные случаи и пользовательские сценарии.
Этические и правовые ограничения на использование некоторых наборов данных	**Этические ограничения.** **Недискриминаци**я. Алгоритмы и наборы данных не должны приводить к умышленной дискриминации отдельных лиц или групп лиц. Рекомендуется не учитывать такие факторы, как национальная, языковая и расовая принадлежность, принадлежность к политическим партиям и общественным объединениям, вероисповедание и отношение к религии. **Непричинение вреда**. Недопустимо использовать технологии ИИ для причинения вреда жизни и здоровью человека, имуществу граждан и юридических лиц, окружающей среде. **Прозрачность и открытость принципов работы.** Рекомендуется публично декларировать факт применения рекомендательных систем, а также раскрывать цели и общие принципы их работы. **Использование данных о пользователе**. Должно основываться на законодательстве о защите персональных данных и учитывать требования иных нормативных актов. **Правовые ограничения.** **Авторские права**. Использование для обучения моделей ИИ наборов данных, которые защищены авторским правом, может нарушать исключительное право владельца авторских прав контролировать воспроизведение своих работ.
Этические и правовые ограничения на использование некоторых наборов данных	В России искусственный интеллект не наделяется авторскими правами, его рассматривают только в качестве инструмента при создании объекта авторских прав. **Защита персональных данных**. Разработчики ИИ должны соблюдать законодательство в области персональных данных и охраняемых законом тайн. В частности, использовать качественные и репрезентативные наборы данных, полученные без нарушения закона из надёжных источников. **Использование псевдоданных**. Например, в законе 123-ФЗ от 24 апреля 2020 года установлен специальный правовой режим в сфере ИИ, в котором указаны требования по защите персональных данных граждан и использования псевдоданных, собираемых в режиме анонимности.
Динамическое изменение входных данных в реальных приложениях	Динамическое изменение входных данных в реальных приложениях ИИ реализуется с помощью адаптивных алгоритмов, которые корректируют своё поведение или структуру в ответ на изменяющиеся входные данные.
Динамическое изменение входных данных в реальных приложениях	**Некоторые области применения таких алгоритмов.** **Обслуживание клиентов**. Чат-боты настраивают тон и контент в зависимости от настроений пользователей, истории или сложности проблемы.

	Здравоохранение. Виртуальные ассистенты персонализируют рекомендации на основе истории болезни пациента, симптомов и контекста. **Образование**. Адаптивные системы обучения изменяют подсказки и пояснения в соответствии с темпом обучения учащегося и пробелами в знаниях. **Автоматизация бизнеса**. Агенты ИИ динамически обновляют инструкции по автоматизации рабочего процесса на основе данных в режиме реального времени или отзывов пользователей. **Автономные автомобили**. Динамическое исполнение помогает ускорить обработку данных от камер и сенсоров, улучшая реакцию автомобилей на дорожные условия. **Голосовые помощники**. Уменьшение задержки в обработке запросов позволяет сделать взаимодействие с ИИ более естественным.
Динамическое изменение входных данных в реальных приложениях	**Облачные сервисы**. Оптимизация моделей в облаке снижает нагрузку на серверы, улучшая производительность и экономя ресурсы. **Кибербезопасност**ь. Быстрая обработка потоков данных помогает обнаруживать угрозы в реальном времени.
Отсутствие формальных спецификаций для требований к данным	Отсутствие формальных спецификаций для требований к данным в системах ИИ — проблема, которая мешает автоматизации процессов, связанных с разработкой и тестированием систем на базе ИИ. Это проявляется в том, что требования к данным не имеют унифицированных шаблонов, чётких ролей, условий и ожиданий, машиночитаемой структуры. В результате ИИ не способен вычленить из требований полноценные тест-кейсы, что срывает автоматизацию на этапе первого же документа. **Причины.** **Отсутствие стандартов**. Каждый проект пишет требования по-своему, исходя из личных предпочтений автора, а не здравого смысла или потребностей автоматизации. Это приводит к:
Отсутствие формальных спецификаций для требований к данным	− отсутствию уникальных ID требований; − непонятным правилам валидации; − правилам заполнения полей формы, которые нужно уметь «читать» между строк. **Сложность формулировки спецификаций**. Спецификации, описывающие «правильное» поведение ИИ-систем, часто трудно сформулировать точно. Это связано с сложным поведением систем и работой в неструктурированном окружении. **Решения.** **Требовать структурированную документацию**. Помогать внедрять шаблоны, изучать инструменты, которые умеют работать с реальной документацией. **Использовать инструменты на базе ИИ для анализа требований.** Они анализируют требования на предмет согласованности, полноты и точности, автоматически отмечают

	несоответствия или конфликты и предлагают улучшения или корректирующие действия. **Автоматизировать отслеживание требований**. ИИ анализирует отношения между требованиями и другими артефактами, такими как проектная документация, тестовые наборы и код. Это помогает поддерживать чёткий контрольный след и выявлять возникающие проблемы.
Отсутствие формальных спецификаций для требований к данным	**Использовать генерацию естественного языка (NLG) для документации требований**. Алгоритмы машинного обучения автоматически создают текстовые описания требований, что снижает объём требуемой ручной работы.
Влияние смещений (bias) в данных на результаты верификации	Смещение (bias) в данных влияет на результаты верификации в ИИ, приводя к искажённым или неточным выводам. **Некоторые примеры влияния смещения**. **Модель, обученная на изображениях дорожного движения в больших городах**. Если её развернуть в сельской местности, она может неправильно классифицировать необычные дорожные разметки или не обнаружить типы автомобилей, которые она никогда раньше не видела. **Модель распознавания лиц, обученная в основном на одной демографической группе**. Она может оказаться не в состоянии точно предсказать всех пользователей. **Нейронная сеть, принимающая решения о выдаче кредита малому предпринимательству** на основании исторически смещённых в пользу белокожих мужчин данных, чаще отказывает в кредите афроамериканцам и женщинам.
Влияние смещений (bias) в данных на результаты верификации	Смещение снижает точность ИИ и, следовательно, его потенциал. В приложениях, где точность очень важна, например в здравоохранении или безопасности, такие ошибки могут быть опасны. Для выявления и устранения смещения в данных в ИИ используют, например, аудит данных, тестирование производительности модели в различных сценариях, сбор большего количества образцов из недопредставленных сценариев.
Организационные причины	
Недостаток специалистов по формальной верификации в области ИИ	Недостаток специалистов по формальной верификации в области ИИ может быть связан с несколькими факторами. **Сложность адаптации традиционных методов тестирования**. Подходы, которые работают хорошо на традиционных программах, сложно применять для тестирования ИИ-моделей из-за их масштаба и отсутствия структуры. **Высокая сложность алгоритмов формальной верификации**. Время выполнения таких алгоритмов зависит от количества параметров нейронной сети, поданной на вход, и структуры свойств, которые нужно проверить.
Недостаток специалистов по формальной верификации в области ИИ	**Сложности с привлечением квалифицированных специалистов**. Например, в 2024 году сообщалось, что регулирующие органы Италии сталкиваются с проблемами при найме экспертов по ИИ. Это связано с низким вознаграждением, длительными сроками найма и бюрократическими препятствиями при получении рабочих виз.

	Формальная верификация — процесс математического доказательства соответствия программы или системы заданным спецификациям. В случае ИИ-сетей это означает доказательство того, что поведение сети соответствует определённым ожиданиям и безопасным границам. Поскольку ИИ продолжает развиваться быстрыми темпами, спрос на квалифицированных специалистов, способных решать сложные этические, юридические и технические проблемы, которые он представляет, будет только возрастать.
Нежелание компаний инвестировать в сложные и дорогостоящие методы	**Снижение окупаемости вложений**. Некоторые разработчики ИИ отмечают, что новые модели не показывают той производительности, которую от них ожидают. При этом затраты на создание и обслуживание новых моделей растут.
Нежелание компаний инвестировать в сложные и дорогостоящие методы	**Сложности с поиском качественных данных**. Для создания продвинутых ИИ-моделей нужно всё больше неадаптированных массивов данных, сделанных человеком. Найти такие данные становится всё сложнее. **Отсутствие квалифицированных специалистов**. Для успешного внедрения ИИ необходимы квалифицированные кадры, а найти таких специалистов трудно. **Необходимость в дорогостоящих центрах** обработки данных. Например, у небольших компаний нет возможности инвестировать в создание больших языковых моделей, так как для этого нужны такие центры. **Расходы на государственную регистрацию**. Например, при внедрении ИИ-решений в медицинскую сферу нужно пройти сложную и затратную процедуру регистрации. Однако, по мнению некоторых экспертов, проблемы с финансированием ИИ-проектов решаемы. Например, источниками финансирования могут быть государственные программы поддержки инноваций, субсидии и гранты.
Конкурентное давление	Конкурентное давление в области ИИ — это стимул для технологических компаний и стартапов бороться за то, чтобы первыми запустить более мощные инструменты ИИ, позволяющие им получать награды (венчурные инвестиции, внимание СМИ, регистрацию пользователей). Это может проявляться в разных аспектах, например: **Борьба за лидерство в разработке ИИ-решений**. Например, западные компании стремятся к доминированию в области ИИ, а китайские конкуренты быстро перенимают и адаптируют перспективные технологии с меньшими затратами. **Формирование стандартов** — общепризнанных принципов разработки и использования ИИ-решений, которые используются как инструмент конкурентной борьбы. Подходы к стандартам отличаются от страны к стране: например, Китай внедряет стандарты, направленные на ограничение экспансии внешних технологий, а США устанавливает собственные стандарты как универсальные и международно-признанные. **Причины.** **Быстрое развитие технологий ИИ** и обещание, которое они несут для преобразования различных аспектов бизнеса и повседневной жизни.

Конкурентное давление	**Стремление к технологическому лидерству** — кто сможет освоить ИИ быстрее других, тот и получит стратегическое преимущество. **Последствия.** Поспешная разработка ИИ — компании выделяют как можно больше ресурсов на повышение мощности своих систем, пренебрегая при этом исследованиями в области безопасности. Например, в феврале 2023 года компания Microsoft поспешила с выпуском чат-бота Bing, в результате он стал угрожать пользователям и оскорблять их. **Неравномерное распределение центров и мощностей развития ИИ** — развитие ИИ идёт в странах, которые имеют ресурсы или возможности в виде высокого научного потенциала. Это влияет на менее развитые страны, которые зависят от чужих технологий и часто привязаны к их поставщикам. **Меры.** **Ускорение разработки продуктов** — компании, создающие флагманские языковые модели, должны выстраивать вокруг себя инфраструктуру, чтобы при появлении китайского аналога пользователям было труднее переходить на него. Например, за счёт глубокой интеграции своих моделей с широко используемыми программными решениями других разработчиков.
Конкурентное давление	**Усиление патентной защиты** — компании, контролирующие базовые модели, могут устанавливать ограничения на то, как другие компании используют их для последующих приложений, и взимать за доступ.
Отсутствие стандартов формальной верификации для ИИ	Сложности с адаптацией традиционных методов формальной верификации для тестирования моделей ИИ. Это связано с масштабом и отсутствием структуры в моделях, которые могут содержать сотни миллионов параметров. Для решения этой проблемы разрабатываются новые подходы, в том числе направленные на поддержание высокой производительности при внедрении криптографической или образцовой верификации. Некоторые из них: **EZKL**. Открытая система для создания, проверяемого ИИ и аналитики с использованием доказательств нулевого разглашения (ZKPs). Позволяет разработчикам доказать, что модели ИИ были выполнены правильно, не раскрывая чувствительные данные или собственные детали модели.
Организационные причины	
Отсутствие стандартов формальной верификации для ИИ	**Proof of Spontaneous Proofs** (PoSP). Обеспечивает эффективную верификацию, способен поддерживать масштабные децентрализованные рабочие нагрузки для ИИ, обеспечивая проверяемость и доверие высокопроизводительных вычислительных и выводных процессов. **EigenLayer**. Протокол повторного стекинга, который позволяет валидаторам Ethereum «переставлять» свои ETH для обеспечения дополнительных децентрализованных служб, в том числе для валидации ИИ.

	Также для обеспечения доверия к системам ИИ используются подходы, которые применимы к обычным информационным системам, но не ограничиваются ими. Например, для ИИ, основанного на машинном обучении, способность вызывать доверие подразумевает непредвзятость (объективность) функционирования системы.
Недостаток академических исследований, направленных на интеграцию формальных методов с ИИ	Недостаток академических исследований, направленных на интеграцию формальных методов с ИИ. Проявляется в разных областях, например в образовании и науке. Большинство исследований в этой области сконцентрировано на применении ИИ для организации образовательного процесса и обучения естественным языкам, а не на интеграции ИИ в обучение другим дисциплинам.
Недостаток академических исследований, направленных на интеграцию формальных методов с ИИ	Также в недостаточной степени рассматривается влияние ИИ на академическую успеваемость студентов разных годов и направлений обучения. **Причины.** **Ограниченность практики использования ИИ**. Технологии стали доступны широкому кругу пользователей сравнительно недавно, а их возможности ещё не до конца осмыслены преподавательским сообществом. **Нехватка методических рекомендаций**. Большинство работ в области интеграции ИИ носят гипотетический характер или базируются на локальном практическом опыте, меньше — на больших выборках и продвинутых методах анализа данных. **Этические и технические проблемы**. Внедрение ИИ в образование сопряжено с рядом юридических и этических вопросов, например, защита персональных данных, прозрачность алгоритмов оценки и принятия решений. **Неравномерный доступ**. Проблема неравномерного доступа к ИИ-технологиям среди студентов из разных регионов и с разными финансовыми возможностями.
Недостаток академических исследований, направленных на интеграцию формальных методов с ИИ	**Направления.** **Образование**. Исследования должны изучать, как ИИ может помочь в персонализированном преподавании, интеллектуальном обучении, создании динамичного образовательного контента (от интерактивных симуляций до адаптивных учебников). **Наука**. Технологии ИИ должны стать не только инструментом, но и объектом исследований, например, в области машинного обучения, автономных киберфизических систем, нейронных сетей. **Анализ данных**. Алгоритмы ИИ должны использоваться для быстрого и эффективного анализа огромных объёмов данных, что позволяет выявлять закономерности, корреляции и тенденции, которые нелегко обнаружить с помощью традиционных методов. **Рекомендации.** **Систематизировать и формализовать практику использования ИИ**. Например, проводить исследования, которые рассматривают интеграцию ИИ в обучение разным дисциплинам, а не только естественным языкам.

Недостаток академических исследований, направленных на интеграцию формальных методов с ИИ	**Разрабатывать методы для объяснения причин, лежащих в основе результатов, полученных с помощью ИИ**. Некоторые алгоритмы ИИ, такие как модели глубокого обучения, можно считать «чёрными ящиками», что затрудняет понимание и интерпретацию их процессов принятия решений. **Учитывать этические аспекты**. Исследователи должны разрабатывать методы для смягчения предвзятости моделей ИИ, обеспечения конфиденциальности данных и прозрачности алгоритмов.
Недостаточная осведомленность разработчиков о формальных методах	Проблема недостаточной осведомлённости разработчиков о важности применения формальных методов в разработке систем на основе ИИ. **Формальные методы** и методы моделирования и тестирования позволяют: **Сформулировать предположения об окружающей среде системы**, которые необходимы для её правильного функционирования. **Указать ожидаемые функциональные возможности и ограничения системы**. **Определить существенные характеристики ИИ**: корректность, надёжность, вероятность ошибок, ложноположительных результатов, чувствительность к неопределённости данных и параметров.
Недостаточная осведомленность разработчиков о формальных методах	Однако, по информации на 2021 год, применение формальных методов при разработке киберфизических систем сдерживалось из-за того, что существующие методологии и инструменты не подходили для формализации корректного поведения таких систем. Недостаточная осведомлённость разработчиков о важности формальных методов может привести к созданию нестабильных, ненадёжных и потенциально опасных систем. Это, в свою очередь, может повлечь некорректные прогнозы и рекомендации, системные сбои, утечки конфиденциальных данных и другие проблемы.
Ограничения современных алгоритмов верификации	
Нехватка автоматизированных инструментов для формальной верификации ИИ	Проблема сложности адаптации традиционных методов тестирования к моделям из-за их масштаба и отсутствия структуры. В 2019 году писали, что потребуется много работы, чтобы создать автоматизированные инструменты, которые гарантируют правильную работу систем ИИ в реальном мире. Некоторые механизмы, которые предлагают для улучшения верификации ИИ. **Аудит третьими сторонами**. Независимая проверка алгоритмов и данных помогает выявить недостатки и устранить ошибки.
Нехватка автоматизированных инструментов для формальной верификации ИИ	**Упражнения по «красной команде».** Симуляция атак на системы ИИ позволяет выявить их уязвимости и улучшить защищённость.

	Премии за выявление предвзятости и безопасности. Такие конкурсы поощряют разработчиков находить и исправлять недостатки, что ведёт к улучшению качества ИИ. **Безопасное аппаратное обеспечение**. Оно предотвращает утечки данных и обеспечивает защиту информации в системах машинного обучения.
Неэффективность SMT-решателей для больших нейросетей	SMT-решатели (Satisfiability Modulo Theories) не всегда эффективны для больших нейросетей ИИ. Это связано с сложностью задач, которые возникают при верификации нейросетей, и ограничениями SMT-решателей. **Проблема.** SMT-решатели позволяют выяснить, выполнима ли формула, и подобрать значения переменных, для которых она выполняется. Однако для больших нейросетей, например, многослойных перцептронов (MLPs), генерируемые кодировки задач верификации могут быть сложными для современных SMT-решателей. Это ограничивает возможность проверки нейросетей на практике, особенно в приложениях, связанных с безопасностью.
Неэффективность SMT-решателей для больших нейросетей	**Причины.** **Неразрешимость задач**. В общем случае задача проверки нейросетей неразрешима, и решатели не всегда могут доказать наличие или отсутствие решения, а иногда и вовсе могут зависнуть. **Сложность кодировок**. Например, для MLPs, которые представляют собой систему взаимосвязанных вычислительных единиц (нейронов), организованных в слои, генерируемые кодировки могут быть сложными для SMT-решателей. **Ограничения теорий**. SMT-решатели не всегда учитывают все теории, которые используются в задачах верификации, и не всегда могут взаимодействовать между теориями при решении ограничений. **Решения.** Использовать другие подходы. Например, комбинировать SMT-решатели с другими методами верификации нейросетей, которые позволяют сузить границы выходного множества, сохранив при этом все взаимосвязи и прослеживаемость переходов между нейронами. Например, **использовать интервальный анализ**, который оценивает входное и выходное множество для каждого нейрона, а также измеряет результаты применения функции активации.
Неэффективность SMT-решателей для больших нейросетей	**Учитывать ограничения теорий**. Например, в некоторых случаях SMT-решатели не могут учитывать ограничения, связанные с типами, которые не являются теорией решателя.
Сложность доказательства устойчивости модели к небольшим изменениям входных данных	Доказать устойчивость модели ИИ к небольшим изменениям входных данных сложно из-за сложности моделей и проблем с данными. Даже небольшие изменения в формулировке запроса или шуме могут давать разные результаты, что требует анализа чувствительности модели к вариациям. Это важно, так как устойчивость модели не гарантирует «правильность» результатов: устойчивые прогнозы могут быть ошибочными.

	Методы. Для анализа устойчивости модели к изменениям входных данных используют, например: **Адверсарное тестирование (Adversarial Testing).** Модель сознательно «обманывают», создавая специальные входные данные, чтобы исследовать, как небольшие изменения влияют на выход модели. Это помогает обнаружить слишком «резкие» переходы между классами, неожиданные паттерны в классификации и области, где модель особенно уязвима.
Сложность доказательства устойчивости модели к небольшим изменениям входных данных	**Тесты на робастность**. Проверяют устойчивость модели к шуму и небольшим изменениям во входных данных, включают имитацию дрейфа и оценку поведения вне распределения. **Визуализация пространства признаков**. Это помогает обнаружить резкие «скачки» предсказаний при минимальных изменениях входов. **Проблемы.** **Сложность моделей**. Даже простые модели могут не справиться с решением простых задач при внесении малозаметных различий. **Проблемы с данными**. Недостаточный объём обучающей выборки, неразмеченные или некорректные данные, смещение (bias), из-за которого ИИ выдаёт ошибочные прогнозы. **Переобучение** (overfitting). Если модель была обучена на «чистом» наборе и переобучена, она оказывается слишком чувствительной к малейшим отклонениям от знакомого. **Рекомендации.** **Использовать методы обучения с использованием противодействующих примеров (adversarial training)**. В модель на старте обучения добавляют данные об атаках, чтобы она могла научиться распознавать их и правильно обрабатывать такие случаи.
Сложность доказательства устойчивости модели к небольшим изменениям входных данных	**Внедрять фильтры на вход (детекция шума, length check и т. д.)**. Это поможет модели выявлять и отклонять подобные запросы до обработки данных. **Использовать синтетические данные для заполнения пробелов**, если определённые сценарии слишком редки или слишком чувствительны, чтобы их можно было зафиксировать в естественных условиях.
Отсутствие поддержки динамических изменений в моделях	Есть проблема, когда модели ИИ не могут адекватно реагировать на динамично меняющиеся условия из-за использования устаревших данных для обучения. **Некоторые последствия такого подхода**. **Финансовые убытки**. Включают затраты на переработку моделей, повторное обучение, покупку или сбор новых данных, а также потери от некорректных решений, которые принимают ИИ-системы. **Неэффективность операционной деятельности**. ИИ-система, обученная на плохих данных, может генерировать неверные прогнозы, оптимизации или автоматизированные действия, что приводит к сбоям, ошибкам и снижению общей производительности.

Отсутствие поддержки динамических изменений в моделях	**Потеря доверия и репутационный ущерб**. Неэффективные или ошибочные ИИ-решения подрывают доверие клиентов, партнёров и инвесторов. **Потеря конкурентного преимущества**. В то время как конкуренты внедряют работающие ИИ-решения, компания, застрявшая в бесконечном цикле отладки данных, теряет свои позиции на рынке. Для решения этой проблемы предлагается, например, **технология Model Context Protocol (MCP)**. Она устраняет «изоляцию» интеллектуальных моделей от данных, стандартизирует обмен информацией между ИИ-ассистентами и системами, где находятся данные. Также рассматривается **архитектура «живого преобразователя»** — гибкой, стохастически модулируемой системы, способной к самонастройке, рефлексии и субъективному восприятию информации.
Проблемы с кодированием нейросетей в логические ограничения	Существует проблема неспособности моделей ИИ адекватно реагировать на динамично меняющиеся условия. **Некоторые причины возникновения этой проблемы**.
Проблемы с кодированием нейросетей в логические ограничения	**Устаревшие данные**. Они не позволяют модели реагировать на изменения, что делает её бесполезной в реальном времени. **Сложность моделей**. Даже небольшие изменения в формулировке запроса могут дать очень разные результаты. **Недостаточное развитие науки об обучении моделей**. Адаптация ИИ к динамичным средам сопряжена со значительными трудностями. Некоторые из них. **Скорость обработки**. Модели должны реагировать в режиме реального времени, не теряя точности своих анализов. **Способность к обобщению**. Системы должны экстраполировать решения на новые ситуации, а не полагаться исключительно на конкретные знания. **Этические критерии и принятие решений**. Алгоритмы должны принимать ответственные решения в непредвиденных ситуациях, избегая нежелательных или пагубных последствий. **Уменьшение смещения**. ИИ должен гарантировать, что данные, используемые для обучения, не вносят предвзятости или систематические ошибки.
Проблемы с кодированием нейросетей в логические ограничения	Без возможности адаптироваться к изменениям системы ИИ быстро устареют в непредсказуемых условиях.
Ограниченные методы поиска контрпримеров	Существуют ограничения методов создания контрпримеров ИИ. По результатам одного из исследований, лучшие модели способны создавать контрпримеры только для менее 9% некорректных решений. При этом способность ИИ к фальсификации отстаёт от его способности генерировать решения. Также есть информация о сложности автоматической проверки контрпримеров для некорректных решений. Однако есть и позитивные новости: например, исследователи из Калифорнийского технологического института разработали модель ИИ, которая ищет неожиданные, сложные пути.

	Для этого использовался **метод обучения с подкреплением**: ИИ начинали с простых задач и постепенно их усложняли. В результате алгоритмы новой модели опровергли ряд потенциальных контрпримеров, которые оставались нерешёнными 25 лет.
Невозможность проверять модели с плавающей запятой с высокой точностью	Невозможность проверять модели с плавающей запятой с высокой точностью в ИИ связана с тем, что числа с плавающей запятой являются приблизительными. Ошибки округления и другие математические неточности могут привести к полной потере значимости выводов нейросети, а также сделать бессмысленными любые сравнения метрик качества. Некоторые причины проблем с точностью. **Уменьшение размера числа**. Например, для оптимизации скорости используются 16-битные числа с плавающей запятой (FP16). Каждое уменьшение размера числа приводит к снижению точности, что уменьшает значимость результата. **Неточность операций**. Все операции с числами имеют некоторые неточности, но самая большая из них возникает в случае вычитания чисел, близких друг к другу. Сравнение значений также может вносить ошибки, поскольку реализуется через вычитание. **Накопление цифровых шумов**. Единичные ошибки обычно не оказывают заметного влияния на работу нейросети, а вот их накопление в конечном итоге приводит к потере значимости результата.
Невозможность проверять модели с плавающей запятой с высокой точностью	Для решения этих проблем разрабатываются новые методы машинного обучения, например COLLAGE, который использует многокомпонентное представление с плавающей запятой (MCF) для точной обработки операций с числовыми ошибками.
Проблема экспоненциального роста вычислительных требований	Проблема экспоненциального роста вычислительных требований в ИИ заключается в том, что количество вычислений, необходимых для обучения современных моделей ИИ, растёт экспоненциально. То есть, чтобы сделать модель вдвое «умнее», может потребоваться не в два, а в десять или сто раз больше вычислений. Это приводит к резкому росту энергопотребления и стоимости, например: **Обучение одной большой модели может потреблять огромное количество энергии**. **Запуск моделей в производстве (вывод) также влечёт за собой постоянные затраты энергии**, часто превышающие затраты на фазу обучения с течением времени. **Причины.** **Увеличение сложности моделей**. Современные модели, особенно те, которые содержат миллиарды или даже триллионы параметров, требуют огромных объёмов оперативной памяти и вычислительной мощности.

Проблема экспоненциального роста вычислительных требований	**Ограничения классических вычислений**. Классические компьютеры работают с битами, которые могут быть только 0 или 1, и каждая операция выполняется последовательно или параллельно, но всегда в конечном числе состояний. Чтобы найти оптимальный выход (обучить модель), классическому компьютеру приходится пробовать один путь за другим, даже если это происходит очень быстро. **Решения.** **Поиск новых вычислительных парадигм и специализированных архитектур**. Например, разработка специализированных чипов для ИИ, которые работают быстрее и потребляют меньше энергии, чем GPU. **Оптимизация хранения и извлечения знаний в моделях ИИ**. Например, подход «масштабируемые слои памяти (SML)» — вместо того, чтобы встраивать всю изученную информацию в параметры с фиксированным весом, вводят внешнюю систему памяти, извлекая информацию только при необходимости. Это снижает вычислительные издержки и улучшает масштабируемость.
Проблема экспоненциального роста вычислительных требований	**Использование квантовых вычислений**. Квантовые компьютеры работают по иным принципам, чем классические, и некоторые алгоритмы могут быть выполнены на них с гораздо меньшими энергозатратами.
Плохая адаптируемость к задачам с неполной информацией	Плохая адаптируемость к задачам с неполной информацией в ИИ может быть связана с несколькими факторами. **Качество данных**. ИИ-модели обучаются на огромных наборах данных, но они могут быть неполными, неправильными или предвзятыми. Если модель обучалась на данных, которые не охватывают все возможные ситуации или содержат ошибки, её предсказания также будут ошибочными. **Сложность поставленной задачи**. Некоторые задачи требуют понимания контекста, ассоциаций и опыта, которых у модели нет. ИИ часто «думает» только в пределах данных, которые получил, и не всегда способен обобщать и адаптироваться к новым условиям. **Технические ограничения архитектуры**. Даже самые передовые архитектуры ИИ, такие как глубокие нейронные сети, имеют технические ограничения. У модели есть определённое количество слоёв, нейронов и весов, которые ограничивают её способность к анализу и синтезу информации.
Плохая адаптируемость к задачам с неполной информацией	**Ограниченность в проверке данных**. Модели ИИ, как правило, не умеют проверять правдивость информации и полагаются на вероятностные вычисления. Это означает, что иногда они «угадывают» ответ, что особенно рискованно при ответах на вопросы, связанных с фактами. Чтобы минимизировать риски получения неточной информации от ИИ, можно следовать ряду рекомендаций. **Задавать более точные и детализированные вопросы**. Это поможет модели лучше понять запрос.

	Проверять полученные данные, используя несколько источников информации, особенно если ответ кажется сомнительным. **Учитывать контекст**, предоставляя модели дополнительную информацию или уточнения. **Разбивать сложные задачи**. Большую задачу нужно разделить на несколько подзадач и решать их последовательно, опираясь на предыдущие результаты.
Проблемы с архитектурой нейросетей	
Глубокие нейросети имеют слишком сложную зависимость между слоями	Глубокие нейросети (глубокое обучение, deep learning) имеют сложную зависимость между слоями, что связано с их многослойной архитектурой. Это позволяет моделям выявлять сложные закономерности в данных, но при этом возникают проблемы, связанные с переобучением и сложностью интерпретации решений. **Принцип работы.** Глубокие нейросети состоят из множества слоёв, каждый из которых обрабатывает данные на разных уровнях абстракции. Некоторые особенности зависимости. **Входной слой принимает исходные данные** — например, изображение или текст. **Скрытые слои производят вычисления на основе входящих параметров**, настраивая свои «решения» через веса — параметры, которые определяют силу связи между нейронами. **Выходной слой выводит результат вычислений**. Чем глубже слой, тем более сложные и информативные признаки извлекает нейросеть. Например, в свёрточных нейросетях для обработки изображений первый слой анализирует «примитивные» детали (грани, края),
Глубокие нейросети имеют слишком сложную зависимость между слоями	второй — геометрические конструкции (прямые, кружки), третий — части объектов (глаза, колёса), а последний — целые объекты (лицо, машину). **Проблемы.** **Переобучение (Overfitting)** — модель чрезмерно адаптируется к обучающим данным, «запоминает» детали, которые не являются важными для общего паттерна, и теряет способность обобщать информацию. **Сложность интерпретации решений** — из-за сложности глубоких нейронных сетей часто трудно понять, как они принимают решения. **Проблема исчезающего градиента** — при обратном распределении ошибки градиент становится слишком малым, чтобы веса обновлялись эффективно. **Неэффективное обновление** весов снижает скорость и качество обучения нейросети. **Решения.** Чтобы избежать проблем, связанных со сложной зависимостью между слоями, важно: **Тщательно выбирать архитектуру сети** — количество слоёв и их характеристики в зависимости от задачи. **Использовать методы регуляризации** — например, Dropout и L2-регуляризацию, которые помогают снизить вероятность переобучения.

Глубокие нейросети имеют слишком сложную зависимость между слоями	**Использовать функции активации** — они добавляют системе нелинейность, что позволяет ей учиться более эффективно и запоминать сложные зависимости данных.
Отсутствие верификационных методов для гибридных систем	Имеется в виду проблема «чёрного ящика» в гибридных системах ИИ. Это отсутствие прозрачности в работе алгоритмов, когда пользователь не понимает, каким образом ИИ пришёл к тому или иному выводу или решению. Такая непрозрачность затрудняет верификацию результатов. Однако есть способы, которые помогают решить эту проблему и обеспечить верификацию гибридных систем ИИ. **Участие человеческих экспертов в разметке данных, генерации информативных признаков и верификации моделей**. Эксперты направляют обучение алгоритмов в сторону более осмысленных и корректных решений. **Использование статистических методов для начальной проверки модели**. Это позволяет сравнить результаты и оценить эффективность адаптированных параметров, полученных с помощью ИИ.
Отсутствие верификационных методов для гибридных систем	**Применение модульной архитектуры при проектировании гибридных систем**. Компоненты разделяют на слои логических правил и машинного обучения, обеспечивают строгий интерфейс обмена данными между ними. **Использование правил как регуляризаторов в обучении**. В функцию потерь нейросети добавляют штрафы за отклонение от экспертных знаний. **Динамическое взвешивание**. Реализуют адаптивный баланс между правилами и ML-моделью через механизм внимания. Также важно учитывать, что качество исходных данных напрямую определяет качество результатов, получаемых от интеллектуальных систем. Недостаточная верификация, наличие смещений или пропусков в данных могут привести к неверным выводам и ошибочным решениям.
Разнообразие топологий нейросетей	**Некоторые виды топологий нейросетей в ИИ**. **Полносвязные**. Каждый нейрон передаёт свой выходной сигнал остальным нейронам, в том числе и самому себе. Все входные сигналы подаются всем нейронам. **Многослойные** (слоистые). Нейроны объединяются в слои. Слой содержит совокупность нейронов с едиными входными сигналами. Число нейронов в слое может быть любым и не зависит от количества нейронов в других слоях.
Разнообразие топологий нейросетей	**Монотонные.** Это частный случай слоистых сетей с дополнительными условиями на связи и нейроны. Каждый слой, кроме последнего (выходного), разбит на два блока: возбуждающий и тормозящий. **Сети без обратных связей**. В таких сетях нейроны входного слоя получают входные сигналы, преобразуют их и передают нейронам первого скрытого слоя, и так далее вплоть до выходного, который выдает сигналы для интерпретатора и пользователя.

	Топологические. Специализированные архитектуры, предназначенные для работы с данными, структурированными в топологических областях. В отличие от традиционных нейронных сетей, адаптированных для сетчатых структур, топологические сети умеют обрабатывать более сложные представления данных. **Сверточные**. Предназначены для обработки двумерных структур данных, прежде всего изображений. **Рекуррентные.** Сети с обратными связями. Также нейронные сети различают по структуре сети (связей между нейронами), особенностям модели нейрона, особенностям обучения сети.
Разнообразие топологий нейросетей	Некоторые отличия (проверены в 2024 г.) сетей Колмогорова-Арнольда (KAN) от перечисленных выше нейронных сетей: **Функциональная декомпозиция**. KAN явно используют теорему Колмогорова-Арнольда для разложения многомерной функции на одномерные функции, в то время как традиционные нейронные сети не имеют такой явной декомпозиции. **Одномерные преобразования**. KAN в значительной степени полагаются на одномерные преобразования, тогда как традиционные нейронные сети используют многослойные персептроны с нелинейными функциями активации. **Связность уровней**. В традиционных нейронных сетях каждый нейрон в одном слое связан с каждым нейроном в последующем слое (полностью связан). В KAN связи структурированы таким образом, чтобы отражать сумму одномерных функций. **Интерпретируемость**. Структура KAN, основанная на одномерной функциональной декомпозиции, иногда может обеспечить большую интерпретируемость по сравнению с природой «чёрного ящика» традиционных нейронных сетей.
Разнообразие топологий нейросетей	**Оптимизированная масштабируемость**. KAN демонстрирует превосходную масштабируемость по сравнению с традиционными нейронными сетями, особенно в сценариях с высокоразмерными данными. **Повышенная точность**. Несмотря на использование меньшего количества параметров, KAN достигает более высокой точности и меньших потерь, чем традиционные нейронные сети, при решении различных задач. Применение сетей KAN приближает к созданию сильного искусственного интеллекта.
Формальные методы плохо адаптируются к сетям с изменяющейся структурой	Формальные методы в ИИ плохо адаптируются к сетям с изменяющейся структурой из-за ограничений, связанных с формализацией задач и отсутствием гибкости в работе с динамическими данными и знаниями. Это характерно для традиционных систем ИИ, которые часто не способны решать плохо формализованные задачи, требующие построения оригинального алгоритма решения в зависимости от конкретной ситуации.

	Причины. **Ограничения формализации**. Формальные методы (например, логические системы) не учитывают динамичность исходных данных и знаний, не позволяют автоматически строить алгоритмы решения в условиях неопределённости.
Формальные методы плохо адаптируются к сетям с изменяющейся структурой	**Жёсткая запрограммированность**. Варианты поведения системы часто запрограммированы в программе, что ограничивает гибкость. Например, в алгоритмах с большим числом конструкций (if, case) варианты поведения жёстко запрограммированы, и при настройке на иные сценарии работы нужно менять исходный код программ. **Сложность автоматического построения представлений**. Если в задачах приобретения знаний представления знаний заданы априори, то в задачах метаобучения ставится вопрос об автоматическом построении самих представлений, детали которых могут сильно меняться в зависимости от предметной области. **Решения.** Для преодоления ограничений формальных методов в ИИ используются, например: **Системы, основанные на моделях.** В таких системах программы и структуры данных генерируются или компонуются из единиц знаний, описанных в репозитарии, каждый раз при изменении модели проблемной области.
Формальные методы плохо адаптируются к сетям с изменяющейся структурой	**Подходы, основанные на машинном обучении**. Система обучается, а не программируется явно: ей передаются многочисленные примеры, имеющие отношение к решаемой задаче, а она находит в этих примерах статистическую структуру, которая позволяет системе выработать правила для автоматического решения задачи. **Использование неклассической логики**. Классическая логика (логика высказываний и логика предикатов) не подходит для решения целого ряда задач, что приводит к созданию других логических систем, которые учитывают динамичность.
Проблемы с проверкой работы слоев нормализации и активации	Некоторые проблемы, которые могут возникать при проверке работы слоёв нормализации и активации в ИИ: **Проблемы с нормализацией**. **Раздельная нормализация обучающих и тестовых данных**. Нормализация предполагает определение новых единиц измерения для проблемных переменных. Нужно выразить каждую запись в одних и тех же единицах измерения, независимо от того, принадлежит ли она к обучающему или тестовому набору.
Проблемы с проверкой работы слоев нормализации и активации	**Проблема внутреннего ковариантного сдвига**. Распределение входов слоя изменяется по мере обновления предыдущих слоёв, что значительно замедляет обучение. Нормализация — одно из перспективных направлений решения этой проблемы. **Увеличение вычислительных затрат**. Для небольших и мелких нейронных моделей с небольшим количеством слоёв

	нормализации это незначительно, но становится серьёзной проблемой, когда базовые сети становятся больше и глубже. **Проблемы с активацией**. **Неправильное применение функции активации**. Функции активации всегда должны следовать линейному преобразованию, чтобы сеть могла эффективно обучаться и представлять сложные шаблоны. **Потеря преимуществ линейного преобразования**. Линейное преобразование предназначено для отображения входных данных в новом пространстве, где функция активации может эффективно вводить нелинейность. Применяя сначала функцию активации, теряют преимущества этого сопоставления, уменьшая способность сети изучать сложные шаблоны.
Проблемы с проверкой работы слоев нормализации и активации	**Вопросы размерности и инициализации**. Функции активации ожидают, что входные данные будут иметь определённую форму, обычно это результат линейного преобразования. Их применение непосредственно к необработанным входным данным может привести к проблемам с размерностью и неправильной инициализации, что снижает производительность сети. **Нарушение градиентного потока**. Во время обратного распространения градиенты должны проходить как через линейное преобразование, так и через функцию активации. Применение функции активации сначала нарушает этот поток, что потенциально приводит к исчезновению или взрыву градиентов, что затрудняет процесс обучения.
Трудности с верификацией сетей, работающих в реальном времени	Некоторые трудности с верификацией сетей, работающих в реальном времени на основе искусственного интеллекта. **Задержка**. Даже при параллельном выполнении проверок синхронизация результатов и достижение согласия вызывает задержки. Процесс проходит только так быстро, как самый медленный узел. **Сложность инженерии**. Оркестровка оценок по нескольким моделям и обеспечение плавной работы механизма консенсуса требуют значительных инженерных усилий.
Трудности с верификацией сетей, работающих в реальном времени	**Требования к вычислениям**. Даже при использовании более маленьких моделей их одновременное выполнение в ансамблях увеличивает вычислительные требования. **Неоднозначные случаи**. В таких случаях ансамбли могут испытывать трудности с выравниванием, что приводит к несогласованным результатам. **Масштаб и отсутствие структуры моделей**. Адаптировать традиционные методы тестирования для моделей, содержащих сотни миллионов параметров, сложно. **Потенциальное мошенничество**. Без надёжной верификации злоумышленники могут представлять неверные или изменённые данные.

	Зависимость от централизованных слабых мест. Зависимость от оффчейн оракулов или частных серверов может подорвать децентрализованную этику, привести к цензуре или единичным точкам отказа.
Неясность, как корректно формализовать понятие «хорошей генерации» для генеративных моделей	Понятие «хорошей генерации» для генеративных моделей в ИИ может быть формализовано как реалистичность и разнообразие выходных данных. Это важно, так как генеративные модели создают новый контент (изображения, тексты, музыку) на основе изученных закономерностей, и качество генерации зависит от качества обучения и разнообразия обучающих данных. **Критерии оценки.** **Реалистичность** — насколько хорошо модель имитирует реальные данные. Например, сгенерированные изображения должны быть похожи на реальные. **Разнообразие** — насколько широк спектр возможных выходов. Идеальная модель должна уметь генерировать и котиков, и собачек, и пейзажи — в зависимости от задачи. **Согласованност**ь с исходными данными — при повторном запуске модели с теми же входными данными результаты должны быть похожими или одинаковыми. **Методы.** Для объективного анализа качества генерации используются специальные метрики, которые измеряют разные аспекты. Например:
Неясность, как корректно формализовать понятие «хорошей генерации» для генеративных моделей	**FID (Fréchet Inception Distance)** — сравнивает распределение сгенерированных изображений с распределением реальных (тестовых) данных. Чем ниже значение FID, тем лучше модель. **Inception Score (IS)** — оценивает два аспекта: качество — насколько изображения узнаваемы (высокая уверенность классификатора), разнообразие — насколько разные классы представлены в генерации. **LPIPS (Learned Perceptual Image Patch Similarity)** — оценивает perceptual similarity (похожесть для человеческого глаза). Важно: ни одна метрика не идеальна, лучше использовать несколько и смотреть на их комбинацию. **Проблемы.** **Субъективность оценки** — восприятие качества генерации может отличаться у разных людей. Учёт различных мнений и перспектив помогает снизить субъективность, но полностью избежать её может быть сложно. **Ограниченность метрик** — они могут не охватывать весь спектр допустимых вариантов ответа модели.
Рекуррентные сети (LSTM, GRU) имеют сложные временные зависимости	LSTM (Long Short-Term Memory) и GRU (Gated Recurrent Unit) — типы рекуррентных нейронных сетей (RNN), разработанные для обработки сложных временных зависимостей в искусственном интеллекте (ИИ). Эти архитектуры решают

	проблему исчезающего градиента, которая возникает в традиционных RNN и препятствует их способности улавливать долгосрочные зависимости в последовательных данных. **LSTM (Long Short-Term Memory)** — специализированный тип архитектуры RNN, разработанный для преодоления ограничений традиционных RNN в обучении зависимостям на дальних расстояниях. **Особенности обработки временных зависимостей**. Использует ячейки памяти, которые сохраняют информацию в течение длительного времени. Использует три «ворота» для регулирования информации, хранящейся в ячейке памяти: **Ворота забывания** — решают, какую информацию из состояния ячейки следует выбросить. **Входные ворота** — решают, какую новую информацию сохранить в состоянии ячейки. **Выходной гейт** — решает, какую часть состояния ячейки выводить.
Рекуррентные сети (LSTM, GRU) имеют сложные временные зависимости	Эти ворота учат, какую информацию важно сохранить или отбросить на каждом временном шаге, позволяя сети сохранять релевантный контекст в течение длительного времени. **GRU (Gated Recurrent Unit)** — упрощённая версия LSTM, разработанная для упрощения и ускорения обучения по сравнению с LSTM. Особенности обработки временных зависимостей: **Ворота обновления** — определяют, какое количество прошлой информации (предыдущее скрытое состояние) должно быть перенесено в будущее состояние. **Ворота сброса** — решают, сколько прошлой информации нужно забыть, прежде чем вычислять новое скрытое состояние кандидата. Эти ворота совместно управляют памятью сети, позволяя ей узнавать, какую информацию следует сохранять или отбрасывать в длинных последовательностях. Важно: хотя GRU также хорошо справляются с долгосрочными зависимостями, они могут быть не столь эффективны, как LSTM, при понимании более долгосрочных зависимостей из-за своей более простой структуры.
Верификация многомодальных моделей	Верификация многомодальных моделей в ИИ — это процесс оценки корректности и надёжности модели, которая обрабатывает и интегрирует несколько типов данных (текст, изображения, аудио, видео и данные сенсоров) одновременно. Цель — определить способность модели эффективно обобщать и давать правильные результаты на новых примерах. Верификация также позволяет выявить потенциальные проблемы, такие как переобучение, смещение или недостаточную производительность в реальных условиях. **Методы.** Для верификации многомодальных моделей используют, например:

	Контрастное обучение — модель учится сближать представления связанных пар модальностей (например, изображение и его описание) и отдалять представления несвязанных. **Механизмы внимания** — выявляют наиболее релевантные части одной модальности, соответствующие определённым элементам другой. **Стратегии слияния** — объединяют данные из разных модальностей в единое представление. Например:
Верификация многомодальных моделей	**Раннее слияние** — объединяет извлечённые векторы признаков сразу после обработки каждой модальности. **Позднее слияние** — сохраняет разделение модальностей до финальных этапов принятия решений, когда прогнозы от каждой модальности объединяются. **Гибридное слияние** — объединяет признаки несколько раз на разных уровнях модели, используя механизмы совместного внимания для динамического выделения и согласования важных кросс-модальных взаимодействий. **Данные.** Для верификации многомодальных моделей используют разнообразные наборы данных, которые охватывают несколько модальностей. Некоторые требования: **Репрезентативность** — набор должен содержать разнообразные сценарии и условия, чтобы верифицировать работу модели на различных ситуациях. **Объективность и независимость от обучающего набора данных** — помогают предотвратить переобучение модели и дают более объективную оценку производительности.
Верификация многомодальных моделей	**Критические случаи и редкие сценарии** — верификационный набор данных должен включать их, чтобы проверить, как модель справляется с экстремальными условиями. **Инструменты.** Для верификации многомодальных моделей используют, например: **Специализированные инструменты сбора данных** — они одновременно фиксируют несколько модальностей. **Протоколы выравнивания на основе временных меток** — помогают обеспечить согласование между различными модальностями. **Оценка.** Результаты верификации многомодальных моделей оценивают по различным метрикам, например: **Точность, полнота**, F1-мера — оценивают производительность модели в разных аспектах. **Семантическая близость между модальностями** — например, для оценки качества описаний к видео используют косинусную близость эмбеддингов.
Ограничения в математических основах	
Отсутствие общепринятой формальной логики для описания поведения ИИ	Общепринятой формальной логики для описания поведения ИИ нет. Это связано с особенностями работы современных систем ИИ, которые не всегда подчиняются законам формальной логики. Например:

	Генеративные модели ИИ (машинное обучение) не способны к эксплицитным рассуждениям. Система нащупывает закономерности в исходных данных и руководствуется ими в своих действиях, но не имеет средств выразить их в доступной человеческому восприятию форме. **Продукты машинного обучения** (например, чат-боты) не имеют заметного отношения к логике и зависят от сочетания учебных корпусов и совокупного опыта обучения. **Причины.** **Трудности с формализацией неформальных знаний**. Логический подход сталкивается с трудностями, когда возникает необходимость описания знаний, которые плохо формализуются. **Неполнота понимания природы человеческого мышления**. Логический подход не может полностью описать процесс мышления и принятия решений, например, такие феномены, как озарение, интуитивный поиск решения или эмоциональные влияния на принятие решений.
Отсутствие общепринятой формальной логики для описания поведения ИИ	**Методы.** Для описания поведения ИИ в отсутствие общепринятой формальной логики используются неформальные модели. Например: **Семантические сети** — описывают набор сущностей и связей между ними, изображаются в виде графа. **Сценарии** — моделируют возможность человека группировать знания о какой-либо деятельности или типичной ситуации в одну целостную структуру. **Фреймовые технологии** — позволяют извлекать закономерности из избыточной информации и создавать логические цепочки. **Исследования.** В 2024 году исследователи корпорации Apple провели исследование, которое показало, что ИИ-модели не способны к подлинному логическому мышлению. Например, чат-боты сбиваются с толку, если добавлять в простые задачи лишние или несуществующие данные. Это объясняется тем, что ИИ-модели следуют уже имеющимся у них шаблонам и логическим связям, которые записаны в их данных, вместо того чтобы проводить анализ новых условий задачи и адаптировать свои выводы.
Трудности с построением исчерпывающих спецификаций для ИИ	Построение исчерпывающих спецификаций для систем ИИ сталкивается с рядом трудностей, которые связаны с техническими ограничениями, сложностью предметной области и нереалистичностью ожиданий. Эти трудности проявляются в разных аспектах: в работе с данными, выборе и настройке алгоритмов, в обеспечении надёжности и устойчивости систем. **Причины.** **Работа с данными**. Объём, качество, репрезентативность и чистота данных определяют потенциал и ограничения ИИ-системы. Предвзятость, неполнота или шум в данных могут

	привести к систематическим ошибкам, которые сложно выявить и устранить на поздних этапах разработки. **Выбор, обучение и валидация моделей**. Существует множество алгоритмов и архитектур, каждый из которых обладает своими преимуществами и ограничениями. Выбор оптимальной модели, её тонкая настройка (гиперпараметры), обеспечение способности к обобщению на новые данные — всё это требует глубоких теоретических знаний и практического опыта.
Трудности с построением исчерпывающих спецификаций для ИИ	**Нереалистичность ожиданий**. Разработчики и заказчики устанавливают недостижимые цели, не учитывая текущие ограничения алгоритмов, доступность данных или сложность предметной области. **Игнорирование реальных условий** эксплуатации. Модель, прекрасно работающая в контролируемой лабораторной среде, может демонстрировать низкие показатели при столкновении с вариативностью и непредсказуемостью реального мира. **Методы.** Для преодоления трудностей в построении спецификаций для ИИ используются, например: **Модульный подход к проектированию**. Позволяет разбивать процессы на более управляемые компоненты, что упрощает разработку и отладку. **Использование готовых библиотек и фреймворков**. Они могут ускорить процесс разработки и предложить проверенные алгоритмы. **Итеративный подход к разработке**. Постоянное тестирование и готовность к адаптации планов позволяют минимизировать риски несоответствия.
Трудности с построением исчерпывающих спецификаций для ИИ	**Примеры.** **Несоответствие между идеальной спецификацией и выявленной**. Например, система ИИ не делает то, что от неё хотят, из-за того, что системы ИИ — не идеальные оптимизаторы или из-за непредвиденных последствий использования проектной спецификации. **Проблема «чёрного ящика».** Многие современные модели, такие как глубокие нейронные сети, работают как «чёрные ящики», где трудно понять, как именно принимаются решения. **Литература.** Проблема построения исчерпывающих спецификаций для ИИ рассматривается в научной статье AI Safety Gridworlds. В ней представлены разные типы спецификаций и проблемы, связанные с ними, например, несоответствие между идеальной и проектной спецификацией.
Неопределенность в трактовке вероятностных моделей	Неопределённость в трактовке вероятностных моделей для ИИ возникает из-за ограниченности информации, сложности прогнозирования будущих событий, а также воздействия случайных факторов. Есть два основных источника неопределённости:

Неопределенность в трактовке вероятностных моделей	**Неопределённость данных (алеаторическая)**. Измеряет «сложность» задачи и может быть вызвана шумной зависимостью таргета от факторов или пересекающимися классами. Эта неопределённость не может быть уменьшена путём сбора большего количества обучающих данных. **Неопределённость знаний (эпистемическая)**. Возникает, когда модель получает входные данные из области, которая либо слабо охвачена обучающими данными, либо далека от них. Поскольку модель мало знает об этой области, она, скорее всего, допустит ошибку. В отличие от неопределённости данных, неопределённость знаний может быть уменьшена путём сбора большего количества обучающих данных из плохо изученных областей. Вероятностные рассуждения обеспечивают математическую основу для представления неопределённости и управления ею. Используя вероятности, системы ИИ могут принимать обоснованные решения даже в условиях неопределённости. Например, в экономике такие модели могут быть использованы для прогнозирования рыночных тенденций, оценки финансовых рисков и оптимизации цепочек поставок, где важна высокая степень адаптации к неопределённости.
Отсутствие единой модели математического представления различных архитектур ИИ	Единой модели математического представления различных архитектур ИИ нет. Исследователи в этой области используют различные подходы, которые не всегда унифицированы. Например, в ИИ нет единой трактовки понятия «искусственный интеллект» (Artificial Intelligence, AI) в научной среде. Представители разных наук привносят в изучение ИИ специфическое, связанное с их первоначальными интересами. **Причины.** **Многообразие интеллектуальных задач и методов**. Методы ИИ разнообразны, они адаптируются под решаемую задачу. **Сложности формализации знаний**. Для различных предметных областей и задач более удобными и эффективными в вычислительном отношении оказываются различные формы представления знаний. **Трудности с построением глобальных истинных моделей**. Попытки создания ИИ с поддержкой глобальных истинных моделей (большая база непротиворечивых аксиом) наталкивались на значительные трудности. **Подходы.** Некоторые подходы к математическому представлению архитектур ИИ:
Отсутствие единой модели математического представления различных архитектур ИИ	**Логический** — базируется на моделировании рассуждений и символьном представлении информации. Теоретическая основа — логика, правила вывода. **Эволюционный** — основное внимание уделяется построению начальной модели и правилам, по которым она может изменяться (эволюционировать). Модель может быть составлена по различным методам, например, по мотивам человеческого мозга.

	Имитационный — объект, поведение которого имитируется, представляет собой «чёрный ящик»: информация о внутренней структуре и содержании отсутствуют, но известны спецификации входных и выходных данных. **Исследования.** Однако есть и тенденции к объединению различных подходов. Например: **Разработка интегрированных и гибридных систем ИИ** — они объединяют преимущества разнородных моделей, например нечётких экспертных систем и нейронных сетей. В таких системах могут поддерживаться различные модели представления знаний, разные типы рассуждений, модели восприятия и распознавания образов.
Отсутствие единой модели математического представления различных архитектур ИИ	**Исследование интерпретируемости и объяснимости моделей** — разработка методов, которые позволят понять, как модель принимает решения и какие факторы влияют на её выводы.
Формальная верификация плохо адаптирована к динамическим системам	Формальная верификация может быть сложно адаптирована к динамическим системам ИИ. Это связано с тем, что модели ИИ могут содержать сотни миллионов параметров и не иметь чёткой структуры. Кроме того, обычная логика высказываний плохо подходит для формулировки утверждений о поведении сложных динамических систем при изменении их состояний во времени. Однако есть исследования, которые пытаются адаптировать формальную верификацию к верификации динамических систем ИИ. Например, в одной из работ авторы предлагают алгоритмы, которые подходят для верификации динамических систем, в том числе, когда архитектуры ИИ используются в качестве контроллеров с обратной связью по состоянию.
Сложность формальной верификации решений, основанных на эвристиках	Формальная верификация решений, основанных на эвристиках в ИИ, сталкивается с сложностями из-за особенностей неформальных рассуждений и специфики задач, для которых используются эвристические методы. Это требует разработки новых подходов, которые объединяют неформальное и формальное мышление, и учёта специфики предметной области. **Причины.** Неформальные рассуждения часто включают сокращения пути и пропущенные шаги, которые формальные системы не могут верифицировать. **Специфический характер задач** — эвристические функции часто адаптируются к конкретным сценариям, что ограничивает их обобщение. Эвристика, которая хорошо работает для одного сценария, может быть неприменима в другом контексте. **Вычислительные издержки** — вычисление сложной эвристики на каждом шаге может привести к дополнительным вычислительным затратам. Если стоимость вычисления эвристики перевешивает её преимущества, это может не повысить общую производительность.

Сложность формальной верификации решений, основанных на эвристиках	**Риск неоптимальных решений** — недопустимые эвристические методы, хотя и более быстрые, могут привести к неоптимальным решениям. **Методы.** Для преодоления сложностей используются, например: **Разбиение сложных проблем на более мелкие, управляемые части** — вместо того чтобы пытаться решить всю проблему сразу, система разбивает её на серию «подцелей» — промежуточных лемм, которые служат ступеньками к окончательному доказательству. **Использование «награды за согласованность»** — это снижает структурное несоответствие и обеспечивает включение всех декомпозированных лемм в окончательные доказательства. **Применение экспертных знаний** — «подсказки» эксперта предметной области используются для направленного поиска, построения абстракций и накладывания ограничений на пространство поиска.
Сложность формальной верификации решений, основанных на эвристиках	**Примеры.** Формальная верификация решений, основанных на эвристиках, применяется в задачах, требующих больших пространств поиска, где эвристические методы направляют поиск по более перспективным путям. Например: **Поиск пути** — эвристический поиск помогает найти кратчайший или наиболее эффективный путь между двумя точками. **Оптимизация** — эвристические методы помогают максимально использовать доступные ресурсы при максимизации эффективности. **Планирование и распределение ресурсов** — эвристические функции направляют поиск решений, удовлетворяющих всем ограничениям. **Исследования.** В области сложности формальной верификации решений, основанных на эвристиках, проводятся, например: **Исследование модели DeepSeek-Prover-V2** — модель объединяет неформальное и формальное мышление, разбивает сложные проблемы на управляемые части, сохраняя при этом точность, необходимую для формальной верификации.
Сложность формальной верификации решений, основанных на эвристиках	Исследование DeepSeek должно указывать, что была использована теория «Управления распределенными базы данных» В.М. Глушкова (СССР 1964 г.). Эта теория определяет распределенное множество ИНС и, как следствие, сложность существенно уменьшается и исчезает проблема «черного ящика».
Формальные методы не учитывают обучение с подкреплением в реальном мире	Имеются сложности, с которыми сталкивается обучение с подкреплением (Reinforcement Learning) при применении в реальном мире. Некоторые из них: **Зависимость от симуляций**. Реальные среды часто слишком дороги или опасны для непосредственного обучения. При этом перенос знаний из виртуальной среды в реальный мир

	остаётся серьёзной проблемой. Даже современные методы обучения не гарантируют успешный перенос в 100% случаев. **Сложность дизайна вознаграждения**. Трудно создать системы вознаграждения, которые бы сочетали немедленные действия с долгосрочными целями. **Высокие вычислительные требования**. Алгоритмы RL требуют большой вычислительной мощности, особенно при использовании в крупномасштабных или сложных ситуациях.
Формальные методы не учитывают обучение с подкреплением в реальном мире	**Эффективность выборки**. Для эффективной работы обучения с подкреплением часто требуется много данных, что является большой проблемой в таких областях, как робототехника или здравоохранение, где сбор данных может быть дорогим или рискованным. Несмотря на эти сложности, метод Reinforcement Learning применим на практике, однако он требует взвешенного подхода к использованию. Современные исследования сосредоточены на создании более эффективных, стабильных и безопасных алгоритмов, а также специализированных инструментов для их промышленного внедрения.
Верификация больших языковых моделей требует новых логических формализмов	Верификация больших языковых моделей (LLM) в ИИ требует новых логических формализмов из-за проблем, связанных с неопределённостью в работе моделей и ошибками в логических рассуждениях. Эти вызовы требуют разработки методов, которые учитывают вероятностный вывод LLM и ошибки, а также использования новых логических формализмов для анализа неопределённости. **Методы верификации.** Для верификации LLM используются, например:
Верификация больших языковых моделей требует новых логических формализмов	**Бенчмарки** — задачи, требующие логического рассуждения и анализа длинного контекста. Например, MuSR — алгоритмически сгенерированные сложные задачи, требующие логического рассуждения. **Запрос к самой модели** — пользователи задают модели дополнительные вопросы для проверки её ответов и получения информации о процессе генерации ответа. Модель предоставляет объяснения или дополнительные данные, которые помогают понять процесс генерации ответа и выявить возможные ошибки. **Методика prompt engineering** — продуманная формулировка запросов (prompts), которые управляют поведением модели для получения необходимых результатов. Это позволяет оптимизировать взаимодействие с моделью, фокусируясь на конкретных задачах и снижая вероятность генерации нерелевантной или недостоверной информации.
Верификация больших языковых моделей требует новых логических формализмов	**Логические формализмы.** Для верификации LLM используются, например: **Вероятностные контекстно-свободные грамматики (PCFG)** — расширение обычных контекстно-свободных

	грамматик, в которых правилам вывода приписываются вероятности. Используются для моделирования неопределённости в формальных спецификациях, сгенерированных LLM. **Метрики, основанные на анализе грамматик** — например, энтропия грамматики, которая показывает, насколько неопределённым является выбор правил вывода, или перплексия, оценивающая, как грамматика предсказывает дальнейшие шаги формализации. Эти метрики помогают обнаруживать ошибки и неопределённости, позволяя проводить выборочную формальную верификацию только тех выводов, которые содержат признаки повышенного риска ошибок. **Исследования**. Некоторые исследования в области верификации LLM с использованием новых логических формализмов: Исследование группы учёных из Amazon и Калифорнийского университета в Лос-Анджелесе (2024).
Верификация больших языковых моделей требует новых логических формализмов	Учёные разработали модель SolverLearner, которая использует двухэтапный подход: отделяет процесс изучения правил рассуждений от процесса их применения к конкретным случаям. Это помогает чётко отличить индуктивные рассуждения от дедуктивных, так как LLM хорошо справляются с индуктивными рассуждениями, но часто не способны к дедуктивным. Исследование исследователей из Case Western Reserve University и Microsoft (2025). Предлагает подход, который меняет представление о неопределённости в работе LLM: вместо того чтобы считать её критическим недостатком, её следует использовать для лучшего понимания и контроля над процессом формализации с помощью LLM.
Проблемы, связанные с интерпретируемостью ИИ	
Формальная верификация не улучшает интерпретируемость моделей	Формальная верификация не улучшает интерпретируемость моделей в ИИ, но формальная верификация важна для обеспечения корректности алгоритмов, но интерпретируемость моделей — задача, которая требует улучшения. Это связано с разными понятиями: формальной верификацией и интерпретируемостью. **Формальная верификация.** Формальная верификация — это доказательство корректности алгоритмов на основе математических методов. Она позволяет:
Формальная верификация не улучшает интерпретируемость моделей	− доказать, что алгоритм соответствует заданным свойствам (например, устойчивости к малым изменениям входных данных, ограничениям по безопасности); − предоставить математические гарантии верной работы модели для любого входа из пространства возможных входных данных. Однако формальная верификация не улучшает интерпретируемость, так как её применимость ограничена масштабом и сложностью системы, а также её стохастической природой.

	Интерпретируемость. Интерпретируемость — это способность понимать и объяснять, как модель машинного обучения или глубокого обучения делает прогнозы или принимает решения. Некоторые особенности интерпретируемости: **Простота** — в традиционных моделях машинного обучения (деревья решений, линейная регрессия) понимание поведения модели относительно простое из-за их прозрачности. **Сложность** — модели глубокого обучения, особенно нейронные сети, работают как сложные многослойные «чёрные ящики», что затрудняет интерпретируемость.
Формальная верификация не улучшает интерпретируемость моделей	**Субъективность** — то, что составляет «интерпретируемую» модель, может варьироваться в зависимости от приложения и аудитории. Модель, которая интерпретируема специалистом по обработке данных, может быть не интерпретируема экспертом в предметной области в другой области. **Взаимосвязь.** Формальная верификация не решает проблему интерпретируемости, так как её цель — обеспечить корректность алгоритмов, а интерпретируемость — обеспечить прозрачность процесса вывода. Например: Формальная верификация не помогает понять, почему модель приняла конкретное решение — она не объясняет логику, приведшую к этому результату. Сложность интерпретации возрастает с увеличением сложности моделей — даже при наличии высокоточного предсказания или обнаружения аномалии понимание логики, приведшей к этому результату, остаётся неочевидным. **Решения.** Для улучшения интерпретируемости моделей в ИИ необходимо:
Формальная верификация не улучшает интерпретируемость моделей	**Разрабатывать модели, которые поддаются интерпретации по своей природе** — например, неглубокие модели (деревья принятия решений, линейные модели). **Использовать методы для улучшения интерпретируемости** — например, разделение процесса принятия решения на отдельные интерпретируемые блоки (для многослойной нейросети — интерпретация работы отдельных нейронов). **Учитывать контекст** — не следует гнаться за максимальной интерпретируемостью результатов без учёта контекста. Например, при низкой стоимости ошибки в предсказании (например, рекомендация фильма пользователю) не стоит прилагать усилия для того, чтобы сделать модель более интерпретируемой.
Трудности с интерпретацией формальных доказательств корректности	В ИИ возникают трудности с интерпретацией формальных доказательств корректности из-за разрыва между неформальными и формальными рассуждениями. Например, большие языковые модели (LLM) испытывают трудности с преобразованием интуитивных рассуждений в формальные доказательства, которые могут проверить машины.

Трудности с интерпретацией формальных доказательств корректности	**Причины.** Неформальные рассуждения часто содержат сокращения и неявные шаги, которые формальные системы не могут проверить. Формальное доказательство теорем требует полной точности: каждый шаг явно указан и логически обоснован без двусмысленности. Преобразование между естественным языком и формальной записью усложняет задачу. **Методы.** Для решения проблемы интерпретации формальных доказательств корректности в ИИ используются, например: **Подход, объединяющий неформальные и формальные рассуждения**. Например, в модели DeepSeek-Prover-V2 система разбивает сложные проблемы на более мелкие, управляемые части, сохраняя при этом точность, требуемую формальной проверкой. **Синтез данных обучения**. Когда все подцели сложной проблемы успешно решены, система объединяет эти решения в полное формальное доказательство.
Трудности с интерпретацией формальных доказательств корректности	**Обучение с подкреплением**. Модель получает обратную связь о том, верны ли её решения или нет, и использует эту обратную связь, чтобы узнать, какие подходы работают лучше всего. **Исследования.** В мае 2025 года исследователи DeepSeek-AI представили DeepSeek-Prover-V2 — модель ИИ с открытым исходным кодом, способную преобразовывать математическую интуицию в строгие, проверяемые доказательства. В работе описан подход, который облегчает преодоление разрыва между человеческой интуицией и проверенными машиной доказательствами. Исследование DeepSeek должно указывать, что была использована теория «Управления распределенными базы данных» В.М. Глушкова (СССР 1964 г.), эта теория определяет распределенное множество ИНС и, как следствие, сложность существенно уменьшается и исчезает проблема «черного ящика».
Формальные методы не могут объяснить, как модель принимает решения	Формальные методы не всегда могут объяснить, как модель принимает решения (**принимать решение может только человек возможностями своего головного мозга, который алгоритмические задачи не решает**) в ИИ из-за сложности понимания внутренней работы алгоритмов и задач, для которых используются эти методы. Однако существует направление исследований — объяснимый искусственный интеллект (Explainable AI, XAI), которое стремится создать системы и модели, способные объяснять свои действия и принимать решения понятным для людей образом. **Формальные методы.** Формальные методы (например, логическое программирование) эффективны, когда входные данные предсказуемы, а поведение можно выразить в виде жёсткой логики. Однако этот

	подход плохо масштабируется в условиях неопределённости, неполноты и нечёткости знаний. Например, невозможно описать правилами, как отличить Луну от круглой лампочки, как читать неразборчивый почерк.
Формальные методы не могут объяснить, как модель принимает решения	**Проблемы.** Отсутствие прозрачности в моделях «чёрного ящика» (системы, которые принимают решения на основе входных данных и алгоритмов) вызывает проблемы: **Кризис доверия** — люди не могут понять закономерности и процессы принятия решений в этих моделях. **Опасения по поводу возможной дискриминации, предвзятости** — отсутствие транспарентности приводит к возникновению этих факторов, заложенных в архитектуру модели. **Сдерживание совершенствования и оптимизации** — отсутствие ясного понимания внутренней работы модели сдерживает её последующее совершенствование. **Методы.** XAI использует специальные методы, которые позволяют отслеживать и объяснять каждое решение, принятое в процессе машинного обучения. Некоторые из них: **Анализ интерпретируемости модели** — изучение внутренней работы модели, чтобы понять, как она приходит к своим решениям.
Формальные методы не могут объяснить, как модель принимает решения	Например, анализ важности признаков, графики частичной зависимости. **Объяснения контрфактуальными данными** — сравнение фактического результата с тем, что произошло бы, если бы определённые условия были другими. **Локальные объяснения** — сегментируют пространство решений и дают объяснения менее сложным подпространствам решений, которые актуальны для всей модели. **Объяснения на примерах** — извлечение репрезентативных примеров, которые улавливают внутренние отношения и корреляции, обнаруживаемые анализируемой моделью данных. **Нормативные требования.** В некоторых юрисдикциях, например в Европейском Союзе с введением Общего регламента по защите данных (GDPR), требуется, чтобы организации (**т. е. люди**) могли объяснить решения, принятые с использованием автоматизированных систем.
Проблема определения «разумных границ» допустимых отклонений	Есть проблема выбора диапазона значений для гиперпараметров при настройке моделей искусственного интеллекта. Неправильно определённые «разумные границы» могут привести к неэффективному поиску и плохой производительности модели. Некоторые аспекты, на которые нужно обращать внимание при определении границ: **Контекст задачи и алгоритма**. Нужно убедиться, что заданные минимальное и максимальное значения имеют смысл в рамках конкретной задачи.

	Специфика алгоритма. Разные алгоритмы имеют разные требования к гиперпараметрам. **Влияние гиперпараметров на производительность модели**. Перед настройкой рекомендуется оценить их влияние на производительность модели с помощью базовых экспериментов. **Вычислительные ресурсы**. Настройка гиперпараметров может быть ресурсоёмким процессом, поэтому при планировании экспериментов нужно учитывать доступные вычислительные ресурсы.
Проблема определения «разумных границ» допустимых отклонений	Также есть мнение, что при работе с ИИ важно соблюдать «разумные границы» и использовать искусственный интеллект только в тех ситуациях, когда это действительно необходимо.
Формальная верификация не дает гарантии, что модель «понимает» проблему	Формальная верификация в ИИ не гарантирует, что модель «понимает» проблему, но может помочь выявить и устранить потенциальные уязвимости и ошибки в поведении системы. Однако этот метод имеет ограничения, и для проверки моделей ИИ существуют альтернативные подходы. **Формальная верификация.** Формальная верификация — это математическое доказательство соответствия программы или системы заданным спецификациям. В случае ИИ она означает доказательство того, что поведение сети соответствует определённым ожиданиям и безопасным границам. Например, верификация помогает: − выявить ошибки в поведении нейронных сетей; − оценить точность и надёжность систем ИИ в критических приложениях (автономное вождение, медицинская диагностика). Входными данными формальной верификации являются модель и её спецификации, описывающие свойства, выполнение которых нужно проверить.
Формальная верификация не дает гарантии, что модель «понимает» проблему	**Ограничения.** Верифицируется только модель, а не сама система. Тот факт, что модель обладает определёнными свойствами, не гарантирует, что финальная реализация будет обладать теми же свойствами. Для проверки финальных реализаций требуются дополнительные методы, такие как систематическое тестирование. **Сложности с масштабируемостью**. Формальная верификация требует исчерпывающего анализа всех возможных состояний системы, что с увеличением размера и сложности становится более сложным. **Невозможность проверять обобщения**. Например, если протокол верифицирован для одного, двух и трёх процессов с использованием формальной верификации, это не даёт результата для другого числа процессов — метод практичен только для частных случаев. **Неопределённость внешних факторов**. Взаимодействие с внешними системами (например, с данными от внешних источников) может вносить дополнительную неопределённость в модель верификации, так как их поведение часто не может быть точно предсказано.

Формальная верификация не дает гарантии, что модель «понимает» проблему	**Альтернативы.** Для проверки моделей ИИ существуют, например: **Тестирование** — оценка сети на большом наборе точек во входном пространстве и определение, соответствуют ли её выходы желаемым на этом наборе. Однако тестирование не может служить гарантированным обоснованием того, что тестируемая модель обладает проверяемыми свойствами. **Анализ границ принятия решений** (Decision Boundary Analysis) — исследование, как небольшие изменения входных данных влияют на выход модели. Это помогает обнаружить слишком «резкие» переходы между классами, неожиданные паттерны в классификации и области, где модель особенно уязвима.
Формальная верификация не решает проблему доверия к модели	Формальная верификация не даёт гарантии, что модель «понимает» проблему в ИИ, так как этот метод ориентирован на доказательство корректности модели, а не на проверку её понимания проблемы. Это связано с особенностями формальной верификации и её ограничениями.
Формальная верификация не решает проблему доверия к модели	**Формальная верификация.** Цель формальной верификации — доказать, что модель корректна относительно заранее определённых требований. Для этого используют математическую модель, которая отражает возможное поведение системы, и формальную спецификацию требований, описывающую желаемое поведение. На базе этих спецификаций формальным доказательством проверяют, действительно ли возможное поведение согласуется с желаемым. **Важно:** верифицируется только модель, а не сама система. Тот факт, что модель обладает определёнными свойствами, не гарантирует, что финальная реализация будет обладать теми же свойствами. Для проверки финальных реализаций требуются дополнительные методы, такие как систематическое тестирование. **Ограничения.** Проблема масштабируемости. Формальная верификация требует исчерпывающего анализа всех возможных состояний модели, что с увеличением размера и сложности системы становится более сложным.
Формальная верификация не решает проблему доверия к модели	**Сложности с моделированием внешних факторов**. Например, асинхронные операции и взаимодействие с другими системами через вызовы — из-за сложности взаимодействий невозможно точно смоделировать все возможные сценарии. **Невозможность проверять обобщения**. Если модель верифицирована для одного, двух и трёх процессов, это не даёт результата для другого числа процессов — проверка моделей практична только для частных случаев. **Неопределённость внешних факторов**. Использование данных от внешних источников может вносить дополнительную неопределённость в модель верификации, так как их поведение часто не может быть точно предсказано.

	Альтернативы. Для проверки моделей ИИ, помимо формальной верификации, используют, например: **Систематический поиск наихудших результатов работы** — помогает найти случаи, в которых модель может отойти от желаемых свойств. **Анализ границ принятия решений** (Decision Boundary Analysis) — исследует, как небольшие изменения входных данных влияют на выход модели, что помогает обнаружить слишком «резкие» переходы между классами и неожиданные паттерны в классификации.
Формальные методы не гарантируют объяснимость решений ИИ в суде или регуляторных органах	Существуют проблемы, связанные с прозрачностью и объяснимостью решений, которые принимают системы ИИ в суде или регуляторных органах. Некоторые из них: **Отсутствие технико-юридического контроля**. Например, ИИ может использовать для принятия решений информацию из источников, которые не проверены на соответствие законодательству. **Невозможность пересмотреть решение**. Технологически ИИ может выдать одно решение, и его нельзя будет изменить. **Неучитывание уникальных особенностей дела**. В общем сформированном судебном решении могут быть не учтены особенности отдельного уголовного дела. **Необъективные решения**. Самообучающиеся алгоритмы могут обучаться на определённых наборах данных, которые содержат необъективные сведения. Чтобы решить эти проблемы, предлагают, например, добавлять в систему ИИ нейронные сети, которые будут формировать текстовые объяснения на основе данных, полученных от сетей, принимающих решения.
Формальные методы не гарантируют объяснимость решений ИИ в суде или регуляторных органах	Также можно выдавать среднестатистические данные по всем решениям, чтобы человек мог сравнить отклонения в индивидуальном решении ИИ для него по сравнению с усреднёнными значениями. Кроме того, важно разработать этические стандарты и контрольные механизмы, чтобы убедиться в том, что ИИ используется ответственно и в соответствии с правовыми и этическими нормами.

2.6. Проблема «черного ящика»

Объяснимый искусственный интеллект (explainable artificial intelligent – XAI) является актуальной потребностью, поскольку все больше и больше моделей ИИ используется для подготовки альтернатив принятия решений. Таким образом, эти решения также воздействуют на многих пользователей. При этом каждый пользователь может получить благоприятное или неблагоприятное воздействие.

На данный момент проблема заключается в том, что специалист по анализу данных, построивший модель, не имеет полной ясности о поведении модели, и не хватает ясности в ее объяснении. В дальнейшем, завершив обучение нейронной сети и проверив ее работоспособность на тестовых данных, по сути, получается «черный ящик». Фактически неизвестно, как вырабатывается ответ, но все же он вырабатывается!

Проблема «чёрного ящика» [36, 85, 86] в ИИ представляет собой одну из наиболее актуальных и обсуждаемых тем в области машинного обучения и анализа данных. Суть данной проблемы заключается в том, что многие современные модели ИИ, особенно те, которые основаны на глубоких нейронных сетях, обладают высокой сложностью и непрозрачностью. Это приводит к тому, что пользователи и разработчики не могут в полной мере понять, как именно принимаются решения, что, в свою очередь, вызывает опасения по поводу их надежности и этичности. В условиях, когда ИИ все чаще используется в критически важных сферах, таких как медицина, финансы, право и безопасность, необходимость в объяснимости и интерпретируемости моделей становится не просто желательной, а жизненно важной.

В 2022–2023 годах в мире произошел новый скачок в развитии технологий ИИ, благодаря совершенствованию больших генеративных моделей в области языка, изображений (включая видеоизображения) и звука.

Большие фундаментальные модели уже сейчас способны писать программные коды по техническим заданиям, сочинять поэмы на заданную тему, давать точные и понятные ответы на тестовые вопросы различных уровней сложности, в том числе из образовательных программ.

Модели искусственного интеллекта за секунды создают изображения на любую тему по заданному текстовому описанию или наброску, что создает угрозу распространения запрещенной информации, нарушения авторских прав и генерации ошибочных сведений.

Большие языковые модели, которые используются для генерации текста, становятся все более похожи на человеческий язык. И поэтому многие люди, и здесь даже среди специалистов нет консенсуса, и многие специалисты говорят, что видят в этих системах уже проблески человекоподобного интеллекта, а некоторые считают и сознание.

В настоящее время стали проблемой появившиеся мультимодальные большие языковые модели, которые добавляют к языку обработку изображений и звука, а иногда и управление физическим или виртуальным телом.

Поскольку человеческое сознание по своей природе мультимодально (смешанное, комбинированное) и связано с активными действиями, эти расширенные системы больших языковых моделей более перспективны в качестве кандидатов на возможное человеческое сознание.

Следует понимать, что повсеместное использование ИИ (GPT 4,5) провоцирует возможность манипулирования с его помощью человеческим сознанием путем изготовления различных дипфейков и интернет-влияния на психику человека. В табл. 14. Приведены характеристики нейронных сетей ИИ.

Таблица 14. Характеристики нейронных сете типа GPT.

Название	**Характеристики**
GPT-3	17,5 миллиардов параметров
GPT- 4	1,76 триллиона параметров
WuDao 2.0	1,76 триллиона параметров
GPT-5	17,5 триллиона параметров
DeepSeek V3	671 млрд параметров
Claude Haiku 3.	20 млрд параметров
YandexGPT 5 Lite.	8 миллиардов параметров

Именной эта характеристика (параметры) является основной в проблеме нейронных сетей с названием «черный ящик». Кто может убедить пользователя, что при таком (триллионном) количестве параметров, обученная ИНС работает правильно? Как убедить, что при триллионном количестве параметров, веса входов нейронов (из них собрана ИНС) являются оптимальными?

Актуальность «черного ящика» ИНС, обусловлена растущим внедрением ИИ в повседневную практику и необходимость обеспечения доверия к этим технологиям.

Например, в медицине алгоритмы ИИ могут принимать решения о диагнозах и назначении лечения, что требует от них не только высокой точности, но и возможности объяснить, на каких основаниях было принято то или иное решение.

В финансовом секторе алгоритмы могут определять кредитоспособность клиентов, и отсутствие прозрачности в таких процессах может привести к дискриминации и юридическим последствиям.

Таким образом, проблема «чёрного ящика» затрагивает не только технические аспекты, но и этические, правовые и социальные.

Проблема «чёрного ящика» в искусственном интеллекте возникает в результате сложных и зачастую непрозрачных механизмов, лежащих в основе работы современных моделей, таких как глубокие нейронные сети. Эти системы принимают решения на основе анализа входных данных, однако сам процесс принятия решений оказывается скрыт от разработчиков и пользователей, что делает результаты нечитаемыми и непредсказуемыми, а также создает потенциальные риски предвзятости и дискриминации [87, 88].

Актуальность проблемы возрастает с каждым годом, поскольку ИИ все активнее проникает в различные сферы – от медицины и финансов до правосудия и систем безопасности. Как показывают исследования, многие алгоритмы не могут объяснить, как они пришли к тем или иным выводам, что вызывает обоснованные опасения относительно их использования в критически важных ситуациях [87]. Кроме того, даже создатели тех систем зачастую не понимают, каким образом они функционируют, что усугубляет этические и правовые вопросы при их использовании [88].

Ключевыми аспектами, касающимися объяснимости ИИ, являются интерпретируемость и прозрачность. Интерпретируемость подразумевает, что человек

должен иметь возможность понять и объяснить, как и почему модель приняла то или иное решение. Прозрачность, в свою очередь, требует, чтобы информация о функционировании алгоритма была доступна и понятна пользователю. Существующие подходы к объяснению моделей ИИ зачастую далеки от удовлетворительных решений, особенно в контексте сложных и многослойных архитектур, характерных для глубокого обучения [89].

Исторически концепция черного ящика начала развиваться с ростом популярности алгоритмов машинного обучения. В своих ранних разработках исследователи стремились к созданию легко интерпретируемых моделей, но с увеличением сложности алгоритмов необходимость в объяснимости стала критической. В настоящее время многие ученые и практики ищут пути решения этой проблемы, разрабатывая различные методологии и подходы, направленные на создание более объяснимых ИИ [85, 86].

Наличие чёрного ящика в системах ИИ создает вызовы, которые требуют не только технических решений, но также глубокого понимания границ знаний и методов, используемых при их разработке. Этические и практические аспекты использования таких систем должны стать предметом серьезного обсуждения и регуляции, так как от этого зависит доверие пользователей и, в конечном счете, успешность внедрения технологий ИИ в общество.

Современные исследования в области объяснимости стремятся найти баланс между сложностью моделей и потребностью в понятности их работы. Важно отмечать, что ни один из подходов не является универсальным, и выбор метода объяснения зависит от специфики задачи и контекста, в котором используется модель. Если в прошлом акцент ставился на собственные объясняющие возможности моделей, то теперь следует учитывать и потребности пользователей, что становится новым вызовом.

Таким образом, понимание исторических основ этой проблемы и ее эволюции позволяет увидеть, как менялись подходы и как росла необходимость в объясняющих инструментах. Это понимание подготовит почву для дальнейшего анализа современных методов и вызовов, которые стоят перед исследователями

и практиками в области достижения более высокой степени прозрачности и объяснимости в ИИ.

Современные подходы к объяснению моделей искусственного интеллекта сосредоточены на повышении степени понимания и доверия к алгоритмическим решениям, используемым в различных областях. Основные направления данного исследовательского процесса включают как наглядные, так и более абстрактные методы анализа.

Надзорное обучение, основанное на заранее размеченных данных, позволяет моделям учиться на конкретных примерах. К числу методов в этой категории относятся линейная и логистическая регрессии, решающие деревья, случайный лес и бустинг. Эти техники предоставляют возможность оценивать важность различных признаков, что упрощает интерпретацию результатов. Например, в линейной регрессии влияние каждой переменной можно проиллюстрировать с помощью коэффициента, а в решающих деревьях визуализация древообразной структуры помогает понять, какие факторы приняты во внимание при классификации [90].

К методам, не требующим надзора, относятся, в частности, k-ближайших соседей и метод опорных векторов. Эти определения могут быть менее прозрачными, однако объяснимость может быть достигнута через сравнительный анализ объектов, основанный на близости в пространстве признаков. В результате, понятие объяснимости для данных подходов может зависеть от условий, при которых модель была создана и использована [36].

Объяснимость моделей ИИ представляет собой сложную задачу, к решению которой специалисты в области машинного обучения сталкиваются с множеством трудностей и барьеров. Одна из основных проблем – это феномен, известный как "shortcut learning". Этот эффект выражается в том, что модели иногда основывают свои выводы на неверных признаках, таких как фон изображений, игнорируя более значимые характеристики объектов [85]. Это приводит к ситуациям, когда, несмотря на обходительные результаты на тестовых наборах, системы могут выдавать неверные результаты в реальных условиях.

Существуют и другие проблемы. Например, утечки данных при обучении, избыточность или недостаток информации также сильно влияют на устойчивость и надежность моделей. Разработка более детализированных и объяснимых систем

требует значительных усилий и понимания сложных архитектур машинного обучения. Исследования показывают, что около 87% проектов в данной области сталкиваются с неудачами, зачастую из-за недооценки важности объяснимости моделей [85]. Эта ситуация требует дополнительных внимания к анализу таких провалов, чтобы оптимизировать подходы к созданию более надежных моделей.

Вопросы объяснимости становятся всё более актуальными, и привести к эффективному решению проблемы без интеграции мнений экспертов в эту область невозможно.

Эксперты в области объяснимого искусственного интеллекта (XAI) высказывают разнообразные мнения о текущих вызовах и перспективах внедрения решений, повышающих объяснимость моделей ИИ. Одной из основных проблем становится недостаточная прозрачность алгоритмов, используемых в критически важных сферах, таких как медицина, финансы и правосудие. Механизмы «чёрного ящика» затрудняют анализ решений ИИ, что может вызывать недоверие со стороны пользователей и регуляторов [86,90].

Дискуссии среди исследователей также подчеркивают необходимость создания единообразной терминологии и разработку практических рекомендаций по XAI. Как отмечает один из экспертов, отсутствие четких стандартов в использовании объяснимых моделей ограничивает их внедрение и применение на практике [91]. Тем не менее, существует общее согласие относительно того, что повышение прозрачности в принятии решений критически важно для обеспечения доверия к ИИ, особенно в случаях, когда алгоритмы влияют на жизни людей [92].

Ключевым аспектом обсуждения является возможность пользователей понимать и интерпретировать действия ИИ. Некоторые эксперты выделяют интерпретируемость как важный элемент XAI, позволяющий пользователям осознанно взаимодействовать с технологиями. Такой подход предполагает, что объяснения должны быть понятны не только разработчикам, но и конечным пользователям [93]. В этом контексте наблюдается растущее внимание к формированию моделей, которые могут не только выдавать результаты, но и объяснять, как они были достигнуты.

Обсуждение XAI также касается уровней объяснимости, которые могут быть достигнуты с помощью разработки специализированных методов. Одни эксперты указывают на необходимость интеграции моделей с возможностью идентификации и исправления ошибок, что, в свою очередь, повышает уровень доверия к системам ИИ. Однако существует и другая точка зрения. Некоторые специалисты говорят о недостаточной информативности даже объяснимых моделей, что требует дополнительного внимания со стороны исследователей и практиков.

Таким образом, несмотря на многообещающие направления исследований в области XAI, остаются значительные препятствия, связанные с практическим применением объясняющих систем. Это подчеркивает важность continuing collaboration между учеными, разработчиками и пользователями для формирования единого подхода к повышению объяснимости и интерпретируемости технологий ИИ.

Повышение прозрачности алгоритмов ИИ требует комплексного подхода, включающего как технологические, так и этические аспекты. Важно, чтобы специалисты по обработке данных осознавали значимость объяснимости моделей, ведь только открытое взаимодействие пользователей с алгоритмами может способствовать доверию к результатам.

Первым шагом к прозрачности является использование композитного ИИ. Это подход интегрирует разнообразные аналитические методы и технологии, что не только улучшает интерпретируемость, но и позволяет уменьшить предвзятость на выходе. Кроме того, рекомендуется применять модели, которые легко объяснить пользователям, такие как линейные модели и деревья решений. Использование модели, понятной для конечного пользователя, уменьшает риск недоверия к системам ИИ и помогает более эффективно оценивать последствия их решений.

Следует использовать принципы справедливости и защиты конфиденциальности в модели ИИ.

Правовая основа также играет критическую роль в обеспечении прозрачности. Существующие законодательства в сфере технологий должны обновляться с учётом появления новых алгоритмических решений и их этических следствий.

Наконец, важно, чтобы все заинтересованные стороны, включая разработчиков, законодателей и пользователей, вели открытый диалог о проблемах и решениях, связанных с прозрачностью ИИ. Создание совместных инициатив по обмену знаниями приведет к более интуитивным и понятным системам. В конечном счете, разработка эффективных механизмов для повышения объяснимости и справедливости систем ИИ будет способствовать не только улучшению качества принимаемых решений, но и укреплению доверия общества к технологиям, изменяющим нашу жизнь [85].

Исторический контекст развития ИИ показывает, что проблема объяснимости не нова. С момента появления первых алгоритмов машинного обучения исследователи и практики сталкивались с необходимостью объяснять результаты работы своих моделей. Однако с развитием более сложных и мощных алгоритмов, таких как глубокие нейронные сети, эта проблема стала более остро ощущаться.

Из проведенного анализа следует, что проблема «чёрного ящика» в ИИ требует комплексного подхода, включающего как технические, так и образовательные меры. Только совместными усилиями можно преодолеть существующие трудности и создать более прозрачные и объяснимые модели, которые будут служить на благо общества. Важно помнить, что объяснимый ИИ — это не просто тренд, а необходимость, которая будет определять будущее технологий и их взаимодействие с человеком.

ГЛАВА 3. СЕТИ КОЛМОГОРОВА: НАЧАЛО НОВОЙ ЭРЫ В НЕЙРОННЫХ МОДЕЛЯХ

Передовые методики машинного и глубокого обучения, включая нейронные сети, стремительно эволюционируют, открывая все более широчайшие перспективы решения комплексных задач в самых разных индустриальных и научных сферах.

Одним из наиболее перспективных векторных направлений данного сектора считаются сети Колмогорова–Арнольда (KAN) [94,95] — инновационная нейросетевая архитектура, использующая принципы теоремы Колмогорова–Арнольда о универсальной аппроксимации.

В отличие от классических многослойных персептронов, такие архитектуры внедряют обучаемые активаторные функции, размещённые прямо на рёбрах графа, благодаря чему устраняются жёстко заданные нелинейности и зависимость от линейных весов.

Вместо традиционных методов в KAN задействуют сплайны при построении одномерных функций, что заметно усиливает гибкость модели и адаптивность всей нейронной архитектуры.

На рисунке 5 представлены теоремы Колмогорова [95].

Академик А. Н. КОЛМОГОРОВ

О ПРЕДСТАВЛЕНИИ НЕПРЕРЫВНЫХ ФУНКЦИЙ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ В ВИДЕ СУПЕРПОЗИЦИЙ НЕПРЕРЫВНЫХ ФУНКЦИЙ ОДНОГО ПЕРЕМЕННОГО И СЛОЖЕНИЯ

Целью заметки является краткое изложение доказательства следующей теоремы:

Теорема. *При любом целом $n \geqslant 2$ существуют такие определенные на единичном отрезке $E^1 = [0; 1]$ непрерывные действительные функции $\psi^{pq}(x)$, что каждая определенная на n-мерном единичном кубе E^n непрерывная действительная функция $f(x_1, \ldots, x_n)$ представима в виде*

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{q=2n+1} \chi_q \left[\sum_{p=1}^{n} \psi^{pq}(x_p) \right], \qquad (1)$$

где функции $\chi_q(y)$ действительны и непрерывны.

При $n = 3$, положив

$$\varphi_q(x_1, x_2) = \psi^{1q}(x_1) + \psi^{2q}(x_2), \quad h_q(y, x_3) = \chi_q[y + \psi^{3q}(x_3)],$$

получаем из (1)

$$f(x_1, x_2, x_3) = \sum_{q=1}^{7} h_q[\varphi_q(x_1, x_2,), x_3], \qquad (2)$$

Рис. 5. Теорема Колмогорова

Современные архитектуры Колмогорова–Арнольда (KAN), названные в честь математиков Андрея Николаевича Колмогорова и его ученика Владимира Игоревича Арнольда, опираются на принципы знаменитой теоремы 1957 года. Согласно данному результату, произвольное отображение можно разложить в конечную сумму непрерывных функций одной переменной, что в рамках глубинного обучения открывает путь к более пластичной репрезентации данных. В KAN слоях акценты смещены: обучаемые нелинейности располагаются на рёбрах графа, тогда как вершины выполняют лишь линейное смешивание сигналов. Такая топология, контрастирующая с фиксированными активациями классических MLP, повышает экспресс-способность и адаптивность моделей при работе с высокоразмерными и нестационарными входами.

Одним из фундаментальных достоинств KAN считается отказ от фиксированных линейных весов и неизменных активаций в пользу богатого семейства сплайновых базисов. Такой переход раскрывает простор для конструирования нестандартных топологий и тонкой настройки сети под конкретные исследовательские задачи. Благодаря этому возрастает прозрачность вычисленного решения, что крайне важно для научных целей, например, при извлечении скрытых

физических закономерностей из экспериментальных наборов [95]. Дополнительная простота и адаптивность KAN позволяют точнее моделировать сложные явления, тогда как классические многослойные перцептроны (MLP) нередко уступают в подобных сценариях.

Ряд недавних исследований демонстрирует, что сети типа KAN эффективно справляются с высокоразмерными задачами. Их вычислимая через внутреннюю математическую топологию энтропия формирует инструментарий для детектирования скрытых закономерностей в масштабных датасетах, тем самым открывая перспективные направления использования нейросетевых моделей в фундаментальной и прикладной науке [96]. При планировании экспериментов необходимо учитывать, что функциональный потенциал KAN существенно превосходит классические архитектуры.

Взаимодействие теоретических основ и прикладных возможностей KAN проявляется, помимо прочего, в их умении обходить ограничения, характерные для классических архитектур нейросетей. Благодаря компактной архитектуре, повышенной интерпретируемости и улучшенной обобщающей способности, а также явной регуляризации исследователи получают более детальное представление о динамике обучения, что критически важно при извлечении скрытых паттернов из высокоразмерных сложных датасетов [97].

Тем самым архитектура KAN представляет собой интеграцию фундаментальных концепций нейрокомпьютинга с актуальными запросами отрасли, формируя качественно новый виток эволюции нейросетевых систем, способный радикально трансформировать методики решения исследовательских и инженерных задач. Подобные топологические новшества повышают не только метрическую точность, но и прозрачность, облегчая верификацию, дебаг и семантическую интерпретацию моделей.

Архитектура KAN представляет собой эволюционное поколение нейронных сетей, радикально отличающееся от классических топологий. Основа

KAN — адаптивные функции активации, которые замещают фиксированные нелинейности, традиционно используемые в многослойных перцептронах (MLP). Эти параметризуемые сплайны обладают высокой гибкостью, локальным управлением формы и способны моделировать сложные нелинейные зависимости, обеспечивая гладкость и непрерывность при обработке данных [97].

В архитектуре KAN узлы (вершины графа) агрегируют входные сигналы, суммируя их и пропуская через параметризуемые, обучаемые функции активации. Ребра выполняют роль коммуникационных каналов, по которым эти агрегированные значения передаются, формируя цепочки композиционных отображений и тем самым расширяя функциональное пространство сети [98]. В отличие от традиционного MLP, где нелинейности (ReLU, tanh и др.) задаются заранее и остаются статичными, в KAN сами функции активации оптимизируются в процессе обучения [99]. Такая адаптивность повышает экспрессивную способность модели и позволяет точнее аппроксимировать сложные распределения и латентные ореолы данных.

Сплайны, используемые в архитектуре KAN, представляют собой класс математических функций, позволяющих сшивать дискретные контрольные узлы гладкими кривыми. Благодаря этой способности они незаменимы при задачах точной интерполяции и высокоточной аппроксимации сложных зависимостей [100]. Свойственная им непрерывность и адаптивная гибкость дают возможность тонко настраивать форму сегментов, сохраняя достоверное отражение данных на любых масштабах сложности. Управление каждым сегментом сплайна обеспечивает требуемую гладкость C^1 или C^2, что принципиально влияет на эффективность процесса обучения модели [101].

Фундаментальной особенностью KAN выступает интеграция обучаемых функций активации, размещённых прямо на рёбрах графа. Подобная конструкция открывает возможность гибко выбирать класс нелинейностей — от B-сплайнов и рациональных кривых Безье до других кусочно-полиномиальных аппроксиматоров, что существенно повышает точность моделирования и интерполяции

сложных распределений [98, 99]. Дополнительное преимущество заключается в том, что параметризуемые активации упрощают адаптивное масштабирование KAN, позволяя обрабатывать высокоразмерные выборки быстрее и с меньшими вычислительными затратами по сравнению с архитектурами на фиксированных активационных ядрах.

Рассматриваемая архитектура предоставляет KAN способность оперативно подстраиваться под разнообразные рабочие сценарии, тем самым делая её оптимальным выбором для широкого круга прикладных задач. Эластичность сплайновых ядер позволяет достигать выдающейся точности при значительно меньшем числе обучаемых параметров, что одновременно сокращает расходы на тренинг модели и снижает вычислительную нагрузку по сравнению с классическими нейросетевыми архитектурами [97]. Сочетание адаптивности, параметрической экономичности и масштабируемости превращает KAN в перспективное направление для дальнейших исследований и разработок в AI и глубоком обучении, открывая возможности интеграции инновационных алгоритмов в разнообразные отраслевые приложения.

Архитектура KAN-сетей открывает перед научным сообществом новые горизонты благодаря своей своеобразной структурной организации и оригинальным принципам функционирования. Ключевое достоинство этих моделей заключается в том, что они изначально формируют ответы в виде явных математических выражений; следовательно, их внутренние механизмы остаются прозрачными и легко поддаются строгой интерпретации, в отличие от типовых глубоких сетей, например классических многослойных персептронов. Подобная формульная репрезентация позволяет не только получать численные прогнозы, но и обосновывать их через проверяемые аналитические выкладки, что имеет решающее значение для фундаментальных исследований [102].

Более того, концептуальный фундамент KAN, опирающийся на теорему Колмогорова–Арнольда о представлении функций, предоставляет возможность

разлагать высокоуровневые нелинейные зависимости на композиции элементарных функций. Подобная декомпозиция не только повышает прогностическую точность моделей, но и придаёт им свойство универсальной аппроксимации, что расширяет спектр решаемых прикладных и исследовательских задач [97]. Так, в современной физике KAN успешно применяются для численного описания нелинейных динамических систем, где классические алгоритмы оказываются несостоятельными. В этих кейсах сети позволяют детально воспроизводить сложные параметрические взаимосвязи, существенно увеличивая практическую ценность результатов для фундаментальных исследований.

Характерным примером, демонстрирующим уже достигнутый эффект от внедрения KAN, является интеллектуальное планирование и комплексная оптимизация энергетических сетей и инфраструктур. Высокий потенциал KAN в этой сфере обусловлен их способностью оперативно обрабатывать массивные датасеты, выделяя из них значимые корреляции, тренды и скрытые зависимости, что остается малодоступным для традиционных архитектур нейронных сетей [103]. Благодаря столь глубокой интерпретируемости и перенастраиваемости, KAN рассматриваются как перспективный инструмент для машинного обучения, компьютерной инженерии, а также data science, где критичны прозрачность и адаптивная точность моделей [104].

Следовательно, Колмогорово-Арнольдовские сети представляют собой новый эволюционный этап в развитии глубинного обучения, расширяя инструментари й для решения особенно комплексных вычислительных задач. Их врождённая способность восстанавливать причинно-следственные зависимости, выполнять символьную регрессию и автоматически выводить аналитические формулы из эмпирических данных обещает серьёзно повлиять на широкий спектр научных и инженерных областей — от физического моделирования до биоинформатики. Инструменты, которые предоставляет KAN, способны радикально трансформировать практики интерпретируемого машинного обучения, анализа и точного моделирования данных уже в ближайшие годы.

KAN-сети знаменуют собой следующий этап развития архитектуры нейронных сетей, поскольку способны устранять ряд ограничений, характерных для классических подходов. С момента анонса в апреле 2024 года они стимулировали интенсивные исследования и практические эксперименты в междисциплинарных областях — от теоретической физики и прикладной математики до молекулярной генетики [102].

Одним из наиболее наглядных кейсов использования KAN является предсказание топологических инвариантов узлов внутри многокомпонентных систем. Группа под руководством докторанта Цимина Лю совместно с теоретиком-физиком Максом Тегмарком продемонстрировала, что KAN могут автоматически формулировать аналитические выражения, описывающие фазовые переходы, которые остаются за пределами возможностей классических многоуровневых нейросетей. Подобная интерпретируемость делает KAN мощным инструментом для исследователей, решающих глубоко аналитические задачи.

Современные эмпирические работы всё более убедительно демонстрируют, что KAN остаются перспективным инструментом для задач, где необходимо достоверно моделировать нелинейные и многомерные взаимосвязи. Так, в анализе хаотических динамических систем KAN помогают раскрывать скрытые закономерности траекторий и строить точные прогнозы, превосходя классические алгоритмы машинного обучения, нередко теряющие эффективность [97].

Прогресс KAN заметен и в практических внедрениях данных алгоритмов при разработке современных криптографических протоколов. Предполагается, что данная архитектура будет и дальше эволюционировать, задавая вектор развития интеллектуальных систем и кардинально меняя методы анализа и компьютерного моделирования высококомплексных наборов данных [105]. По мере усиления интереса к KAN всё больше научных коллективов привлекают эту технологию для решения интегративных задач, демонстрируя её высокую адаптивность [106].

Тем самым результаты исследований сетей Колмогорова–Арнольда подтверждают их универсальную аппроксимирующую способность и значительный потенциал для разнообразных научных направлений. Каждый новый эксперимент углубляет понимание внутренней структуры этих архитектур и ускоряет их переход к реальным инженерным приложениям, что способно радикально преобразовать современные методы машинного обучения и искусственного интеллекта [102, 105].

KAN-сети наглядно проявляют свои специфические преимущества в ряде прикладных задач, успешно преодолевая недостатки классических архитектур глубокого обучения. Их внедрение уже дало ощутимый эффект в области компьютерного зрения. Так, эти сети использовались как для реконструкции изображений с пониженным разрешением, так и для классической задачи super-resolution. В одном из исследований модель KAN сумела достоверно восстановить мелкие текстурные элементы, поскольку адекватно моделирует нелинейные пространственно-коррелированные зависимости между пикселями, тогда как простые интерполяционные алгоритмы нередко демонстрируют более слабые показатели [102].

Еще одним весомым направлением применения KAN является интеллектуальный анализ крупных массивов данных социальных сетей. Ряд исследований демонстрирует, что эти модели способны детально выявлять сложные паттерны взаимодействия между аккаунтами, с высокой точностью прогнозировать будущую активность пользователей и динамически сегментировать аудиторию в более информативные кластеры, чем традиционные алгоритмы [96]. Такая аналитика формирует основу для высокоточной персонализированной рекламы и тонкой настройки пользовательских интерфейсов.

В экологических исследованиях KAN активно применяются как инструмент для компьютерного моделирования популяционной динамики флоры и фауны. Такие сети позволяют выполнять сценарное прогнозирование эволюции экосистем под влиянием антропогенных факторов и глобальных климатических изменений.

Благодаря высокому разрешению аналитики KAN детально отслеживают демографические параметры, выявляя потенциальные угрозы исчезновения видов ещё на ранних стадиях. Это, в свою очередь, предоставляет учёным и специалистам по консервационной биологии возможность заранее разрабатывать стратегии, направленные на поддержание и восстановление биоразнообразия [107].

Интересным примером внедрения KAN является их использование в производственных секторах, где эти сети служат инструментом для тонкой настройки технологических цепочек и одновременного роста качества выпуска. Так, путём прогнозирования потенциальных дефектов оборудования и узких мест удаётся практически исключить перерасход сырья и образование брака на конвейере. Благодаря архитектурам, способным анализировать не только стандартные метрики, но и многомерные нелинейные зависимости, KAN рассматриваются как устойчивое средство повышения общей эффективности [108].

Дальнейшая эволюция применения KAN-сетей раскрывает новые горизонты для междисциплинарных исследований, формируя прочную основу грядущих прорывов в ИИ. Каждый удачный кейс интеграции подтверждает растущую значимость этих архитектур в медицине, экологии и высокотехнологичном производстве.

Современные исследования архитектур KAN значительно расширяют потенциал нейронных сетей, открывая перспективные направления для решения высококомплексных вычислительных задач. Благодаря впечатляющему превосходству в вычислительной производительности перед классическими многослойными персептронами (MLP) — прежде всего по метрикам точности предсказаний и параметрической экономичности — KAN заслуженно оказываются в центре внимания научного сообщества, специализирующегося на глубоком обучении и смежных дисциплинах. Экспериментальные результаты свидетельствуют, что KAN могут демонстрировать до стократного выигрыша в точности и ресурсной эффективности в сравнении с четырёхслойными MLP, что делает их оптимальным выбором для задач, требующих предельной прецизионности и строгой оптимизации [109].

Хотя обучение KAN занимает больше времени, чем классические MLP, оно прокладывает путь к более глубокой интерпретации и наглядной визуализации высокоразмерных данных. Его оригинальная методика, основанная на решении задач аппроксимации вдоль ребер графа сети, помогает модели эффективно обходить проклятие размерности и преодолевать иные сложности многомерных пространств [97]. Благодаря этому открываются перспективы использования KAN в междисциплинарных проектах AI + Science — от точной подгонки экспериментальных кривых до прямого решения нелинейных уравнений в частных производных, что уже демонстрируют недавние работы исследовательских групп MIT и ряда других институтов [110].

Опираясь на актуальные экспериментальные данные и аналитические обзоры, можно прогнозировать, что KAN вскоре превратятся в ключевой инструмент эволюции ИИ, прежде всего в сегменте high-performance scientific computing и обработки массивов данных. Модульная, адаптивная архитектура сети позволяет динамически подстраиваться под разнородные рабочие нагрузки, тем самым расширяя потенциал приложений в климатическом моделировании, биоинформатике, квантово-финансовой аналитике и прочих вычислительно ёмких дисциплинах [101].

Будущие исследования KAN сосредоточатся, в частности, на оптимизации алгоритмов обучения и сокращении связанных с ним временных издержек; это откроет путь к ещё более массовому применению методологии в реальных задачах. Современные достижения, включая параллельные вычисления и аппаратные ускорители, обещают новые методы бесшовной интеграции KAN в действующие экосистемы машинного обучения, углубляя инструментальный арсенал академии и индустрии [111].

Следовательно, предстоящие исследования Kolmogorov–Arnold Networks (KAN) сосредоточатся на повышении вычислительной эффективности, адаптации под отраслевые домены и разнородные аппаратные архитектуры, разработке продвинутых средств интерпретируемости и доверительного анализа, а также на

человекоориентированном взаимодействии, раскрывая новые горизонты экосистемы глубокого обучения.

Сопоставление KAN-сетей с прочими актуальными алгоритмами машинного обучения, в частности с классическими многослойными персептронами (MLP), подчёркивает оригинальные преимущества KAN-подхода. В отличие от MLP, основанных на заранее заданных нелинкностях ReLU, Sigmoid или Tanh, KAN оперируют параметрическими, то есть обучаемыми, функциями активации. Такая динамическая нелинейность позволяет модели точнее подстраиваться под латентную геометрию выборки и удобно описывать композиционные структуры, что, в свою очередь, снижает влияние проклятия размерности и повышает устойчивость на высокоразмерных задачах [109].

Современные эмпирические работы свидетельствуют, что архитектуры KAN зачастую обеспечивают сопоставимую, а нередко и превосходящую точность относительно значительно более громоздких MLP при решении идентичных задач. Такой эффект объясняется их способностью параллельно извлекать латентные структуры признаков и адаптивно корректировать функции вывода. В частности, при решении задач анализа информации и прогнозирования модели KAN значительно уменьшили величину прогностической ошибки и сократили время обучения, что критично для прикладных исследований в науке и инженерных технологиях [107].

Помимо прочего, ключевой характеристикой KAN является их развитый механизм детализированной визуализации, благодаря которому платформа превращается в эффективный коллаборативный инструмент. Подобная функциональность критична для научных коллективов, занятых формулировкой новых математических и физических принципов. Отображая сетевые структуры и иерархические связи, KAN помогают исследователям глубже анализировать динамику и эволюцию комплексных систем, что, в свою очередь, ускоряет процесс генерации гипотез и их верификации [102].

Следовательно, нейросетевые структуры Колмогорова–Арнольда, кардинально отличающиеся от канонических глубинных моделей, формируют прорывную парадигму, способную усилить, интегрировать и оптимизировать существующие архитектурные решения, тем самым расширяя спектр их применений и открывая перспективные направления в современном машинном обучении.

KAN-сети являются важным этапом эволюции нейронных систем, открывая более эффективные методы представления, обработки и аналитики данных. Опираясь на теорему Колмогорова-Арнольда о суперпозиции непрерывных функций, такая модель демонстрирует нетипичную топологию, отличную от классических глубоких MLP. Вместо фиксированных нелинейностей здесь применяются обучаемые активационные модули, что повышает экспрессивность, адаптивность и способность сети к точному аппроксимированию сложных зависимостей.

Ключевым достоинством архитектуры KAN считается способность задавать активационные функции при помощи сплайнов. Подобный подход не только повышает стабильность и точность процесса обучения, но и обеспечивает сети лучшую адаптацию к высокоразмерным и неоднородным данным, присутствующим в задачах компьютерного зрения и аналитики больших данных. Эксперименты подтверждают, что KAN заметно превосходят традиционные архитектуры по основным метрикам, тем самым расширяя спектр практических применений в научных, инженерных и индустриальных областях.

Современные исследования в сфере KAN подтверждают высокий потенциал этих архитектур, выступающих перспективной альтернативой классическим алгоритмам. Научное сообщество активно изучает их применимость в медицине, финансовой аналитике, робототехнических системах, кибербезопасности и других доменах. Первые внедрения KAN в прикладных проектах уже демонстрируют убедительные результаты, подчеркивая технологическую значимость и практическую ценность подхода.

Тем не менее, при всех своих достоинствах KAN-сети остаются объектом интенсивных исследований. Требуются дополнительные работы, ориентированные на тонкую настройку их топологии, усовершенствование процедур обучения, включая адаптивный градиентный спуск, а также на расширение спектра практических задач. Кроме того, необходимо более детальное сопоставление KAN с альтернативными cutting-edge подходами, чтобы определить их ключевые преимущества, ограничения и оптимальные области применения.

Перспективы применения KAN в грядущих исследованиях представляются исключительно многообещающими. На фоне стремительного прогресса информационных технологий и непрерывного роста массивов обрабатываемых данных эти нейросетевые архитектуры способны стать ключевым инструментарием для решения высокоуровневых вычислительных задач. Их динамическая пластичность, устойчивость к изменяющимся параметрам среды и способность к инкрементальному обучению особенно ценятся академическим сообществом и отраслевыми специалистами.

Несомненно, Kolmogorov–Arnold Networks (KAN) становятся следующим значимым этапом в развитии глубокого обучения. Их оригинальная топология и функциональные преимущества относительно классических архитектур открывают исследователям и инженерам принципиально новые горизонты для экспериментов и внедрения в прикладные задачи самых разных доменов. При нынешних трендах и запросах индустрии машинного обучения KAN способны занять ключевые позиции, ускоряя появление более гибких, масштабируемых и энергоэффективных интеллектуальных систем.

ГЛАВА 4. СИЛЬНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ СИИ

Согласно Национальной стратегии развития искусственного интеллекта до 2030 года [1, 2], все имеющиеся сегодня технологические реализации относятся к категории слабого (узкоспециализированного) ИИ; именно эволюция этих систем должна привести к формированию сильного, универсального ИИ.

Технологические решения, создаваемые на основе методов машинного обучения по пункту 8 настоящей Стратегии, представляют собой тип искусственного интеллекта, ориентированный на выполнение ограниченного круга специализированных задач; такой ИИ относят к категории слабого (или узкоспециализированного) искусственного интеллекта. Напротив, разработка универсального, так называемого сильного или общего ИИ, — системы, способной, аналогично человеку, решать широкий спектр разноплановых задач, самостоятельно рассуждать, эффективно взаимодействовать с окружающей средой и оперативно адаптироваться к динамично меняющимся условиям, — представляет собой чрезвычайно сложную междисциплинарную научно-техническую задачу. Её решение лежит на стыке естественных наук, инженерных дисциплин и социально-гуманитарных исследований. Успешный прорыв в этой области способен кардинально преобразовать ключевые сферы жизнедеятельности общества, однако одновременно несёт в себе риски отрицательных социальных и технологических последствий, неизбежно сопровождающих эволюцию технологий искусственного интеллекта. [1]

Сильный, или общий, искусственный интеллект — это разновидность ИИ, способная решать разноплановые задачи, вести коммуникацию с человеком и автономно подстраиваться под динамику окружающей среды.

Ситуацию в сфере искусственного интеллекта к летнему периоду 2025 года можно описать следующим образом.

Прежде всего, с позиции адаптивности следует отметить, что современные интеллектуальные системы, разработанные на базе любых существующих методологий, рассчитаны на работу внутри узкого спектра задач и контекстов и не умеют автономно перенастраиваться под радикально новые проблемные ситуации.

Иначе говоря, система остаётся программируемой, однако трудоёмкость её кодирования существенно уменьшилась, а интеллектуальная обучаемость значительно увеличилась.

Во-вторых, если говорить об автономности ИИ, то ни одна из современных интеллектуальных систем не является полностью самодостаточной: для запуска, остановки, регламентного обслуживания, целеполагания и выбора эксплуатационных режимов ей неизбежно нужен человек-оператор. Иными словами, контур управления по-прежнему внешнее. Особенно ярко эта зависимость заметна в сферах с повышенными рисками или высокой исходной неопределённостью: достаточно вспомнить, что даже обычный робот-пылесос без человеческой поддержки не сможет работать дольше пары дней в типовой квартире.

В-третьих, если говорить об интегративности — способности разрабатывать автономные программные модули, без труда сочетающиеся с другими элементами — то нынешние ИИ-платформы, по сути, всё ещё не представляют собой полноценные (даже узкоспециализированные) интеллектуальные системы, а лишь:

- системы компьютерного зрения;
- распознавание в пространстве;
- распознавание целей;
- компьютерной обработки естественного языка;
- анализа данных методами машинного обучения;
- обработки символических данных, включая вывод и рассуждения из знаний и т. п.

Следовательно, современные системы нельзя назвать в полной мере интегративными.

Дальнейший прогресс ИИ тесно связан с созданием сильного ИИ [113].

Термин «сильный ИИ» (Strong AI) вошёл в научный оборот в ходе споров о предельных когнитивных возможностях искусственных систем, разгоревшихся в 1980-е. Формальное различие между «сильным» и «слабым» ИИ ввёл философ

Джон Сёрл в программной статье «Minds, Brains, and Programs» (1980) [114], дополнившей дискуссию мысленным экспериментом «Китайская комната», иллюстрирующим проблему семантического понимания алгоритмов.

Сёрл ввёл понятие «strong AI» (сильный ИИ), обозначая тезис, что вычислительная система, будучи снабжена подходящим алгоритмом, способна не только имитировать когнитивную деятельность, но и обладать подлинным сознанием, семантическим пониманием и субъективными переживаниями уровня Homo sapiens. В противовес этому «weak AI» (слабый ИИ) описывает машины, демонстрирующие разумоподобные функции без феноменального осознания и глубинного семантического содержания.

Данные термины продолжают функционировать в современном научно-философском дискурсе в качестве базовых категорий, оставаясь концептуально неизменными.

Мысленный опыт «Китайская комната» [114], сформулированный Джоном Сёрлом в 1980 году, является наглядной критикой тезиса о сильном ИИ. Его суть следующая:

Вообразите оператора, запертого в отдельной комнате. Он абсолютно не владеет китайским языком. Сквозь узкую прорезь в двери ему поступают записки, содержащие вопросы, записанные иероглифами. В его распоряжении имеется детальный алгоритм на родном языке, описывающий, какие последовательности китайских символов нужно сопоставлять и какие знаки выводить в ответ. Ригористически следуя этому алгоритму, оператор формирует ответы и возвращает их наружу всё теми же китайскими иероглифами.

С точки зрения внешних наблюдателей создаётся впечатление, будто в помещении находится человек, свободно владеющий китайским языком, поскольку его реплики выглядят осмысленными. Однако оператор системы на самом деле не улавливает ни смысла входящих фраз, ни содержания собственных ответов, а лишь механически применяет заданные синтаксические алгоритмы.

Ключевая позиция Сёрла такова: даже если вычислительная система безупречно имитирует владение языком и проходит тест Тьюринга, это не свидетельствует о наличии у неё семантического понимания. Машина, как оператор в «китайской комнате», выполняет лишь синтаксическую обработку символов по формальным алгоритмам, не соотнося их со значениями.

Данный мысленный эксперимент превратился в весомый философский контраргумент против тезиса, будто одна лишь компьютерная программа способна породить подлинное сознание и истинное понимание.

С момента, когда Сёрл сформулировал концепцию «сильного ИИ» и аргумент «китайской комнаты», данная парадигма существенно эволюционировала (см. табл. 15).

Таблица 15. Ключевые концепции «китайской комнаты»

Дискуссии и контраргументы	
Системный ответ	Хотя находящийся в комнате оператор не владеет китайским, объединённая система – человек, алгоритмическая инструкция и процедура обработки – может демонстрировать понимание.
Аргумент о виртуальном мозге	Даже на достаточно базовом нейронном уровне мозг способен обрести сознание.
Биологический натурализм	Сёрл уточнил свою концепцию, подчеркивая, что сознание представляет биологический феномен, обусловленный специфической нейробиохимической субстратной базой.
Современные интерпретации	
Общий (СИИ) узкий ИИ	Концепция сильного ИИ постепенно трансформировалась в представление об общем искусственном интеллекте (СИИ) — универсальной архитектуре, способной решать любые когнитивные задачи человеческого уровня.
Проблема сознания	Обсуждение эволюционировало от дилеммы «способен ли алгоритм мыслить?» к более глубокому вопросу: «может ли искусственный интеллект переживать сознательные qualia?»
Интеграция с нейронаукой	Современные исследователи стремятся увязать философские модели сильного ИИ с новыми нейрофизиологическими открытиями о функционировании мозга.
Практические аспекты	
Этические соображения	Перспектива разработки сильного ИИ вызвала дискуссии о юридических правах, этических ценностях и моральном положении гипотетических мыслящих машин.
Сроки достижения	По оценкам специалистов, вероятность появления искусственного общего интеллекта колеблется от десятилетий до столетий, вплоть до гипотезы о его принципиальной недостижимости.

Концепция Сёрла и сегодня подпитывает философские, инженерные и нормативно-этические дебаты вокруг искусственного интеллекта. Она же вдохновила и наше исследовательское коллективное усилие. Хотя авторы в целом принимают прогнозируемые временные рамки, остаются веские оговорки.

По мнению исследователей, эволюция сильного искусственного интеллекта способна радикально трансформировать не только сферу технологий, но и социальные структуры, экономические модели и этические нормы общественной жизни.

На фоне стремительного развития генеративных моделей искусственного интеллекта, таких как ChatGPT, особенно важно осознавать их потенциал для повышения уровня и устойчивости жизни, оптимизации производственных процессов и комплексного преодоления междисциплинарных вызовов, стоящих перед современным обществом.

Тем не менее, помимо очевидных преимуществ, интеграция СИИ порождает и ряд угроз, охватывающих этические дилеммы, социально-экономические эффекты и сугубо технологические вызовы, которые требуется всесторонне учитывать.

Исследователи определяют сильный искусственный интеллект (СИИ) как концептуальную модель, в рамках которой вычислительные системы способны к автономному формированию когнитивных стратегий, глубинному обучению и дедуктивно-индуктивному рассуждению, достигая результатов, сравнимых с компетенцией высококлассных специалистов-людей в различных предметных областях. В отличие от «слабого» ИИ, функционирующего на основе статичных алгоритмов и ограниченного реакций исключительно предзаданными сценариями, СИИ опирается на динамическое самообучение, метакогнитивную рефлексию и адаптивную реконфигурацию собственной архитектуры. Указанные различия формируют основу для понимания масштабного потенциала СИИ, который сегодня интенсивно исследуется и развивается благодаря прорывам в

нейронных сетях, нейроморфных вычислениях и высокопроизводительной вычислительной инфраструктуре.

К ключевым характеристикам самообучающихся интеллектуальных систем относится способность адаптироваться к новым данным и формировать методики решения задач в условиях неопределённости.

Указанные механизмы обеспечивают содержательное и адаптивное взаимодействие с внешней экосистемой, одновременно стимулируя у системы когнитивный анализ и формирование собственных суждений.

В отличие от узкоспециализированного ИИ с фиксированными алгоритмами, сильный ИИ демонстрирует широту функционала.

По распространённому мнению, настоящий искусственный общий интеллект обязан уметь оперировать абстрактными концептами и демонстрировать высокую творческую активность; достижение этого уровня всё ещё требует существенных исследований и ресурсов [115].

Разработка СИИ сегодня рассматривается как одна из самых дерзких и масштабных инженерных амбиций человечества. Несмотря на отсутствие универсального методологического маршрута к созданию сильного ИИ, уже обозначились ключевые исследовательские векторы — от нейроморфных архитектур до когнитивных моделей — и научно-технические предпосылки, чье целенаправленное развитие способно приблизить воплощение данной концепции.

Ключевым условием появления сильного искусственного интеллекта является обеспечение достаточного объёма вычислительных ресурсов. Согласно эмпирическому правилу, получившему название закона Мура, число транзисторов на интегральной схеме удваивается примерно каждые полтора-два года, что влекло за собой лавинообразный рост производительности. Однако по мере исчерпания возможностей кремниевых процессов набирают обороты альтернативные технологические парадигмы — квантовые вычисления, нейроморфные чипы и фотонные процессоры.

Узкоспециализированные аппаратные ускорители, такие как GPU, TPU и нейроморфные процессоры, специально заточены под алгоритмы машинного обучения и нейровычисления.

Квантовые вычисления, опирающиеся на квантовые биты (кубиты), квантовую суперпозицию и явление запутанности, теоретически способны экспоненциально ускорять решение специфических вычислительных задач.

Нейроморфные вычислительные архитектуры, воспроизводящие организацию и функционирование биомозга, способны обеспечивать ультранизкое энергопотребление и масштабируемую параллельную обработку данных.

Гетерогенные распределённые вычислительные системы, интегрирующие различные устройства в цельную программно-аппаратную среду.

Тем не менее до сих пор не решён ключевой вопрос: какой минимальный объём вычислительных ресурсов — будь то пикофлопсы или петафлопсы производительности, объём оперативной памяти и пропускная способность шины — требуется, чтобы воплотить в жизнь полноценный сильный искусственный интеллект?

За годы эволюции искусственного интеллекта разработано немало архитектурных парадигм, каждая из которых обладает собственными преимуществами и ограничениями. При проектировании СИИ наибольшую актуальность приобретают такие архитектуры, как:

Современные глубокие нейронные сети уже продемонстрировали впечатляющие успехи в компьютерном зрении, обработке естественного языка и ряде других узкоспециализированных дисциплин.

Архитектуры класса Transformer, лежащие в основе GPT и родственных моделей, вывели понимание и генерацию текстов на качественно новый уровень.

Тем не менее текущие нейросетевые решения всё ещё страдают от отсутствия интерпретируемости, ограничены в абстрактных выводах и слабо переносят накопленные знания на новые задачи.

Символьные ИИ и логические модели сильны в явном описании знаний, строгой дедукции и прозрачности выводов. Но им трудно работать с неопределённостью, обучаться на «сырых» данных и воспринимать мир.

Гибридные нейро-символьные архитектуры нацелены на синтез сильных сторон обоих парадигм, сочетая способность глубоких нейронных сетей к обработке неструктурированных сенсорных потоков и обучению представлений с логико-символьными механизмами структурированного вывода, формализации правил и эксплицитного хранения знаний.

Когнитивные архитектуры моделируют многообразные процессы человеческого мышления, стремясь объединить сенсорное восприятие, память, механизмы обучения, рассуждение и принятие решений в целостную вычислительную платформу.

В многоагентных системах интеллект трактуется как итог кооперации множества элементарных агентов; такая распределённая самоорганизация порождает эмерджентное поведение, характерное для сложных адаптивных структур.

Многие учёные считают, что путь к полноценному искусственному интеллекту лежит в синтезе символических, нейросетевых, эволюционных и иных методологий, а не в опоре на единственную парадигму.

Ключевым направлением исследований становится создание алгоритмов, гарантирующих прозрачность, интерпретируемость и кибербезопасность ИИ, что приобретает особую значимость при развитии сильного искусственного интеллекта, способного потенциально радикально трансформировать современное общество.

Современные ИИ-технологии и алгоритмы машинного обучения, включая глубокие нейронные архитектуры, добились впечатляющего прогресса, тем не менее идея полноценного сильного искусственного интеллекта всё ещё вызывает интенсивные научные дебаты.

На передний план в дискуссиях о потенциале разработки подобных систем выходят острые этические дилеммы, онтология сознания и сложнейшие инженерные ограничения.

Как отмечают исследователи, финальная задача создания СИИ выходит далеко за рамки простого копирования поведенческих паттернов, стремясь к глубокому моделированию когнитивных механизмов мышления и решений, что открывает новые возможности для науки и высоких технологий [113].

Следовательно, современные исследования СИИ, так же как его практические парадигмы реализации и философско-эпистемологические основания, продолжают оставаться ключевыми объектами научного внимания.

Эти направления подчёркивают необходимость углублённого интердисциплинарного анализа и непрерывного теоретико-прикладного поиска, способных обеспечить устойчивый прогресс в данной сложной, многогранной области [116].

За последние годы технологии генеративного искусственного интеллекта стремительно эволюционировали; кульминацией стал 2023-й, когда крупные языковые модели вроде ChatGPT продемонстрировали коммерческий и исследовательский прорыв.

Современные модели демонстрируют впечатляющий потенциал в генерации текстов, синтезе изображений и прочего мультимедийного контента.

В образовательных и исследовательских организациях фиксируется всесторонняя интеграция цифровых и ИКТ-технологий, радикально модернизирующая учебный процесс.

Согласно аналитическим прогнозам, ожидается, что вложения в технологии генеративного искусственного интеллекта возрастут с нынешних 40 млрд долларов (2023 г.) до свыше 140 млрд к 2030 г., демонстрируя их колоссальный потенциал радикально трансформировать образовательные методики, дидактические практики, цифровое контент-создание и персонализированные модели обучения [117].

Генеративные модели ИИ находят применение во множестве доменов. В образовании всё большую значимость приобретают решения на базе ChatGPT: они обеспечивают студентам адаптивную обратную связь и оптимизируют подготовку письменных работ.

Современные цифровые решения трансформируют педагогические стратегии и расширяют потенциал персонализации образовательных траекторий, тем самым усиливая мотивацию и вовлечённость студентов в учебный процесс [118].

Помимо этого, в академической сфере генеративные модели искусственного интеллекта задействуются для высокопроизводительной обработки массивов экспериментальных данных и автоматизированного выдвижения обоснованных гипотез, что ускоряет аналитические этапы и оптимизирует процесс научного решения задач, повышая результативность проектов [119].

К примеру, всё больше корпораций интегрируют генеративные ML-модели в производственные цепочки, повышая эффективность операций, минимизируя операционные расходы и укрепляя свои конкурентные позиции на рынке.

Тем не менее, по мере стремительного распространения генеративных нейросетей возникают этические и правовые вызовы.

Вопросы этичности использования современных технологий становятся всё более значимыми.

Опасения исследовательского сообщества и разработчиков вызывает вероятность подмены авторства, а также целенаправленное использование генеративных нейросетей для фабрикации дезинформации и фейковых новостей [120]. Дополнительно следует учитывать присущую ряду генеративных архитектур непредсказуемость и дефицит стабильности, что может спровоцировать нежелательные эффекты при их практическом внедрении в различных приложениях, включая журналистику, образование и цифровой маркетинг.

Итоги прогресса в сфере генеративного ИИ обозначили качественно новую веху эволюции цифровых технологий. Теперь обществу необходимо сформировать ответственный, регулируемый и этически выверенный механизм его интеграции, чтобы раскрыть потенциал инноваций, одновременно снижая сопутствующие риски и угрозы.

Интеграция систем искусственного интеллекта в рабочие процессы радикально трансформирует мировую экономику, переопределяя методы выполнения задач и формируя новые профессиональные роли.

СИИ способен автоматизировать рутинные операции, что позволяет работникам освободить время для более сложной и креативной работы. Это развитие

открывает новые горизонты для специалистов в разных областях, особенно в тех, где требуется креативное мышление и инновационные подходы.

По мере того, как СИИ уменьшают спрос на персонал, занятый рутинными операциями, параллельно трансформируются и ценностные ориентиры на рынке труда. Сегодня работодатели особенно высоко оценивают адаптивность и творческое мышление, что обязывает сотрудников оперативно осваивать новые компетенции и регулярно проходить переквалификацию.

Современные технологии, функционирующие на базе искусственного интеллекта, особо подчёркивают значимость междисциплинарной компетентности и умения эффективно работать в команде, что делает актуальной быструю трансформацию образовательных программ под динамику рынка занятости [9].

Внедрение систем искусственного интеллекта позволяет компаниям заметно увеличивать операционную эффективность. Усиленная обработка big data и адаптивные алгоритмы машинного обучения обеспечивают более оперативную реакцию на колебания рынка. Специалисты получают расширенный набор аналитических инструментов, повышают точность решений и предлагают клиентам более высокую добавленную стоимость.

Вместе с тем распространение решений, основанных на системах искусственного интеллекта, сопровождается рядом вызовов. Вопрос сокращения рабочих мест остаётся существенным источником общественной тревоги. Необходимо не только пересмотреть стратегии профессионального образования и программ рескиллинга, но и укрепить социальную поддержку, облегчающую гражданам переход к цифровой экономике. Дефицит компетентных специалистов в изменённой производственной среде всё острее ощущается работодателями, что требует координированных усилий государства, бизнеса и академического сообщества.

Внедрение систем искусственного интеллекта способно вызвать серьёзные социальные изменения.

В ряде ситуаций технологическая автоматизация способна усугубить положение наименее защищённых слоёв, в частности работников с низким уровнем квалификации, что потенциально провоцирует рост социальных напряжений и неравенства. Противодействовать данным рискам можно, развернув более адаптивную

архитектуру рынка труда, предусматривающую системную переподготовку, развитие цифровых компетенций и поддержку сотрудников, которые массово переходят к инновационным форматам занятости и организации процесса [123].

Следовательно, совокупность преимуществ и рисков, сопровождающих интеграцию СИИ в производственные и организационные процессы, диктует потребность в всестороннем исследовании и конструктивном общественном диалоге.

Для максимально продуктивного внедрения СИИ требуются комплексные стратегии, которые, наряду с экономической рентабельностью, анализируют и социальные последствия, обеспечивая устойчивое и долгосрочное развитие профессий и способствуя прогрессивной трансформации общества в целом [124].

Бурное развитие искусственного интеллекта неизбежно приводит к переосмыслению существующих этических парадигм, регулирующих жизнь цифрового социума.

Проблематика конфиденциальности данных обостряется, поскольку интеллектуальные алгоритмы систематически агрегируют, анализируют и архивируют персональные сведения. Большинство пользователей не отдает себе отчёта в масштабе собираемых метаданных и потенциальных сценариях их дальнейшей обработки.

Эмпирические исследования свидетельствуют, что значительная доля аудитории пропускает лицензионные и сервисные соглашения, а следовательно, недооценивает угрозы несанкционированного доступа, репрофилирования или коммерциализации информации [125].

Помимо этого, вопрос алгоритмической предвзятости порождает весомые морально-правовые дилеммы. Модели машинного обучения, тренированные на неполных либо искажённых выборках, способны генерировать дискриминационные выводы. Так, применение подобных систем в кредитном скоринге или в сфере правоприменения рискует обернуться несправедливым ущемлением определённых социальных групп [126].

Соответственно, проблема распределения ответственности разработчиков оказывается в центре этического дискурса: кто должен отвечать за просчёты искусственного интеллекта? Один из предлагаемых путей — введение детальных профессиональных этических кодексов, стимулирующих алгоритмическую транспарентность и надёжность, а также обязательного независимого аудита моделей и регламентов explainability [127].

С этической позиции особого внимания заслуживает дискуссия о формировании у ИИ автономной системы ценностей. Ключевой вопрос — какой алгоритмический каркас задаст машине моральные принципы и каким образом он будет коррелировать с человеческими этическими установками. Создание подобных регламентов неизбежно требует междисциплинарной экспертизы философов, программистов и правоведов [128].

Помимо технических, существуют и юридические аспекты эксплуатации ИИ-систем. В сегодняшней цифровой среде часто образуются правовые лакуны, порождающие дискуссии о том, каким образом внедрять регуляторные механизмы и стандарты комплаенса, не подавляя инновационную динамику. Ситуацию усложняет то, что классические нормы права традиционно отстают от бурного технологического прогресса [129].

Значимость подобных дебатов акцентирует потребность в выработке единого, междисциплинарного свода этических норм и правовых механизмов, регулирующих искусственный интеллект. Дальнейшая эволюция цифровых технологий во многом определится тем, насколько эффективно социум урегулирует эти комплексные вызовы и создаст нормативную инфраструктуру для защиты цифровых прав личности и поддержания алгоритмической справедливости; в противном случае общество и отдельные граждане рискуют столкнуться с серьёзными негативными последствиями.

Интеграция сильного искусственного интеллекта радикально трансформирует общественную структуру, затрагивая экономику, культуру и политические институты. Социальные эффекты этой трансформации окажутся одновременно конструктивными и деструктивными. Эксперты прогнозируют резкие изменения

в распределении благ из-за неравномерного доступа к когнитивным технологиям. Индивиды с высоким уровнем дохода и человеческого капитала быстрее осваивают новые инструменты, тогда как у малообеспеченных слоёв возникает технологический разрыв, усиливающий стратификацию и потенциально провоцирующий масштабные протесты и социополитические потрясения.

Одним из наиболее ощутимых последствий станет сокращение числа рабочих мест в результате масштабной автоматизации производственных и сервисных процессов. Интеллектуальные системы исполняют операции быстрее, точнее и дешевле человека, что чревато исчезновением целых профессиональных ниш. Наибольшую уязвимость демонстрируют низкоквалифицированные кадры, чьи функции легко реплицируются алгоритмами. Появится насущная потребность в массовой переквалификации персонала, однако далеко не все смогут адаптироваться, что усилит уровень безработицы и усугубит социально-экономическое расслоение в глобальном масштабе [131].

С другой стороны, интеграция СИИ способна ощутимо поднять уровень благополучия граждан, оптимизируя сервисы и рабочие потоки в сферах образования, здравоохранения и управления городским хозяйством. Так, внедрение умных педагогических платформ и алгоритмов адаптивного обучения открывает путь к более персонализированному, доказательно эффективному образовательному процессу, тогда как в медицине интеллектуальные системы поддержки клинических решений улучшают диагностику и терапию.

Однако итоговый эффект зависит от тщательного этико-гуманитарного анализа, призванного предотвратить технологическую зависимость и усиление цифрового неравенства.

Помимо этого, важно учитывать культурные трансформации, которые сопровождают экспансию сильного ИИ. Форматы человеческого взаимодействия будут постепенно меняться, и эти сдвиги отразятся на структуре социального капитала. Если часть коммуникации сместится к цифровым агентам, люди могут снизить ценность живого контакта.

Следовательно, обществу придется выработать новые социальные нормы и адаптивные практики, ориентированные на гармоничное сосуществование людей и автономных систем.

В конечном итоге технологические аспекты развертывания сильного искусственного интеллекта требуют особой взвешенности. Рост масштабов его использования вызывает потребность в многоуровневых системах кибербезопасности и в механизмах нейтрализации рисков, обусловленных устранением человеческого контроля в критически значимых секторах. Эффективная интеграция СИИ требует не только инженерной, но и этической готовности общества, позволяющей сохранить базовые гуманистические ценности и поддерживать устойчивое развитие на глобальном, национальном и локальном уровнях.

Следовательно, эволюция сильного ИИ сулит широкие общественные выгоды и одновременно ставит острые вызовы, которые следует учитывать при формировании регуляторной политики и практических моделей его применения.

В процессе разработки систем искусственного интеллекта (СИИ) приходится преодолевать многочисленные технологические барьеры для достижения заметных результатов.

Ключевым вызовом остаётся разработка решений, органично встраивающихся в устоявшуюся ИТ-инфраструктуру предприятия. Большинство передовых технологий предполагает модификацию или рефакторинг наследственных платформ и интерфейсов, что усложняет процесс деплоя и миграции данных. Кроме того, многокомпонентные архитектуры подвержены каскадным отказам, поэтому растут требования к отказоустойчивости, кибербезопасности, совместимости между сервисами и обеспечению непрерывности бизнес-процессов.

Вопрос безопасности эксплуатации СИИ остается фундаментальным. В условиях высокорисковой сферы здравоохранения даже единичная ошибка способна повлечь катастрофические клинические последствия.

Рост степени автоматизации одновременно усиливает технологическую зависимость и усложняет контроль качества. Особенно тревожат так называемые

«галлюцинации» — непредсказуемые сбои, формирующие недостоверные выводы и тем самым подрывающие доверие пользователей и регуляторов к интеллектуальным системам.

Чтобы минимизировать эти угрозы, разработчики обязаны внедрять многоуровневую валидацию данных, объяснимые модели и механизмы fail-safe, а также постоянно донастраивать алгоритмы с учетом реальных клинических сценариев.

Внедрение СИИ в уже функционирующие экосистемы документооборота, корпоративных БД и сетевой инфраструктуры сопряжено с серьёзными трудозатратами.

Практически на каждом этапе всплывают дополнительные задачи, требующие временных, финансовых и людских ресурсов. Необходимо обеспечить программно-аппаратную совместимость, унификацию форматов данных, корректное маппирование атрибутов и надёжно защищённый обмен пакетами информации между гетерогенными узлами [138].

Не менее значимо продвигать когнитивные науки, на фундаменте которых строятся передовые алгоритмы сильного искусственного интеллекта.

Чтобы вывести ИИ на по-настоящему высокий, «сильный» уровень, необходимы углублённые междисциплинарные исследования в психологии, нейробиологии, когнитивной лингвистике и смежных областях, призванные вскрыть нейронные и когнитивные механизмы мышления и восприятия.

Последовательное развитие этих направлений позволит проектировать более гибкие, самоадаптирующиеся когнитивные архитектуры, адекватно реагирующие на разнообразие задач и внешних условий.

В конечном счёте для преодоления обозначенных вызовов требуется вести интенсивные фундаментальные и прикладные исследования, охватывающие несколько взаимосвязанных направлений.

Одновременно с созданием инновационных алгоритмов, архитектур и аппаратно-программных решений следует формировать унифицированные стандарты, регуляторные нормы и адаптивную правовую базу, обеспечивающие без-

опасное и продуктивное внедрение СИИ в будущем. Достижение этой цели предполагает междисциплинарное взаимодействие учёных, инженеров и законодателей, минимизирующее риски, связанные с непредвиденными факторами.

Прогресс систем искусственного интеллекта открывает широкие перспективы в медицине, образовательном секторе и экономической сфере.

В медицинской отрасли системы искусственного интеллекта (СИИ) способны радикально преобразить процесс постановки диагнозов и ускорить разработку индивидуализированных терапевтических стратегий.

Применение алгоритмов машинного обучения при обработке клинических, лабораторных и геномных данных существенно повышает точность диагностических выводов и прогнозирования исходов лечения [133].

Так, ИИ-платформы анализируют рентгенологические снимки и МР-изображения, фиксируя микроскопические аномалии, которые традиционные методики нередко упускают.

Сфера образования стоит на пороге радикальных перемен благодаря интеграции систем искусственного интеллекта. Персонализированные траектории обучения, формируемые через образовательную аналитику, учитывают потребности, текущий уровень компетенций и динамику продвижения каждого учащегося, тем самым существенно повышая результативность освоения материала.

Алгоритмы СИИ создают адаптивные e-learning-платформы, которые, реагируя на индивидуальный стиль восприятия, предлагают релевантные ресурсы, упражнения и оценки, обеспечивая более глубокое закрепление знаний.

В сфере экономики потенциал внедрения систем искусственного интеллекта чрезвычайно высок. Их применение для автоматизации обработки рыночных данных, построения предиктивных аналитических моделей и динамического ценообразования делает управление компаниями заметно более результативным.

Платформы, анализирующие массивы информации о поведении потребителей, формируют рекомендации по оптимизации прибылей и минимизации инвестиционных рисков, что сокращает издержки и ускоряет управленческие решения.

Помимо этого, в областях транспортной и городской инфраструктуры системы искусственного интеллекта способны кардинально изменить устоявшиеся практики администрирования.

Внедрение интеллектуальных платформ, отвечающих за адаптивную координацию дорожного трафика и диспетчеризацию общественного транспорта, способствует сокращению пробок, уменьшению времени поездок и, как следствие, повышению качества городской среды.

Использование СИИ предполагает интеграцию алгоритмов машинного обучения, средств предиктивной аналитики и механизмов динамического перераспределения ресурсов, что позволяет формировать более устойчивые, гибкие и ориентированные на пассажира транспортные экосистемы.

Потенциал сильного искусственного интеллекта практически безграничен, охватывая едва ли не все сферы общественного бытия — от экономики и образования до культуры и управления. Интеграция таких когнитивных систем способна не просто повысить продуктивность и точность операций, но и радикально трансформировать саму логику человеческих коммуникаций и кооперации.

Тем не менее, наряду с технологическим рывком необходимо всесторонне анализировать этические, правовые и социогуманитарные риски применения СИИ, чтобы гарантировать, что грядущая техно-социальная парадигма служит развитию человека и усилению социальной устойчивости.

СИИ представляет собой концепцию, которая нацелена на создание машин, способных к самостоятельному мышлению и обучению, что открывает перед человечеством как уникальные возможности, так и серьезные вызовы. В ходе исследования были рассмотрены ключевые аспекты, касающиеся как достижений в области генеративного ИИ, так и потенциальных последствий его внедрения в различные сферы жизни.

Ключевым выводом является то, что прогресс в сфере генеративного искусственного интеллекта — к примеру, систем уровня ChatGPT — уже сегодня

демонстрирует ощутимый потенциал для радикального роста производительности труда.

Модели такого класса позволяют ускорять автоматизацию рутинных операций, повышать качество клиентского опыта за счёт мгновенной обработки запросов и даже инициировать совершенно новые форматы контент-генерации.

Тем не менее, помимо очевидных преимуществ, следует тщательно оценивать и сопутствующие риски эксплуатации СИИ. В частности, речь идёт о вероятности вытеснения определённых рабочих ролей, а также о формировании технологической зависимости, которая способна негативно отразиться на человеческом факторе и устойчивости профессиональных сообществ.

Этические вызовы, сопровождающие внедрение СИИ, заслуживают пристального рассмотрения. Проблематика сохранения приватности данных, распределения ответственности за автономные решения алгоритмов и риска целенаправленной манипуляции массовым сознанием становится критически значимой.

Требуется выстроить жесткий свод этических стандартов, процедур аудита и регуляторного комплаенса, способный снизить угрозы и гарантировать безопасную эксплуатацию ИИ-технологий. Принципиально, чтобы разработчики, интеграторы и конечные пользователи СИИ осознавали свою роль и последовательно воплощали принципы безопасности, транспарентности и справедливости.

Влияние внедрения систем искусственного интеллекта на социум носит амбивалентный характер. С одной стороны, такие решения способны повышать качество жизни граждан, облегчая доступ к инновационным сервисам, информационным ресурсам и персонализированным возможностям.

С другой — они рискуют усилить цифровое неравенство, если технологические преимущества останутся прерогативой узких социальных групп. Следовательно, государству и бизнесу важно разрабатывать программы инклюзивной цифровой трансформации, чтобы все слои населения получили равный доступ и реальную выгоду.

Недооценивать технологические барьеры, сопровождающие эволюцию систем искусственного интеллекта, крайне рискованно. Для обучения моделей требуются обширные массивы очищенных данных, масштабируемые GPU-кластеры и устойчивые алгоритмические архитектуры, что влечёт серьёзные финансовые, инфраструктурные и кадровые затраты компаний и академических центров.

Лишь тесная синергия лабораторий, отраслевых партнёров и венчурного капитала позволит преодолеть эти ограничения и поддерживать долгосрочный прогресс технологий.

Подводя итог, перспективы развития сильного искусственного интеллекта (СИИ) открывают для человеческой цивилизации колоссальный спектр возможностей, одновременно актуализируя целый ряд серьёзных вызовов.

Чтобы полноценно раскрыть потенциал СИИ, нам предстоит углублять фундаментальные исследования, формировать гибкую систему этических и правовых протоколов, а также оттачивать механизмы безопасного и прозрачного внедрения.

Лишь при таком подходе получится выстроить симбиотическое партнёрство человека и машины, ускоряющее прогресс и повышающее качество жизни каждого.

Проблематика верификации сильного СИИ заслуживает пристального изучения, поскольку сочетает в себе многогранную природу явления и целый спектр философских, этических и инженерных дилемм.

Масштабы, а также принципиальные различия между СИИ и сегодня доминирующим слабым ИИ невозможно недооценить. СИИ, по каноническому определению, способен решать произвольные задачи на уровне человека, проявляя сознание и интенциональность, тогда как слабый ИИ – это лишь совокупность алгоритмов, которые симулируют когнитивные действия, оставаясь без подлинного понимания и самосознания [134].

Отсутствие единой, признанной научным сообществом методики измерения свойств СИИ подчеркивает острую потребность в выработке интегральных

критериев и регламентов, охватывающих когнитивные, поведенческие и этические измерения предполагаемого интеллекта.

Современные ИИ-платформы, опирающиеся преимущественно на глубокие нейронные сети, демонстрируют впечатляющую функциональность, но всё ещё остаются в категории узкоспециализированных систем.

Их работа базируется на выявлении статистических корреляций и программных откликах, что ставит под сомнение наличие подлинного семантического понимания. Отсюда возникает ключевой вопрос: какие индикаторы позволят убедительно зафиксировать момент перехода к подлинному СИИ?

Философские рассуждения и мысленные эксперименты, например знаменитая «Китайская комната» Джона Серла, ярко демонстрируют многослойность данной проблемы.

Рассмотрение сильного искусственного интеллекта как концептуальной конструкции поднимает вопрос о принципиальной разнице между тем, что алгоритм способен выполнять, и тем, обладает ли он подлинным «пониманием», — различие, критически важное для критерия оценки и архитектурного проектирования подобных систем. Поэтому дальнейшие исследования должны уделять внимание не только поведенческим функциям, но и глубинному анализу природы сознания внутри когнитивных процессов.

Формирование универсальной методики верификации крайне важно: она задаёт согласованные критерии и позволяет достоверно судить, достиг ли прототип требуемого уровня когнитивной сложности и рефлексивного самосознания. Актуальные исследования подчёркивают, что модель обязана интегрировать вероятностно-статистические инструменты с философско-онтологическими основаниями и этико-правовым анализом перспектив и рисков сильного ИИ.

Процедура оценки ИИ охватывает широкий спектр аспектов, затрагивающих как эксплуатационные параметры, так и потенциальные социально-этические эффекты.

Современные методологии проверки ВИ (высокоинтеллектуальных) систем строятся из нескольких фаз: от формализации задач и формулирования метрик до долговременного аудита эффективности в реальном окружении.

Каждая схема включает типовой набор шагов. В первую очередь необходимо строго зафиксировать целевую функцию, уточнив, какую конкретную задачу должен решать алгоритм.

При построении моделей применяют разнообразные алгоритмы машинного обучения, которые гибко настраиваются под специфику решаемых задач.

На этапе валидации используют такие показатели, как precision, recall, F1-score и баланс классов. Затем успешные решения переводят в продуктив, где необходим непрерывный мониторинг их производительности и оперативное обновление обучающего корпуса при выявлении деградации [135].

Тем не менее, подобные решения не лишены недостатков. Так, традиционные индикаторы, такие как точность (accuracy), полнота (recall) или даже F1-мера, далеко не всегда отражают реальную полезность моделей, особенно для многогранных задач с неравномерными классами.

Более того, стремление «выжать» максимальную точность нередко приводит к переобучению и падению эффективности при эксплуатации в полевых условиях. Наконец, чрезмерно жесткие бенчмарки и объемные тест-корпусы влияют на общественное восприятие технологии и осложняют стратегию её практического внедрения.

Трудность проявляется ещё и в том, что необходимо согласовать работу нескольких взаимосвязанных систем, где каждая метрика апеллирует не только к строгим математическим критериям, но и к нормативно-этическим параметрам — равноправию и справедливости. Это особенно критично при анализе наборов данных, которые зачастую оказываются неполными или искаженными, а значит, необъективными.

Системы метрик и бенчмарков, создаваемые для валидации качества искусственного интеллекта, становятся фундаментом его эксплуатационной надежности и безопасности.

Эти процедуры аудита и стресс-тестирования необходимо непрерывно совершенствовать, своевременно дополняя их новыми алгоритмами, объективными эталонами и сценариями испытаний, отражающими эволюцию технологий.

Такой динамический мониторинг минимизирует риски социального ущерба — от алгоритмического неравенства до скрытой дискриминационной предвзятости данных.

Следовательно, современные методики валидации сильного искусственного интеллекта строятся на интегральном, многофакторном подходе и требуют непрерывного совершенствования как на этапах проектирования и эксплуатации алгоритмов, так и при обеспечении нормативно-этической безопасности их развертывания, формируя базу для углубленного анализа и формализации универсальных метрик оценки.

СИИ определяется через несколько ключевых критериев, которые помогают различать его от слабого искусственного интеллекта. Главными из них являются сознание, понимание, самообучение, решение задач и эмпатия [113].

Понятие сознания в контексте ИИ предполагает, что алгоритмическая система обладает рефлексивным самосознанием, то есть осознаёт и собственное бытие, и динамику внешней среды. Следовательно, ей необходимо не просто перерабатывать входные данные, но и поддерживать внутреннюю метакогнитивную модель, формулировать автономные цели, прогнозировать возможные эффекты своих интеракций и корректировать стратегии.

Научные работы свидетельствуют: такой уровень мета-представлений критичен для построения полноценного сильного искусственного интеллекта.

Для ИИ когнитивная компетенция «понимание» означает умение активно взаимодействовать с внешней средой и корректно трактовать поступающие данные.

Этот процесс опирается на контекстуальный анализ, алгоритмы распознавания паттернов, а также на ассоциативные модели мышления. При отсутствии таких механизмов система теряет способность глубоко осмыслять и прогнозировать события, что радикально снижает её функциональную ценность.

Самообучение в качестве ключевого критерия описывает способность искусственного интеллекта гибко адаптироваться и эволюционировать, опираясь на поступающие датасеты и потоковые данные. Такое свойство достигается с помощью множества техник машинного обучения: глубоких нейронных сетей, обучения с подкреплением, трансфер-обучения и непрерывного обучения.

Грамотная самоадаптация даёт СИИ возможность выводить закономерности, оптимизировать модели и генерировать оригинальные решения, превращая систему в универсальный инструмент.

Способность искусственного интеллекта к решению задач охватывает не только этапы распознавания и семантической интерпретации данных, но и формирование оптимальных, контекстно-зависимых стратегий в соответствии с заданными ограничениями.

Это включает комплексные аналитико-прогностические процедуры, где система обязана вырабатывать многоуровневые, стратегические решения и динамически адаптировать их в реальном времени к изменяющейся среде.

В рамках сильного искусственного интеллекта эмпатия определяется как способность улавливать, интерпретировать и внутренне моделировать эмоциональные состояния других агентов — будь то человек или, в перспективе, другой ИИ. Такой когнитивно-аффективный механизм является фундаментом эффективного взаимодействия с пользователями и машинами, где доминируют социально-эмоциональные аспекты. Эмпатичная система создает более естественный диалог и укрепляет доверие к цифровым технологиям [137].

Каждый из перечисленных критериев формирует надежную основу для систематического анализа и многоуровневой валидации сильного искусственного интеллекта.

Комплексная оценка по этим метрикам демонстрирует, насколько архитектура приближена к человеческому когнитивному порогу и каким образом алгоритмы могут интегрироваться в разнообразные социальные, производственные и научные контексты.

Современные интеллектуальные системы — от алгоритмов машинного обучения до глубоких нейронных сетей — проектируются с учётом множества критериев на всех стадиях жизненного цикла: от формулировки задачи и подготовки датасетов до внедрения и сопровождения. В качестве методологической опоры служит национальный стандарт ГОСТ Р 59898-2021, регламентирующий комплексную оценку качества ИИ-решений.

Документ выделяет показатели надёжности, информационной безопасности, функциональной полноты и сферу этических рисков, что критично для формирования пользовательского доверия [136]. Регулярные аудиты по стандарту позволяют своевременно обнаруживать и корректировать дефекты, тем самым стабилизируя работу моделей и повышая их совокупную эффективность.

Хорошим примером успешного соблюдения подобных критериев являются решения в области компьютерного зрения, например Google Vision. Платформа демонстрирует исключительную точность детекции и классификации объектов благодаря тщательному препроцессингу данных, применению продвинутых нейросетевых архитектур и регулярному переобучению. При этом датасеты проходят многоэтапную валидацию и стресс-тесты, что гарантирует соответствие заданным стандартам качества, а также надежный контроль версионирования как моделей, так и данных.

Ряд программно-аппаратных комплексов оказался уязвимым, поскольку не обеспечил должный уровень надежности и отказоустойчивости. В особенности это проявилось в решениях, обученных на предвзятых или несбалансированных выборках: алгоритмы распознавания лиц выдали значительное число ложных срабатываний и пропусков, главным образом в отношении крупных пользовательских групп, таких как этнические и культурные меньшинства.

Подобные инциденты демонстрируют критическую необходимость внедрения продвинутых протоколов тестирования, процедур валидации и комплекс-

ного метрического аудита, чтобы удостовериться, что рассматриваемые технологии не только выполняют целевые функции, но и остаются надежными, безопасными и социально ответственными инструментами [139].

При анализе практических кейсов необходимо выделить направления, в том числе рекомендательные системы и разнообразные чат-боты, демонстрирующие широкий спектр качественных показателей. Так, специализированные ассистенты для клиентской поддержки способны генерировать высокий уровень удовлетворённости аудитории благодаря контекстному пониманию и точным ответам.

В то же время упрощённые архитектуры нередко испытывают трудности при обработке многоступенчатых запросов и актуализации данных, что характерно для бот-платформ, слабо владеющих механизмами обработки естественного языка [138].

Методики оценки, базирующиеся на стандарте ГОСТ, дают возможность не просто измерять качество информационных систем на всех стадиях их жизненного цикла, но и совершенствуют процессы обработки данных, валидации и тестирования. Гибкая адаптация специфических требований к датасетам и контрольным сценариям является ключевым фактором, повышающим уровень доверия конечных пользователей.

Данные принципы важны для любой индустрии, однако особого внимания требуют решения с признаками «общего» искусственного интеллекта, где надежность, воспроизводимость и безопасность выходят на первый план.

Следовательно, становится всё более очевидной потребность в внедрении унифицированных методик валидации и бенчмаркинга систем искусственного интеллекта, что позволит не только повысить их надёжность и эффективность, но и укрепить общественное доверие к подобным решениям.

ГЛАВА 5. ЕСТЕСТВЕННЫЙ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ КАК ИНСТРУМЕНТ ХРАНЕНИЯ ДАННЫХ

За последние десятилетия наблюдается бурный прогресс в технологиях, ориентированных на высокоэффективную обработку и долговременное хранение данных. В этом контексте повышенное внимание уделяется как ЕИ, так и системам ИИ.

Человеческий естественный интеллект охватывает умение усваивать новые знания, проводить глубокий аналитический разбор, формировать обоснованные решения и интуитивно постигать многоуровневые ситуации.

Искусственный интеллект, по сути, представляет собой комплекс программно-аппаратных решений, предназначенных для решения задач, требующих когнитивных способностей, включая интеллектуальную аналитику, машинное обучение, распознавание паттернов и роботизацию бизнес-процессов. Проанализируем, каким образом природный и вычислительный интеллект способны синергично взаимодействовать при хранении, интеграции и обработке данных.

ЕИ и ИИ — две альтернативные парадигмы когнитивных систем, знания и обработки данных.

Первая форма проявляется в когнитивных компетенциях биологических организмов, охватывая процессы обучения, сенсорного восприятия, логического мышления и стратегического решения задач.

В отличие от этого, искусственный интеллект создаётся человеком специально для воспроизведения указанных когнитивных операций. Разработчики ИИ формируют математические модели и программные алгоритмы, способные инициировать сопоставимые поведенческие акты без обращения к биологическим либо психологическим субстратам, характерным для природного разума.

Методологические стратегии, применяемые для изучения этих форм интеллекта, разнятся.

Естественный интеллект традиционно анализируется через призму философских и фундаментально-научных парадигм, тогда как искусственный интеллект в основном проектируется под конкретные прикладные задачи и инженерные требования, часто лишённые глубокой онтологической базы. Уже в первой главе авторы отмечали, что значительная часть современных ИИ-разработок не соотносится с универсальными законами, характерными для ЕИ. Отсюда возникает необходимость переосмыслить их взаимосвязь и сформулировать общие методологические принципы для обеих дисциплин.

Стоит подчеркнуть, что в современной образовательной экосистеме активно изучаются как ИИ, так и ЕИ. Их сопоставительный анализ помогает глубже понять преимущества и ограничения каждого подхода, выявить синергетические эффекты и перспективы внедрения в учебные практики. Таким образом, образование превращается в экспериментальную площадку для коэволюции этих форм интеллекта, открывая новые векторы их междисциплинарного применения.

Взаимодействие ЕИ и ИИ создает мощный синергетический эффект, расширяющий возможности оптимизации архитектур хранения и многоуровневой аналитики данных. Комплементарное сочетание когнитивных способностей человека и алгоритмических мощностей машин обеспечивает адаптивное решение задач, сопровождающих обработку экстремальных массивов информации.

ЕИ, обладая творческим мышлением и критическими навыками, умеет ставить ключевые вопросы и формулировать аргументированные гипотезы.

Современный искусственный интеллект, применяя алгоритмы машинного обучения и Big Data, быстро обрабатывает данные, выявляя скрытые закономерности и связи.

Хорошим примером эффективной синергии считается внедрение гибридных образовательных платформ, где искусственный интеллект берет на себя рутинные задачи, тем самым высвобождая время преподавателя для творческих и методически сложных аспектов обучения [138]. Когда учебные заведения под-

ключают ИИ-модули для анализа данных слушателей, адаптивного планирования курсов и автоматизации оценивания, это одновременно уменьшает рабочую нагрузку персонала и поднимает общий уровень образовательного результата.

В научной сфере всё чаще комбинируют ЕИ и ИИ: естественный интеллект формулирует теоретические модели и гипотезы, тогда как алгоритмы машинного обучения и нейросетевые методы берут на себя высокоскоростную обработку массивов эмпирических данных. Такая синергия ускоряет цикл исследований, будь то секвенирование и интерпретация геномов, виртуальный скрининг молекул или квантово-химический расчёт, открывая новые инструментарии для решения сложных биотехнологических и химико-фармацевтических задач [20].

Кроме того, необходимо подчеркнуть, что онтологические, философские и нормативно-этические аспекты коэволюции ЕИ с ИИ играют столь же значимую роль в данном процессе.

Осознание взаимной роли человека и высоких технологий в конструировании грядущей социотехносферы превратилось в приоритетное исследовательское направление для современных учёных [139]. Ключевая задача — наладить синергию между ЕИ и ИИ, создавая киберфизические системы, которые не вытесняют, а усиливают когнитивные возможности человека, обеспечивая тем самым более ответственное, этикоцентричное и устойчивое применение цифровых решений в общественной практике.

Значимость подобного междисциплинарного взаимодействия трудно переоценить: полученные эмпирические данные и валидированные методики формируют платформу для дальнейших фундаментальных и прикладных исследований.

Глубокое освоение тонкостей и микродинамики взаимодействия этих двух разновидностей интеллекта откроет дополнительные перспективы для разработки продвинутых когнитивных архитектур, способных анализировать и агрегировать данные значительно эффективнее.

Клинические системы поддержки принятия решений, сконструированные на принципах синергии между естественным и искусственным интеллектом,

смогут формировать обоснованные рекомендации даже в условиях крайней клинической неопределённости, тем самым расширяя когнитивные возможности врача и обеспечивая более осмысленный выбор терапевтической стратегии [13].

Сегодня прогнозирование перспектив совместной эволюции ЕИ и ИИ приобретает ключевое значение как для фундаментальной науки, так и для прикладных разработок. Их грядущий симбиоз формирует новые парадигмы в сфере многослойной обработки, передачи и долговременного хранения информации. Предполагается, что ИИ, совершенствуясь с помощью машинного обучения и нейроморфных алгоритмов, будет тесно увязываться с биологическими когнитивными механизмами, открывая гибридные методики решения комплексных задач, включая роботизированную автоматизацию и продвинутую аналитику больших данных [140].

Формирование инновационных цифровых платформ и мультидисциплинарных кластеров становится ключевым фактором дальнейшей эволюции искусственного и эмоционального интеллекта. Скоординированное сотрудничество государства, корпоративного сектора и академического сообщества позволит выработать единые регламенты, протоколы безопасности и этические стандарты, обеспечивающие наилучшую интеграцию этих технологий в экономическую и социальную ткань [32]. В результате возрастёт системная устойчивость, отвечающая идеологии Общества 5.0, где технологические решения призваны одновременно повышать продуктивность и ускорять общественный прогресс, а также формировать ценностно-ориентированные модели взаимодействия человека и машины.

Перспективным вектором исследований становится системное рассмотрение узкоспециализированных и общеинтеллектуальных способностей искусственного интеллекта. Подобный фокус позволит детально вскрыть различия механизмов работы ИИ и биологического разума. Глубинный анализ этих расхождений открывает горизонты для развития более комплексных архитектур, способных осуществлять тонкую адаптацию и выбирать многовариантные модели взаимодействия с пользователем и динамичной средой [142].

С быстрым развитием технологий глубокого обучения и продвинутых методов обработки естественного языка становится реальным встраивание компонентов естественного интеллекта в архитектуры ИИ, что открывает принципиально новые сценарии их практического использования. Предполагается, что современные модели машинного обучения смогут автоматически выполнять задачи, которые раньше отнимали у специалистов огромное количество времени. Поскольку мир преобразуется ускоренными темпами, синергия ЕИ и ИИ позволит формировать более продуктивные системы хранения, управления и аналитики данных, критически необходимые для обоснованного принятия решений в разнообразных отраслях.

Ключевым вектором синергии ЕИ и ИИ становится формирование гибридных платформ, обеспечивающих высокоэффективное управление информационными ресурсами; многочисленные кейсы демонстрируют, как данные анализируются, структурируются и архивируются посредством когнитивных технологий и адаптивных алгоритмов.

В целом, синергия экспертного и ИИ с процессами хранения и анализа данных открывает новые возможности для ускорения операций, повышения точности и гибкости ИС.

Технологии искусственного и расширенного интеллекта вступают в критическую фазу эволюции, трансформируя методики хранения и управления данными в разных секторах. Их синергетическое применение ускоряет обработку массивов в дата-лейках и на edge-инфраструктуре, снижая углеродный след и сокращая операционные издержки при лавинообразном росте информации.

ЕИ как совокупность когнитивных способностей человека играет ключевую роль при интуитивном анализе данных и экспертной валидации. Благодаря эмпирическому опыту специалист связывает разрозненные датасеты, обнаруживая скрытые корреляции и редкие паттерны, ускользающие от моделей ML. В условиях, когда алгоритмы не формулируют однозначных выводов, именно ин-

туитивное суждение помогает принимать стратегические решения. Синергия человеческого паттерн-распознавания и вычислительной мощности обеспечивает более рациональное распределение ресурсов и повышение надежности архитектур хранения информации.

Опираясь на практический опыт оценки и картографирования углерода, фиксируемого лесными экосистемами [142], авторы отмечают, что ИИ, включая машинное обучение и анализ биг-дата, существенно автоматизирует процессы и улучшает управление информацией.

ИИ-платформы, опирающиеся на машинное обучение и нейронные сети, способны непрерывно усваивать поступающие данные, динамично адаптируясь к новым условиям, тем самым существенно оптимизируя временные и трудозатраты при информационной обработке.

Широкое внедрение предиктивных алгоритмов машинного обучения и продвинутой аналитики больших данных ускоряет формирование отказоустойчивых систем хранения с минимизированной вероятностью критических ошибок.

Тем не менее, внедрение ИИ в инфраструктуру хранения данных ограничивается эксплуатационными и регуляторными факторами.

Низкое качество исходных наборов данных, усложнённость алгоритмических моделей и дефицит объяснимости остаются ключевыми барьерами для массового внедрения ИИ в данной области. Подобные ограничения способны спровоцировать масштабные сбои и падение производительности, вследствие чего вопрос о синергетической конкуренции и сотрудничестве между естественным и искусственным интеллектом становится ещё более насущным.

Междисциплинарная парадигма, объединяющая экспертные компетенции, творческую интуицию и аналитический аппарат ИИ, формирует предпосылки для появления гибридных, высокоадаптивных систем хранения данных. Слаженная интеграция ЕИ и ИИ трансформирует не только технические алгоритмы бэкап-архитектур, но и матрицу корпоративного управления информациями, где

эти решения разворачиваются. Такая синергия формирует плацдарм для прорывных научных исследований и масштабируемых индустриальных внедрений, способных радикально изменить весь цикл обращения, защиты и интеллектуальной эксплуатации данных.

Комплексное изучение данной области критически важно для формирования специалистов нового поколения, владеющих ключевыми компетенциями и глубокими знаниями в сфере передовых систем хранения данных, включая облачные репозитории и распределённые архитектуры. Перспективы исследований сулят прорывные открытия и инновации, обеспечивающие устойчивое, энергоэффективное управление информационными потоками в стремительно трансформирующейся цифровой экосистеме.

Стратегии хранения данных выступают ключевым фактором, обеспечивающим оперативную обработку информации и оптимальное принятие управленческих решений.

С внедрением ЕИ и ИИ формируются передовые методы структурирования, трансформации и защиты информационных ресурсов. К ключевым инструментам относятся классические реляционные СУБД, решения на базе PostgreSQL, распределённые реестры блокчейна для децентрализованного хранения, а также высокопроизводительная СУБД Adabas, не ограниченная по объёму хранимых данных [142].

Реляционные СУБД базируются на строгих табличных структурах и идеальны, когда требуется чётко определённая, неизменная схема данных. Тем не менее, их эффективность снижается, когда приходится часто изменять модель или горизонтально масштабировать систему при росте объёмов. Интеграция ИИ позволяет интеллектуально управлять СУБД: автоматизировать миграции, адаптивно перестраивать схемы и обеспечивать обработку информации в режиме реального времени [143].

СУБД PostgreSQL характеризуется расширяемой архитектурой и поддержкой форматов JSONB, HSTORE, что придаёт ей высокую адаптивность и делает

удобной при работе с полу- и неструктурированными данными. В связке с алгоритмами машинного обучения можно оперативно обрабатывать массивные датасеты, выявлять латентные взаимосвязи и повышать точность аналитики. Минусы: сложность интеграции в устоявшуюся инфраструктуру и требование узкопрофильных компетенций администраторов и дата-инженеров.

Использование блокчейн-технологии при хранении данных фиксирует записи в защищённом и неизменяемом виде, тем самым обеспечивая высокий уровень безопасности.

У каждого подхода существует собственный набор сильных и слабых сторон, обусловленных спецификой бизнес-требований и практического сценария. Поэтому ставится в приоритет компетентный выбор технологического стека и его гибкая адаптация под стратегические цели компании. Существенное значение приобретает корректная оценка потребностей в обработке и интеграции данных для оптимизации аналитических пайплайнов и процессов управления информацией.

Интеграция ИИ-ориентированных решений для хранения и аналитики данных подразумевает детальное владение вопросами архитектуры, кибербезопасности и нормативного регулирования, что для многих компаний превращается в существенный вызов. Реальные результаты достигаются лишь при выработке прозрачного Roadmap, увязывающего корпоративные процессы, модели управления данными и выбор облачных или локальных платформ с отраслевой спецификой.

На практике ИИ всё чаще задействуют для совершенствования систем хранения и обработки данных во множестве отраслей. Так, в сфере логистики технологии машинного обучения помогают не только оптимизировать построение маршрутов, но и заметно сократить углеродный след перевозок. Эксперименты демонстрируют, что аналитика big data позволяет алгоритмам выявлять более устойчивые, ресурсоэффективные транспортные схемы.

Синергия алгоритмов машинного обучения с системами обработки Big Data выводит автоматизацию бизнес-процессов на новый уровень точности и фокусировки, демонстрируя образцовый баланс технологий и стратегических целей компании.

Интеграция ИИ в инфраструктуру хранения данных существенно повышает управляемость и кибербезопасность корпоративной информации. Модели машинного обучения, анализируя поведенческие паттерны пользователей и систем, оперативно выявляют аномалии, прогнозируют потенциальные инциденты и помогают снизить совокупные риски. Параллельно реализация реляционных практик, включая нормализацию схем БД, предотвращает избыточность и повышает скорость доступа. Такая синергия особенно востребована на фоне стремительного роста объёмов данных, генерируемых современными организациями ежегодно.

Информация породила ЭВМ: их внедрение признано повсеместно, и отменить этот триумф человеческого технического разума невозможно.

Согласно данным аналитической компании Gartner (американского исследовательского центра, специализирующегося на анализе мирового ИТ-рынка), на сегодняшний день по всему миру функционирует:

– свыше 2 млрд персональных ПК по всему миру;

– в мире свыше 14 млрд мобильных устройств; к середине 2025-го аналитики прогнозируют 18,22 млрд;

– сейчас в мире около 1000 публичных суперкомпьютеров;

По диапазону уникальных IPv4-адресов (при допущении «1 сервер = 1 IP») потенциальное число серверов оценивается в ~4,2 млрд.

Сегодня десятки миллиардов ЭВМ повсеместно продуцируют сотни, а то и тысячи, миллиардов информационных потоков.

Объём глобального цифрового контента растёт экспоненциально: ежедневно человечество производит около 328,77 млн терабайт данных, пополняя распределённые облачные репозитории и дата-центры [144].

Это действительно сфера Big Data! Пока экосистема больших данных ещё не располагает полнофункциональными решениями для их сбора, хранения, обработки и аналитики, поэтому основное внимание и ожидания связаны с ИИ.

Дополнительным существенным фактором является углублённая автоматизация операций, позволяющая высвободить персонал от монотонных действий. Развёртывание интеллектуальных систем не только повышает общую производительность, но и открывает сотрудникам пространство для решения комплексных задач, требующих творческого и стратегического подхода. Таким образом, внедрение ИИ в традиционные бизнес-процессы трансформирует не только механизмы хранения данных, но и организационную архитектуру компаний.

В конечном итоге внедрение ИИ в процессы управления и хранения данных становится стратегически значимым этапом для организаций, ориентированных на рост операционной эффективности и усиление информационной безопасности. Практические примеры показывают, что AI-алгоритмы существенно оптимизируют архитектуру хранения, повышая отказоустойчивость и защищённость от киберугроз, превращая эти характеристики в весомое конкурентное преимущество на рынке.

ИИ выступает ключевым технологическим драйвером в сфере управления, хранения и аналитической обработки данных. Интеграция интеллектуальных алгоритмов способна радикально повысить производительность существующих ИТ-инфраструктур, автоматизируя повторяющиеся операции и высвобождая специалистам время для решения приоритетных, стратегически значимых бизнес-вопросов. Благодаря такой интеллектуальной автоматизации ускоряется обработка массивов информации и растёт общая операционная эффективность.

Одно из главных достоинств интеграции ИИ в инфраструктуру хранения заключается в более глубокой аналитической обработке информации. Продвинутые AI-алгоритмы умеют извлекать закономерности, тренды и аномалии из разнородных и неструктурированных массивов данных, что особенно важно при экспоненциальном росте их объёмов. Модели машинного обучения не только выявляют скрытые взаимосвязи, но и строят прогнозы, позволяя бизнесу обоснованно планировать стратегии, оптимизировать процессы и открывать новые рыночные ниши.

Персонализированный клиентский опыт сегодня превращается в заметное конкурентное преимущество. Алгоритмы машинного обучения и аналитика больших данных позволяют детально исследовать поведенческие паттерны аудитории, сегментировать пользователей и в режиме реального времени генерировать релевантные рекомендации, настраивая сервисы и контент под конкретные потребности и цели каждого клиента. Точечная кастомизация повышает удовлетворённость, усиливает эмоциональную привязанность к бренду и, как следствие, положительно влияет на ключевые финансовые показатели компании.

Следует подчеркнуть, что внедрение систем ИИ способно значительно повысить уровень обоснованности управленческих решений, практически исключив фактор человеческой ошибки. Практика показывает: алгоритмы машинного обучения и методы predictive analytics обрабатывают и интерпретируют массивы больших данных точнее человека, обеспечивая более взвешенные и качественные выводы.

Помимо прочего, внедрение технологий ИИ в корпоративные информационные системы обладает существенным экономическим эффектом. Глубокая автоматизация рабочих потоков сокращает издержки на персонал, оптимизирует операционные расходы и укрепляет финансовую устойчивость предприятий. Прогнозируется, что масштабирование ИИ-решений повысит рентабельность и обеспечит конкурентное преимущество в ближайшей перспективе, что многие компании уже рассматривают как ключевую стратегию.

Несмотря на все вышеперечисленные преимущества, необходимо помнить о существующих ограничениях технологий, которые будут доступны для обсуждения в дальнейшем. Важно осознавать как возможности, так и вызовы, связанные с внедрением ИИ в информационные системы для достижения максимального эффекта от его использования.

Интеграция технологий ИИ в цепочку хранения больших данных и их аналитической обработки сталкивается с множеством ограничений, способных существенно снизить производственную эффективность инфраструктуры.

Ключевая трудность заключается в том, что искусственному интеллекту приходится надёжно обрабатывать и интерпретировать массивы разнотипных

данных. Обученные нейросетевые модели нередко теряют точность при гетерогенных входных потоках, поэтому необходимо пересматривать стратегии обучения, включая регуляризацию, аугментацию и контрфактический анализ.

Существенным барьером остаётся необходимость в достоверных, высококачественных датасетах для обучения алгоритмов. При наличии ошибок, дисбалансов или скрытой предвзятости обучающие выборки порождают искажённые выводы и прогнозы, снижая доверие к ИТ-системам на этапе промышленной эксплуатации. Переход к более универсальным, обобщающим архитектурам требует одновременного решения вопросов доступа к конфиденциальным данным, соблюдения норм приватности и согласования множества организационных и технических факторов, что дополнительно усложняет процесс.

Следует подчеркнуть, что внедрение ИИ в инфраструктуру хранения данных предполагает существенные вычислительные мощности, расширенные сетевые ресурсы и высокие издержки на тренинг, калибровку и поддержку ML-моделей. Подобные требования нередко становятся барьером для малых предприятий, которым трудно зарезервировать бюджет и временные ресурсы на научно-исследовательские и пилотные этапы. Проектирование жизнеспособных алгоритмов требует междисциплинарности, объединяющей технические метрики и принципы этического комплаенса.

Современный рынок труда трансформируется под влиянием масштабной автоматизации бизнес-процессов, что нередко провоцирует дефицит профильных экспертов в ряде высокотехнологичных ниш. Такой кадровый вакуум тормозит эволюцию digital-инфраструктуры и обостряет борьбу компаний за talent pool, напрямую отражаясь на качестве и темпах интеграции ИИ в корпоративные контуры.

Важно учитывать, что продуктивность ИИ напрямую обусловлена пропускной способностью современных GPU-кластеров и другой высокопроизводительной вычислительной инфраструктуры. Аппаратные ограничения, включая дефицит видеопамяти и недостаточную параллелизацию, способны заметно тормозить внедрение ИИ, особенно при ресурсоёмких нейросетевых вычислениях. Дополнительно архитектурные ограничения алгоритмов, их обучаемость и адаптивность требуют непрерывной валидации, оптимизации и разработки инновационных подходов.

Российские исследователи разработали уникальную, не имеющую мировых аналогов технологию производства процессоров следующего поколения. Основой служит прорывной метод интегрального формирования логических узлов с субатомной точностью 0,2 Å (0,02 нм), что сопоставимо с размером отдельных межатомных связей. Подробное описание представлено в статье Smirnov и соавт. в журнале Science Advances (Sci. Adv. 11, eads9744, 7 мая 2025). Решение уже защищено отечественным патентом, сейчас идёт международная заявочная кампания. Ожидается, что инновация увеличит производительность ЭВМ на порядки.

Для продуктивного преодоления обозначенных вызовов следует внедрять более детализированные регламентации и проводить комплексные исследования, учитывая непрерывную эволюцию цифровых технологий и трансформацию рыночных запросов. Это диктует потребность в создании гибко-адаптивных моделей, способных интегрироваться в мультисекторную среду при контролируемых рисках и операционных ограничениях. Дополнительно требуется углублённый анализ, фокусированный на совершенствовании архитектур ИИ-систем и увеличении их отказоустойчивости.

В процессе исследования взаимодействия ЕИ и ИИ установлено, что обе когнитивные системы обладают особыми достоинствами, способными существенно повысить эффективность управления данными. ЕИ, основанный на человеческой интуиции, эмпирическом опыте и способности к дедукции, обеспечивает глубокое понимание контекста, скрытых взаимосвязей и смысловых оттенков, которые нередко теряются при автоматизации. Искусственный интеллект, напротив, демонстрирует исключительную продуктивность в скоростной обработке массивов данных, выявлении статистических паттернов и бесшовной автоматизации повторяющихся процедур.

Комплексное исследование систем хранения информации выявило, что классические реляционные БД постепенно сдают позиции, уступая инновационным платформам — таким как PostgreSQL, Adabas, а также облачным сервисам.

Современные архитектуры хранения данных, включая распределённые базы и облачные репозитории, предоставляют гораздо более адаптивные и ресурсосберегающие механизмы управления информационными потоками, что

критично в эру стремительных изменений. Синергия ЕИ и ИИ формирует основу гибридных платформ, объединяющих когнитивные преимущества человека с вычислительной мощью машин для улучшенного принятия решений.

С каждым годом алгоритмы ИИ становятся всё более продвинутыми: теперь они способны не просто оперативно обрабатывать массивы данных, но и самостоятельно обучаться, извлекая закономерности из накопленного опыта. Такое самообучение открывает путь к созданию адаптивных вычислительных платформ, которые тонко реагируют на изменения внешних параметров среды и формируют решения, основанные на углублённой аналитике Big Data. Показательные кейс-стади демонстрируют, что грамотная интеграция естественного и искусственного интеллекта, построенная на принципах когнитивной синергии, обеспечивает существенные прорывы в медицине, финтехе и иных областях.

Тем не менее, при всех очевидных плюсах внедрения ИИ в информационно-технические платформы остаются и существенные барьеры. На передний план выходят аспекты цифровой этики, кибербезопасности и интерпретируемости моделей. Следует помнить, что чрезмерная автоматизация бизнес-процессов способна обесценить человеческий надзор и понимание логики решений, что порождает скепсис и утрату доверия к умным приложениям. По этой причине необходимо системно разрабатывать, нормативно закреплять и внедрять этические кодексы и стандарты, которые бы четко регламентировали эксплуатацию ИИ в разных отраслях.

ЕИ и ИИ, выступающие как высокоэффективные средства накопления, структурирования и скоростной обработки массивов данных, обладают потенциалом радикально повысить качество и обоснованность управленческих и аналитических решений. Синергия между этими формами интеллекта открывает целый спектр возможностей для углубленной оптимизации информационных потоков, однако одновременно обостряет вопросы цифровой этики, конфиденциальности и кибербезопасности. Перспективы эволюции ИС во многом будут зависеть от успешности комплексной интеграции этих технологий, способной обеспечить более устойчивые, точные и безопасные механизмы управления данными.

ГЛАВА 6. ВЛИЯНИЕ ТЕХНОЛОГИЙ BIG DATA И АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ НА IT-РЕШЕНИЯ И КОРПОРАТИВНЫЕ БИЗНЕС-ПРОЦЕССЫ

Сегодня, когда данные превращаются в ключевой стратегический актив, концепции Big Data и машинного обучения (ML) выходят на передний план. Современные алгоритмы обработки массивов информации и инструменты предиктивной аналитики радикально трансформируют методы анализа, открывая путь к инновационным бизнес-моделям и повышая операционную эффективность, гибкость и конкурентоспособность компаний. Экспоненциальный прирост объёмов создаваемых ежедневно данных заставляет организации внедрять масштабируемые облачные хранилища, технологии распределённых вычислений и интеллектуальные модели извлечения знаний, что делает глубокое освоение Big Data и ML критически важным.

Big Data — это масштабные, разнородные и стремительно пополняющиеся массивы информации, с которыми традиционные СУБД уже не справляются. Ключевые параметры, описываемые парадигмой 3V (Volume, Velocity, Variety) и дополняемые концепциями Veracity и Value, превращают такие данные в стратегический актив для здравоохранения, финансов, ритейла и научных исследований. Современные системы распределённой обработки — Hadoop, Spark, NoSQL-хранилища — помогают выявлять скрытые зависимости, поддерживая data-driven принятие решений и повышая эффективность бизнес-операций.

Машинное обучение представляет собой специализированную ветвь искусственного интеллекта, сосредоточенную на разработке математических моделей и алгоритмов обработки данных, благодаря которым системы способны автономно выявлять закономерности и формировать прогнозы, опираясь на входные наборы признаков. К числу базовых методов относятся линейная регрессия, алгоритмы дерева решений, а также их ансамблевое расширение — случайный лес; эти инструменты применяются для задач классификации, регрессионного прогнозиро-

вания, кластеризации и глубокой аналитики больших массивов информации. Подобные модели востребованы не только в корпоративной среде для оптимизации бизнес-процессов, но и в сфере здравоохранения, где они повышают точность диагностических выводов и ускоряют обнаружение заболеваний [150].

По мере стремительного роста массивов данных ключевым фактором успеха становится не только их хранение, но и высокая скорость обработки. Оптимизация аналитических процедур и внедрение автоматизированных алгоритмов Big Data позволяют рационально задействовать вычислительные ресурсы и оперативно извлекать ценные инсайты. В динамичной рыночной среде компании, способные выполнять real-time аналитику на масштабируемой инфраструктуре, получают весомое конкурентное преимущество.

Следует подчеркнуть, что интеграция методов машинного обучения с экосистемой больших данных расширяет горизонты аналитики. Современные нейронные сети и другие алгоритмы глубокого обучения обнаруживают многомерные паттерны и скрытые корреляции, недоступные традиционным статистическим подходам. В результате растёт точность предиктивной аналитики, позволяя компаниям оперативнее управлять рисками и реагировать на динамику рынка.

Помимо этого, прорывные цифровые решения и прогрессивные методы бизнес-аналитики на базе big data задают свежие векторы развития компаний, обязывая их непрерывно мониторить рыночную конъюнктуру и динамику потребительских предпочтений. Создание и реализация гибких стратегий, опирающихся на продвинутые алгоритмы машинного обучения и предиктивной аналитики, уже стали критически важным условием конкурентоспособности. Организации, активно инвестирующие в искусственный интеллект и связанные платформы, не только оптимизируют операционные цепочки, но и формируют дополнительную ценность для клиентов и конечных пользователей [150].

Начиная с 2023 года наблюдается стремительный прогресс в областях машинного обучения и искусственного интеллекта, что открывает для бизнеса, промышленности и академического сообщества ещё более широкие горизонты. Ключевые направления включают тотальную роботизацию процессов, углублённую аналитическую обработку данных, а также повсеместное внедрение само-

обучающихся и адаптивных алгоритмов. Передовые подходы, в частности глубокие нейронные сети и трансформер-архитектуры, уже активно применяются в здравоохранении, финтехе, логистических цепочках и smart-manufacturing, значительно повышая операционную эффективность. Одновременно возрастающее внимание к работе с Big Data помогает принимать обоснованные, предиктивные управленческие решения.

Само настраиваемые алгоритмы, способные в режиме реального времени корректировать гиперпараметры, проводить онлайн-оптимизацию, заметно расширяют практический арсенал машинного обучения. По сравнению с классическими методами, они точнее прогнозируют, адаптируясь к концепт-дрифту и нестационарным данным внешней среды. Подобная гибкость критична при управлении бизнес-процессами и оптимизации KPI в условиях высокой неопределённости.

Возросшая доступность разнородных датасетов стала фундаментальным фактором эволюции машинного обучения. Современные методики агрегации, хранения и быстрой обработки информации дают компаниям возможность задействовать продвинутую аналитику для более детального прогнозирования клиентского поведения и совершенствования сервисов. Однако внедрение AI-инструментов в корпоративные процессы — это не только IT-инициатива, но и необходимость перестройки управленческой структуры и культурных практик.

Сегодня ключевой акцент смещён к практическому внедрению передовых решений при создании автономных систем. Так, алгоритмы машинного обучения и компьютерного зрения уже внедряются в транспорт для разработки беспилотных автомобилей и дронов, что в ближайшие годы трансформирует логистику и персональную мобильность. При этом траектория развития неотделима от вопросов кибербезопасности, отказоустойчивости и способности подобных платформ динамически адаптироваться к широкому спектру дорожных и погодных сценариев.

Слабые стороны классических алгоритмов машинного обучения становятся всё очевиднее: переобучение, смещение выборки и drift данных усиливают

разрыв между ожиданиями и реальной производительностью. Поэтому компаниям, внедряющим AI-решения, необходимо заранее оценивать риски и формировать стратегии аудита, мониторинга и смягчения негативных эффектов.

Ключевые тенденции 2023 года подчёркивают стремительный рост применения предиктивной аналитики и технологий машинного обучения для data-driven прогнозирования спроса. Эти инструменты позволяют точнее выявлять потребительские предпочтения, оптимизировать планирование ресурсов, управление складскими запасами и логистическими цепочками. Практика показала: компании, внедрившие такие решения, уже укрепили позиции и повысили конкурентоспособность благодаря более достоверным прогнозам.

Глубокое понимание актуальных трендов позволит нам полнее оценить удачные кейсы внедрения цифровых технологий в бизнес-процессы.

Методы машинного обучения активно внедряются в самые разные отрасли — от финансов до медицины — позволяя автоматизировать анализ массивов данных и строить прогнозы. С точки зрения типа обучения принято выделять три большие группы моделей: supervised learning, unsupervised learning и reinforcement learning. В supervised learning используются маркированные выборки, где алгоритм сопоставляет входные векторы признаков с известными целевыми метками. Классическим представителем данного подхода служит модель линейной регрессии, применяемая для количественного прогнозирования на основании набора значимых факторов.

Методы обучения без учителя, наоборот, ориентированы на обнаружение латентных структур в сырых выборках без наличия меток. Самой распространённой техникой здесь считается кластеризация, например K-means: алгоритм группирует объекты с близкими признаковыми векторами. В бизнес-аналитике его часто применяют для поведенческой сегментации клиентов.

Методы обучения с частичным контролем объединяют достоинства супервизионного и несупервизионного подходов, позволяя одновременно задействовать размеченные и сырые выборки. Так, модели типа случайного леса успешно

решают задачи классификации и регрессии, демонстрируя высокую устойчивость, масштабируемость и точность при анализе больших корпоративных массивов данных и бизнес-процессов.

В последние годы наблюдается устойчивый рост интереса к гибридным и ансамблевым алгоритмам машинного обучения. Градиентный бустинг, реализующий идею итеративного построения слабых базовых моделей, зарекомендовал себя как мощный механизм повышения точности и снижения дисперсии прогнозов. Особенно востребован он в финансовых задачах, где минимизация рисковых экспозиций и борьба с переобучением критичны [146].

Современные нейросетевые архитектуры, включая методы глубокого обучения, служат высокоэффективными инструментами для обработки массивных и высокоразмерных датасетов, что открывает возможность создания гибких, самообучающихся систем, функционирующих в режиме реального времени. Их внедрение в компьютерное зрение и акустическое распознавание демонстрирует практическую интеграцию машинного обучения с концепцией Big Data.

Разнообразие алгоритмов и их высокая адаптивность создают расширенные перспективы для data mining и аналитики, а их практическое внедрение открывает свежие горизонты цифровой трансформации бизнес-процессов.

Практическое внедрение технологий Big Data и алгоритмов машинного обучения в бизнес-решениях уже доказало свою результативность, охватывая спектр отраслей — от финансов до здравоохранения. Передовые компании применяют продвинутую аналитику и предиктивное моделирование для оптимизации процессов, повышения эффективности и разработки инновационных продуктов, подтверждая стратегическую ценность этих инструментов.

В современной рознице анализ big data о вкусах и поведении покупателей дает возможность не только прогнозировать колебания спроса, но и конструировать персонализированные офферы, повышая качество сервиса и укрепляя клиентскую лояльность. Применяя алгоритмы машинного обучения и методы пре-

дикативной аналитики, компании оперативно перерабатывают массивы информации, чтобы выявлять наиболее результативные sales-стратегии, уточнять сегментацию аудитории и подстраивать маркетинговые кампании под актуальные рыночные тренды [145].

В промышленном сегменте технологии Big Data служат ключевым инструментом для глубокой оптимизации цеховых процессов. Когортный анализ и машинное обучение помогают сокращать издержки, повышать OEE и тем самым поднимать суммарную производительность. Например, предиктивная аналитика, опираясь на данные IoT-сенсоров, заранее выделяет потенциальные узкие места, предотвращая простои линии.

Сегодня банки и иные финансовые организации всё активнее интегрируют инструменты Big Data и машинного обучения в процедуры кредитного скоринга. Глубокая обработка массивов поведенческой и транзакционной информации позволяет совершенствовать риск-менеджмент: снижать вероятность дефолтов, оперативно выявлять фрод и недобросовестных контрагентов [146], одновременно ускоряя и точнее сегментируя привлечение надёжных заёмщиков.

В современном маркетинге компании задействуют Big Data для проектирования, персонализации и последующей оценки рекламных активностей. Глубокая аналитика массивов информации выявляет офферы, которые лучше всего резонируют с целевой аудиторией, тем самым повышая ROI, позволяя точечно распределять медиабюджет и стимулировать конверсию. Постоянная data-driven адаптация стратегий приводит к ощутимому бизнес-росту.

Интеграция технологий Big Data и machine learning предоставляет компаниям обширный спектр возможностей: от data-driven улучшения управленческих решений и точного прогнозирования рыночных трендов до автоматизированной оптимизации внутренних бизнес-процессов и ресурсов.

Широкий доступ к массивам Big Data дает компаниям возможность глубже исследовать операционные процессы и клиентское поведение. Такой продвинутый аналитический подход позволяет формировать решения, опирающиеся на

объективные метрики и прогнозные модели, а не на субъективные предположения. Как следствие, совершенствуются маркетинговые и коммерческие стратегии, повышается качество клиентского опыта и масштабируется выручка.

Прогнозная аналитика, основанная на методах машинного обучения, дает возможность обнаруживать скрытые тренды и закономерности, недоступные традиционному человеческому анализу. Интегрируя разнородные датасеты — от транзакционных логов до соцмедийных метрик, компании способны заранее распознавать сдвиги в потребительских предпочтениях и оперативно корректировать продуктовую линейку под актуальные рыночные условия. Так, поведенческий анализ аудитории подсказывает, какие товары нужно масштабировать, какие улучшать, а какие рационально вывести из ассортимента.

Оптимизация бизнес-процессов остаётся ключевым приоритетом. Алгоритмы машинного обучения и инструменты predictive analytics детально диагностируют узкие места цепочек поставок и сервисных операций, позволяя радикально сокращать операционные издержки. Одновременно RPA и Big Data автоматизируют повторяемые задачи, перераспределяя человеческий капитал на инновационные инициативы и увеличивая совокупную эффективность.

Благодаря встроенной способности алгоритмов машинного обучения самообучаться на потоках входящих данных, компании получают инструмент для оперативного реагирования на рыночные колебания и внедрения гибкого менеджмента. Углублённая аналитика Big Data способствует совершенствованию производственных линий и логистических цепочек, тем самым укрепляя их конкурентные позиции.

При всей очевидной пользе инноваций важно помнить, что их внедрение сопряжено с масштабными капитальными вложениями, а порой и с полным обновлением инфраструктуры. Трудности при интеграции цифровых решений и реорганизации бизнес-процессов нередко провоцируют кадры на сопротивление переменам.

Технологические решения в области Big Data и machine learning открывают бизнесу принципиально новые возможности для аналитики и построения data-driven стратегий, однако процесс их интеграции сопряжён с внушительными препятствиями. Ключевой фактор риска — высокая цена систем хранения, ETL-конвейеров и вычислительных кластеров, поэтому капитальные вложения в инфраструктуру должны быть обоснованы детальной экономической и TCO-экспертизой до старта инициативы.

Вопрос информационной безопасности ныне вынесен на передний план, тем более в отношении персональных данных, строго регулируемых профильными законами и отраслевыми стандартами. Сохранение целостности, доступности и конфиденциальности сведений обязывает компании внедрять комплексные системы киберзащиты и регулярно проводить аудит рисков. Аналитика инцидентов показывает: провалы грозят существенными финансовыми потерями, ущербом бренду и утратой клиентского доверия.

Дефицит квалифицированных кадров остаётся серьёзным вызовом для бизнеса. Современные модели машинного обучения, особенно в сфере глубоких нейронных сетей и анализа Big Data, требуют инженеров с фундаментальными знаниями математики, вероятностных методов и программирования. Однако рынок труда ощущает заметный недостаток таких data-scientists, что тормозит внедрение ИИ-проектов. Университетские программы не успевают оперативно перестраиваться под динамику отрасли, усугубляя разрыв.

Отдельного внимания заслуживают этические вызовы, сопровождающие интеграцию Big Data и методов машинного обучения. Применение слабо контролируемых алгоритмов и генеративных моделей способно усилить дискриминационный bias при автоматизированном принятии решений — особенно там, где затрагиваются здоровье, безопасность и человеческое достоинство. Ошибочные или некорректно интерпретированные выводы predictive analytics способны напрямую затронуть фундаментальные права и качество жизни граждан.

Помимо этого, следует учитывать культурно-организационную специфику каждой компании, поскольку внедрение цифровых решений нередко сталкивается с сопротивлением персонала. Чтобы нововведения прижились, требуются грамотный change-management, системное повышение квалификации сотрудников и соответствующие затраты времени и ресурсов.

Устранение обозначенных вопросов станет критической предпосылкой для эффективного внедрения проектов.

По мнению специалистов, симбиоз IoT-экосистем и Big Data приведёт к появлению ещё более интеллектуальных и производительных платформ аналитики. Сенсоры и смарт-устройства станут непрерывно формировать терабайты потоковых данных, обработка которых потребует edge-computing, data-lakes и глубоких алгоритмов машинного обучения. Такое вычислительное ядро позволит бизнесу точнее сегментировать аудиторию, выявлять скрытые паттерны поведения и с высокой точностью прогнозировать рыночную динамику.

В ближайшие годы всё больше организаций будет интегрировать искусственный интеллект в корпоративные экосистемы, чтобы глубже автоматизировать и совершенствовать операционные процессы. Платформы AI и алгоритмы машинного обучения дадут возможность не только оперативно обрабатывать массивы big data, но и обнаруживать скрытые паттерны и корреляции. Это критически значимо для отраслей здравоохранения, финансов, логистики и индустриального производства. Быстрая цифровая адаптация обеспечит бизнесу устойчивую конкурентоспособность и позволит формировать решения на базе продвинутой аналитики и прогнозных моделей.

Недостаточная продуктивность классических алгоритмов дополнительно стимулирует развитие квантовых вычислений, способных радикально ускорить pipeline обработки данных. Квантовые процессоры обещают перевернуть парадигму анализа Big Data, предоставляя ресурсы для работы с огромными массивами в экспоненциально сжатые сроки. Синергия квантового машинного обучения и действующих Big Data-платформ откроет широкие перспективы для фундаментальной науки, бизнес-аналитики и предиктивного моделирования.

В перспективе ключевой фокус, вероятно, сместится к ответственному обращению с данными. Всё больше компаний будут интегрировать принцип алгоритмической прозрачности, демонстрируя, как формируются выводы и рекомендации. Подобная открытость укрепит клиентское доверие, углубит взаимоотношения и станет неотъемлемым элементом data-governance, превращаясь в существенный конкурентный дифференциатор.

Комплексный обзор концепций Big Data и машинного обучения подтвердил, что эти технологические домены не просто взаимодополняемы, а формируют мощный синергетический тандем, обеспечивающий компаниям извлечение стратегически значимых инсайтов из экстремально больших массивов разнотипных данных. Текущее состояние индустрии и ключевые тренды ML демонстрируют ускоренную эволюцию моделей и методик — от глубоких нейронных сетей до самообучающихся систем, — чья растущая сложность и точность расширяют возможности аналитики и прогнозирования, позволяя бизнесу принимать data-driven решения с минимизацией риска.

Авторы, исследовав спектр алгоритмов машинного обучения, выявили целый арсенал методик, каждая из которых обладает собственными сильными сторонами и ограничениями. Систематизация решений — регрессий, деревьев решений, ансамблевых методов, нейронных сетей и других техник — помогает глубже понять, какие модели оптимальны для конкретных бизнес-кейсов. Интеграция Big Data, продвинутой аналитики и машинного обучения в операционные процессы, как показывают кейс-стади, существенно повышает производительность, сокращает издержки и улучшает клиентский опыт.

Выигрыши от интеграции таких цифровых решений очевидны: начиная с более точного прогнозирования спроса благодаря алгоритмам машинного обучения и заканчивая повышением потребительского опыта через персонализированные сервисы. Тем не менее остаются и слабые места: дефицит data-science специалистов, риски кибербезопасности, а также юридические и этические коллизии, связанные с обработкой персональных данных. Для нивелирования угроз необходима системная стратегия управления изменениями.

Перспективы Big Data и machine learning остаются весьма вдохновляющими. По мере эволюции искусственного интеллекта, IoT и облачно-нативных платформ спектр инструментов для продвинутой аналитики, data lakes и предиктивного моделирования будет шириться. Компании, быстро внедряющие такие решения и развивающие культуру data-driven, получат ощутимое конкурентное преимущество.

Следовательно, Big Data и алгоритмы машинного и глубокого обучения – это уже не просто модные концепции, а ключевые рычаги цифровой трансформации, задающие вектор развития бизнеса и ИТ-сферы: их внедрение в корпоративные процессы раскрывает возможности предиктивной аналитики, инноваций и устойчивого роста.

ЗАКЛЮЧЕНИЕ

Область исследования авторов была определена на рис. 1, но главным было – естественный и искусственный интеллект.

Наиболее подходящим для понимания, что такое ЕИ, интеллект, интеллектуальная деятельность, разум, сознание является исследование И. Канта «Критика чистого разума». Общее определение интеллекта И. Канта представлено на рис. 2. Важнейшим в концепции И. Канта является то, что интеллект, интеллектуальная деятельность, разум и сознание существуют и определяются не только опытом, а в основном априорными знаниями.

Концепция априорных знаний И. Канта имеет значительную актуальность для современных систем ИИ особенно в контексте: ограничений систем чисто индуктивного обучения; необходимости структурных ограничений для эффективного обучения; стремления к созданию ИИ с причинным мышлением и абстрактным пониманием реальности. Исследования авторов показали, что многие современные исследования в области ИИ фактически возвращаются к кантианским идеям о необходимости априорных структур для осмысления опыта, хотя и в технологически переосмысленном виде.

Авторы считают, что интеллект, интеллектуальная деятельность, разум и сознание проявляются и определяются головным мозгом человека и средой его функционирования. Авторам не удалось найти научно-практически доказанное определение, его структура функционирования, как из материальных компонент, образующих мозг, возникает, позиционируется и функционирует интеллект, интеллектуальная деятельность, разум и сознание.

Для авторов является бесспорным что в организме человека основным в позиционирование и поддержание функционируя интеллекта, интеллектуальной деятельности, разума и сознания является его головной мозг. Авторы используют доказанные достижения биологов и физиологов о содержание в головном мозге нейронов и нейронной сети, а также, что нейроны функционирую парал-

лельно, непрерывно, с постоянным химизмом в каждом нейроне и возникающими электронными импульсами. Каждый нейрон — это биологическая система и ней действуют биологические законы известные на текущий момент времени. В настоящее время не доказано, что мозг человека исполняет какой-то алгоритм, т. е. мозг человека не алгоритмичен.

Авторы показали, что при создании искусственных нейронных сетей и систем ИИ инициаторы значительно упростили представление биологических нейронов, что привело к ряду существенных ограничений: биологические нейроны обрабатывают информацию во временной области (спайки), в то время как большинство искусственных нейронов оперируют усредненными значениями; биологические нейроны имеют сложные дендритные деревья, способные выполнять нелинейные вычисления, а искусственные нейроны обычно сводят этот процесс к простому суммированию с весами; реальные нейроны имеют сложные нелинейные характеристики активации, зависящие от множества факторов; мозг человека потребляет около 20 Вт, тогда как системы ИИ требуют гигаватты энергии для подобных задач; современные ИИ находят корреляции в данных, но не понимают причинности; существующие системы ИИ генерируют новое содержание на основе статистических закономерностей, а не путем подлинного понимания.

Авторы считают, что эти недостатки указывают на необходимость разработки более биологически достоверных моделей нейронов и нейронных сетей для преодоления текущих ограничений систем ИИ.

В настоящее время в мире происходит ускоренное внедрение технологических решений, разработанных на основе искусственного интеллекта, в различные отрасли экономики и сферы общественных отношений.

Начиная с 1956 г. под ИИ понимались программы для ЭВМ, моделирующие деятельность мозга человека. Появилась цифровая модель нейрона мозга человека и искусственная нейронная сеть цифровых нейроном моделирующих мозг человека. Авторы показали достаточно подробную историю развития ИИ.

Первой проблемой с появление стало определение ИИ. Авторы собрали более 20 определения ИИ. В подавляющем множестве определений отсутствуют

ЭВМ, программы для ЭВМ, данные, это то, что и определяет ИИ и его применение. Авторы считают, что самое подходящее определение, присутствует в «Национальной стратегию развития искусственного интеллекта на период до 2030 год.», оно полностью приведено в начале работы.

Авторы провели анализ созданных и внедренных технологий ИИ и определили фундаментальные проблемы современного ИИ. Устранение их путь к созданию сильного искусственного интеллекта.

В целом проведенные исследования позволяет сделать ряд важных выводов относительно роли естественного и искусственного интеллекта в преобразовании данных.

Естественный интеллект сохраняет свои уникальные преимущества, главным из которых является творческая составляющая и способность к эмоциональному восприятию. Человеческий разум обладает способностью к социальному взаимодействию и интуитивному принятию решений, что пока недоступно искусственным системам.

Искусственный интеллект демонстрирует впечатляющие возможности в обработке и анализе данных. Современные системы способны: объединять информацию из различных источников; проводить мгновенный анализ больших массивов данных; работать с различными форматами информации (текст, изображения, неструктурированные данные); обеспечивать высокую скорость обработки информации.

Синергия естественного и искусственного интеллекта создает мощный инструмент преобразования данных. ИИ выступает не как замена человеческому интеллекту, а как его эффективный помощник, расширяющий возможности в различных сферах деятельности.

Перспективы развития данной области связаны с дальнейшим совершенствованием алгоритмов обработки данных и расширением сфер применения ИИ. При этом важно сохранять баланс между автоматизацией процессов и сохранением человеческого фактора в принятии решений.

Таким образом, естественный и искусственный интеллект представляют собой взаимодополняющие системы преобразования данных, совместное использование которых открывает новые возможности для развития технологий и общества в целом.

Авторы в своей интеллектуальной деятельности (наука), под руководством и личном участии профессора, д. т. н. Лабунца В.Г., попытались определить интеллект и сознание в технологиях распознавания. Была построена и апробирована, алгебраическую модель в алгебре Клиффорда [153]. Полный текст доступен по ссылке vikchas.ru/Uploads/Biblioteka/MonografMI.pdf

Авторы апробировали задачу, сочетающею интеграцию естественного и искусственного интеллекта в неопределенных таксационных данных [142, 154].

ЛИТЕРАТУРА

1. Указ Президента Российской Федерации от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации». Собрание законодательства РФ, 14.10.2019, № 41, ст. 5700

2. О внесении изменений в Указ Президента Российской Федерации от 10 октября 2019 г. № 490 "О развитии искусственного интеллекта в Российской Федерации" и в Национальную стратегию, утвержденную этим Указом: указ Президента РФ от 15 февраля 2024 г. № 124// Собрание законодательства РФ, 19.02.2024, № 4, ст. 1102.

3. Кант И. Критика чистого разума / Пер. с нем. Н. Лосского сверен и отредактирован Ц. Г. Арзаканяном и М. И. Иткиным, прим. Ц. Г. Арзаканяна. М.: Мысль, 1994. 591 с.

4. А.И. Аветисян, Академик РАН, директор института системного программирования РАН. Искусственный интеллект по Иммануилу Канту. [Видео]. 24 июня 2024 г. Видеохостинг RUTUBE, Россия. [Электронный ресурс]. URL: https://rutube.ru/video/9c34f69d8a8d16b526f8841bfe40686b (15 июня 2025).

5. Николлс Джон, Мартин Роберт, Валлас Брюс, Фукс Пол. От нейрона к мозгу / Пер. с англ. П. М. Балабана, А.В.Галкина, Р. А. Гиниатуллина, Р.Н.Хазипова, Л.С.Хируга. — М.: Едиториал УРСС, 2003. — 672 с.

6. Eric R. Kandel, James H. Schwartz, and Thomas M. Jessell Principles of Neural Science, Fifth Edition. – McGraw-Hill Education / Medical, the sixth and latest edition was published on March 8, 2021. p. 1760.

7. Между скальпелем и сознанием. С. Беляева. – Поиск №24-25(1878-1879) 20 июня 2025, стр. 8-9.

8. Алексей Квашенкин. Известный нейрохирург Коновалов отмечает 90-летие. [Видео]. 12 декабря 2023 г. Архив НТВ, Россия. [Электронный ресурс]. URL: https://www.ntv.ru/novosti/2803500/?ysclid=mdbznv6c3r80104582 (15 июня 2025).

9. Владимир Кожемякин. Мозг себе на уме. Что за «существо» живёт в нашей черепной коробке? – Еженедельник «Аргументы и Факты» № 5. Будем жить в чистой стране 01/02/2017[Электронный ресурс]. URL: https://aif.ru/health/psycho-logic/mozg_sebe_na_ume_chto_za_sushchestvo_zhivyot_v_nashey_cherep-noy_korobke?ysclid=mdc2fgc11f944608919

10. Мозг против компьютера. 11 апреля 2025 г. // dzen.ru [сайт]. URL: https://dzen.ru/a/Z_kiZt5bDiq7sJ-P (дата обращения: 15.07.2025).

11. Е. Понарина. Разве мы просто роботы? «Поиск» №10, 2025 от 06 марта 2025, с. 4-5. URL: https://poisknews.ru/releases/razve-my-prosto-roboty/

12. Игорь Каляев. Попытки создать аналог человеческого мозга с помощью компьютерных технологий — это путь в никуда. ЭКСПЕРТ №2(3) 9 февраля 2024. – URL: https://expert.ru/mnenie/igor-kalyaev-popytki-sozdat-analog-chelovecheskogo-mozga-s-pomoshchyu-kompyuternykh-tekhnologiy-eto-put-v-nikuda/

13. Оливейра, Арлиндо. Цифровой разум: как наука меняет человечество/Арлиндо Оливейра; переводе английского К. Чистопольской; под научной редакцией М.Фаликман, — Москва: Издательский дом «Дело» РАНХиГС, 2022. — 448 с.

14. О’Коннелл, Марк. Искусственный интеллект и будущее человечества / Марк О’К ; [перевод с английского М. Кудряшова]. — Москва: экс: 2019. — 272 с.

15. Нейман, Джон фон. Вычислительная машина и мозг / Джон фон Нейман; (пер. с англ. А. Чечиной]. — Москва: Издательство АСТ, 2018. — 192 с. — (Наука: открытия и первооткрыватели).

16. Пиковер К. Искусственный интеллект / Клиффорд Пиковер; [пер. с англ. А. Ефимовой]. — М.: Синдбад, 202L — 224 с.

17. Идеи, определившие облик информатики / под ред. Гарри Р. Льюиса; пер. с англ. А. А. Слинкина. – М.: ДОЖ Пресс, 2023. – 616 с.: ил.

18. Сейновски, Терренс.Антология машинного обучения: важнейшие исследования в области ИИ за последние 60 лет / Терренс Сейновски; [перевод с английского М. А. Райтмана, Е. В. Сазановой]. — Москва: Эксмо, 2022. — 304 с.

19. Рассел, Стюарт, Норвнг, Питер. Искусственный интеллект: современный подход, 2-е изд.: Пер. с англ. — СПб.: ООО «Диалектика», 2020. — 1408 с.: ил.

20. Дэвид Чалмерс. Сознающий ум: в поисках фундаментальной теории. – Москва: URSS, «Либроком», – 2015. -509 с.

21. Бостром, Ник. Искусственный интеллект. Этапы. Угрозы. Стратегии. — Москва: «Манн, Иванов и Фербер», 2016. – 490 с.

22. Т.В. Черниговская. Чеширская улыбка кота Шрёдингера: Язык и сознание. – Москва: АСТ, 2025. – 496 с.

23. Сьюзан Шнайдер. Искусственный ты. Машинный интеллект и будущее нашего разума. – Москва: Альпина нон-фикшн, 2022. -243 с.

24. Джон Сёрл. Открывая сознание заново. – Москва: Идея-Пресс, – 2002.- 240 с.

25. Чалмерс, Дэвид Джон. Сознающий ум: в поисках фундаментальной теории. Пер. с англ. В. В. Васильева. — Москва: URSS, Либроком. 2013. — 509 с.

26. Джулио Тонони и Кьяры Чирелли. Убирая лишнее. В мире науки (Scientific American). – 2013. – № 10. – с.42-48.

27. G. M. Edelman, G. Tononi. A Universe of Consciousness: How Matter Becomes Imagination. – New York: Basic Books, 2000. – 274 с.

28. Роджера Пенроуза. Новый ум короля. О компьютерах, мышлении и законах физики. – Москва: URSS, ЛЕНАНД, 2024. – 416 с

29. J. R. Lucas. Minds, Machines and Gödel. – Philosophy, Apr. – Jul., 1961, Vol. 36, No. 137, pp. 112-127. – Cambridge University Press on behalf of Royal Institute of Philosophy. – URL: https://www.jstor.org/stable/3749270

30. Уоллмарк, Л. Ада Байрон Лавлейс – первый программист. – М.: пешком в историю, 2017. – 40 с.

31. Глушков В. М. Основы безбумажной информатики. Изд. 2-е, испр.— М.: Наука. Гл. ред. физ.-мат. лит., 1987.— 552 с.

32. Дубынин В. А. Основные нейромедиаторы. Часть 1. 21 мая 2020 г. // scientificrussia.ru [сайт]. URL: https://scientificrussia.ru/articles/osnovnye-nejromediatory-chast-1 (дата обращения: 15.07.2025).

33. Покровский, В. М. Физиология человека: учебник / Под ред. В. М. Покровского, Г. Ф. Коротько – 3-е изд. – Москва: Медицина, 2011. – 664 с. – Текст: электронный // URL : https://www.rosmedlib.ru/book/ISBN9785225100087.html (дата обращения: 25.06.2025).

34. Питер Уайброу. Мозг. Тонкая настройка. Наша жизнь с точки зрения нейронауки. -Москва, «Альпина Паблишер, 2016. -350 с.

35. Основы физиологии человека. Том 1. под ред. акад. РАМН Б.И.Ткаченко. – СПб: Международный фонд истории науки,1994. – 567 с.

36. Хайкнн, Саймон.Нейронные сети: полный курс. — М.: Диалектика, 2020. — 1104 с.

37. Благинин В. А., Соколова Е. В., Адакава М. И. Достижения и тен денции в области нейротехнологий и искусственного интеллекта в Российской Феде рации: комплексный наукометрический анализ // Цифровые модели и решения. 2023. Т. 2, № 4. С. 13–29. DOI: 10.29141/2949-477X-2023-2-4-2. EDN: MDCONT.

38. А.А. Симакова , М.А. Горбатова , Д.С. Русанов , А.А. Карякин , А.М. Гржибовский ПРИМЕНЕНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ В ПРАКТИКЕ ВРАЧАОРТОДОНТА: ОБЗОР РОССИЙСКИХ ИССЛЕДОВАНИЙ ЗА ПЕРИОД 2013-2023 // Вестник новых медицинских технологий. Электронное издание. 2024. №4. URL: https://cyberleninka.ru/article/n/primenenie-iskusstvennyh-neyronnyh-setey-v-praktike-vrachaortodonta-obzor-rossiyskih-issledovaniy-za-period-2013-2023 (18.05.2025).

39. Владислав Олегович Саяпин ИНТЕЛЛЕКТУАЛЬНЫЕ НЕЙРОСЕТИ — БУДУЩИЙ ПОТЕНЦИАЛ ЦИВИЛИЗАЦИОННОГО РАЗВИТИЯ ЦИФРОВОГО МИРА // Вестник Челябинского государственного университета. 2023. №7 (477). URL: https://cyberleninka.ru/article/n/intellektualnye-neyroseti-buduschiy-potentsial-tsivilizatsionnogo-razvitiya-tsifrovogo-mira (25.02.2025).

40. Тюрина Д.А., Пальмов С.В. ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ В ОБРАБОТКЕ ЕСТЕСТВЕННОГО ЯЗЫКА // Журнал прикладных исследований. 2023. №7. URL: https://cyberleninka.ru/article/n/primenenie-neyronnyh-setey-v-obrabotke-estestvennogo-yazyka (10.12.2024).

41. Авдонин Владимир Сергеевич, Силаева Виктория Леонидовна НЕЙРОСЕТИ НОВОГО ПОКОЛЕНИЯ В КОНТЕКСТЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА, ФИЛОСОФИИ И СОЦИАЛЬНО-ПОЛИТИЧЕСКИХ НАУК // Политическая наука. 2023. №4. URL: https://cyberleninka.ru/article/n/neyroseti-novogo-pokoleniya-v-kontekste-tehnologiy-iskusstvennogo-intellekta-filosofii-i-sotsialno-politicheskih-nauk (27.07.2025).

42. Наумов И.М. ТЕНДЕНЦИИ РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ПОЛНЫЙ ОБЗОР ТЕХНОЛОГИЙ, РЫНКА И БУДУЩЕГО // Вестник науки. 2025. №6 (87). URL: https://cyberleninka.ru/article/n/tendentsii-razvitiya-iskusstvennogo-intellekta-polnyy-obzor-tehnologiy-rynka-i-buduschego (25.07.2025).

43. Адылова Фатима Туйчиевна УСПЕХИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЗДРАВООХРАНЕНИИ в 2023 ГОДУ И ПРОГНОЗ на 2024 ГОД: АНАЛИТИКА МИРОВЫХ ТРЕНДОВ // Raqamli iqtisodiyot (Цифровая экономика). 2024. №7. URL: https://cyberleninka.ru/article/n/uspehi-iskusstvennogo-intellekta-v-zdravoohranenii-v-2023-godu-i-prognoz-na-2024-god-analitika-mirovyh-trendov (10.12.2024).

44. М.А. Мирошниченко, А.А. Абдуллаева, А.В. Джунь ТЕНДЕНЦИИ РАЗВИТИЯ В РОССИИ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В ПЕРИОД ЦИФРОВОЙ ТРАНСФОРМАЦИИ // Естественно-гуманитарные исследования. 2023. №5 (49). URL: https://cyberleninka.ru/article/n/tendentsii-razvitiya-v-rossii-tehnologiy-iskusstvennogo-intellekta-v-period-tsifrovoy-transformatsii (12.01.2025).

45. Гринин Леонид Ефимович, Гринин Антон Леонидович, Гринин Игорь Леонидович ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: РАЗВИТИЕ И ТРЕВОГИ. ВЗГЛЯД В БУДУЩЕЕ. Статья первая. Информационные технологии и искусственный интеллект: прошлое, настоящее и некоторые прогнозы // Философия и общество. 2023. №3 (108). URL: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-razvitie-i-trevogi-vzglyad-v-buduschee-statya-pervaya-informatsionnye-tehnologii-i-iskusstvennyy-intellekt (09.12.2024).

46. Асадуллина Анна Викторовна, Белоусов Виталий Сергеевич Глобальные тренды в развитии и регулировании технологий искусственного интеллекта // Российский внешнеэкономический вестник. 2023. №9. URL: https://cyberleninka.ru/article/n/globalnye-trendy-v-razvitii-i-regulirovanii-tehnologiy-iskusstvennogo-intellekta (10.12.2024).

47. Галина Васильевна Ярошенко, Иван Андреевич Савушкин Социальные последствия применения систем искусственного интеллекта в образовании // Государственное и муниципальное управление. Ученые записки . 2023. №3. URL: https://cyberleninka.ru/article/n/sotsialnye-posledstviya-primeneniya-sistem-iskusstvennogo-intellekta-v-obrazovanii (11.12.2024).

48. Н.А. Калашникова, О.А. Родин ДОЛГОСРОЧНЫЕ ПОСЛЕДСТВИЯ ИНТЕГРАЦИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ОБЩЕСТВЕННУЮ ЖИЗНЬ: РИСКИ И УГРОЗЫ // Международный журнал гуманитарных и естественных наук. 2024. №10-2. URL: https://cyberleninka.ru/article/n/dolgosrochnye-posledstviya-integratsii-iskusstvennogo-intellekta-v-obschestvennuyu-zhizn-riski-i-ugrozy (13.12.2024).

49. Вантяева Анастасия Сергеевна СОЦИАЛЬНЫЕ РИСКИ ВНЕДРЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА // Теория и практика общественного развития. 2022. №7 (173). URL: https://cyberleninka.ru/article/n/sotsialnye-riski-vnedreniya-tehnologiy-iskusstvennogo-intellekta (13.12.2024).

50. А.А. Шпак, Н.М. Лещинская Социальные последствия внедрения искусственного интеллекта в области искусства // Журнал Сибирского федерального университета. Гуманитарные науки. 2024. №8. URL: https://cyberleninka.ru/article/n/sotsialnye-posledstviya-vnedreniya-iskusstvennogo-intellekta-v-oblasti-iskusstva (04.06.2025).

51. Крылова Мария Николаевна Социальные угрозы, которые вызывает искусственный интеллект, и реакция на них общества // Studia Humanitatis. 2024. №1. URL: https://cyberleninka.ru/article/n/sotsialnye-ugrozy-kotorye-vyzyvaet-iskusstvennyy-intellekt-i-reaktsiya-na-nih-obschestva (20.12.2024).

52. Тришечкин Сергей Николаевич Data Mining и метод нейронных сетей // Вестник науки и образования. 2019. №8-1 (62). URL: https://cyberleninka.ru/article/n/data-mining-i-metod-neyronnyh-setey (10.01.2025).

53. Лапина М. А., Подручный Н. В., Багаутдинова А. Р., Шидловский А. С., Кущенко А. А. ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ АНАЛИЗА И ОБРАБОТКИ ВИЗУАЛЬНЫХ ДАННЫХ // Auditorium. 2025. URL: https://cyberleninka.ru/article/n/primenenie-neyronnyh-setey-dlya-analiza-i-obrabotki-vizualnyh-dannyh (30.05.2025).

54. Хрищатый А.С. ИССЛЕДОВАНИЕ ИСПОЛЬЗОВАНИЯ НЕЙРОСЕТЕЙ ДЛЯ АНАЛИЗА ДАННЫХ И ПРИНЯТИЯ БИЗНЕС-РЕШЕНИЙ: АНАЛИЗ ЭФФЕКТИВНОСТИ ИСПОЛЬЗОВАНИЯ НЕЙРОСЕТЕЙ ДЛЯ ОБРАБОТКИ БОЛЬШИХ ОБЪЕМОВ ДАННЫХ И ПРЕДОСТАВЛЕНИЯ ЦЕННЫХ ИНСАЙТОВ ДЛЯ ПРИНЯТИЯ РЕШЕНИЙ // Инновации и инвестиции. 2023. №7. URL: https://cyberleninka.ru/article/n/issledovanie-ispolzovaniya-neyrosetey-dlya-analiza-dannyh-i-prinyatiya-biznes-resheniy-analiz-effektivnosti-ispolzovaniya (10.12.2024).

55. Наркевич Артем Николаевич, Виноградов Константин Анатольевич, Параскевопуло Константин Михайлович, Гржибовский Андрей Мечиславович ИНТЕЛЛЕКТУАЛЬНЫЕ МЕТОДЫ АНАЛИЗА ДАННЫХ В БИОМЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ: НЕЙРОННЫЕ СЕТИ // Экология человека. 2021. №4. URL: https://cyberleninka.ru/article/n/intellektualnye-metody-analiza-dannyh-v-biomeditsinskih-issledovaniyah-neyronnye-seti (16.02.2025).

56. Т.А. Кузовкова, М.М. Шаравова, Д.А. Катунин АНАЛИЗ ПЕРСПЕКТИВ РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА // Экономика и качество систем связи. 2024. №1 (31). URL: https://cyberleninka.ru/article/n/analiz-perspektiv-razvitiya-iskusstvennogo-intellekta (10.12.2024).

57. Дробышевская Л.Н., Молодцова А.В. ТЕНДЕНЦИИ И ПЕРСПЕКТИВЫ РАЗВИТИЯ ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА // Экономика и бизнес: теория и практика. 2020. №11-1. URL: https://cyberleninka.ru/article/n/tendentsii-i-perspektivy-razvitiya-tehnologii-iskusstvennogo-intellekta (10.12.2024).

58. Назаров Д.М., Назаров А.Д. Перспективы искусственного интеллекта и их влияния на цифровую экономику России // Умная цифровая экономика. 2022. №3. URL: https://cyberleninka.ru/article/n/perspektivy-iskusstvennogo-intellekta-i-ih-vliyaniya-na-tsifrovuyu-ekonomiku-rossii (11.12.2024).

59. Т.А. Кузовкова, М.М. Шаравова, Д.А. Катунин АНАЛИЗ ПЕРСПЕКТИВ РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА // Экономика и качество систем связи. 2024. №1 (31). URL: https://cyberleninka.ru/article/n/analiz-perspektiv-razvitiya-iskusstvennogo-intellekta (10.12.2024).

60. Гринин Леонид Ефимович, Гринин Антон Леонидович, Гринин Игорь Леонидович ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: РАЗВИТИЕ И ТРЕВОГИ. ВЗГЛЯД В БУДУЩЕЕ. Статья первая. Информационные технологии и искусственный интеллект: прошлое, настоящее и некоторые прогнозы // Философия и

общество. 2023. №3 (108). URL: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-razvitie-i-trevogi-vzglyad-v-buduschee-statya-pervaya-informatsionnye-tehnologii-i-iskusstvennyy-intellekt (09.12.2024).

61. Утемуратова Асем Назарбековна ОБЛАСТИ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ПРИНЦИП ЕГО РАБОТЫ // Raqamli iqtisodiyot (Цифровая экономика). 2023. №3. URL: https://cyberleninka.ru/article/n/oblasti-primeneniya-iskusstvennogo-intellekta-i-printsip-ego-raboty (10.12.2024).

62. А.А. Поляков, Т.А. Мамаджарова, Е.С. Балашова ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ РЕВОЛЮЦИЯ В СОВРЕМЕННЫХ ОТРАСЛЯХ ПРОМЫШЛЕННОСТИ // Естественно-гуманитарные исследования. 2024. №5 (55). URL: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-revolyutsiya-v-sovremennyh-otraslyah-promyshlennosti (15.12.2024).

63. Попов В.В. Влияние применения искусственного интеллекта на различные отрасли экономики России // Прикладная статистика и искусственный интеллект. 2024. №2. URL: https://cyberleninka.ru/article/n/vliyanie-primeneniya-iskusstvennogo-intellekta-na-razlichnye-otrasli-ekonomiki-rossii (02.02.2025).

64. Нурымова О., Тойлыев Б. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ЕГО ПРИМЕНЕНИЕ В РАЗЛИЧНЫХ СФЕРАХ ЖИЗНИ // Символ науки. 2024. №4-1-1. URL: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-i-ego-primenenie-v-razlichnyh-sferah-zhizni (16.02.2025).

65. Аширова Б., Шыхыева М., Бабаджанова М., Гулмырадов А., Бердиев Ш. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: СФЕРЫ ПРИМЕНЕНИЯ // Символ науки. 2024. №5-1-3. URL: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-sfery-primeneniya (13.05.2025).

65. Владислав Олегович Саяпин ИНТЕЛЛЕКТУАЛЬНЫЕ НЕЙРОСЕТИ — БУДУЩИЙ ПОТЕНЦИАЛ ЦИВИЛИЗАЦИОННОГО РАЗВИТИЯ ЦИФРОВОГО МИРА // Вестник Челябинского государственного университета. 2023. №7 (477). URL: https://cyberleninka.ru/article/n/intellektualnye-neyroseti-buduschiy-potentsial-tsivilizatsionnogo-razvitiya-tsifrovogo-mira (25.02.2025).

66. cyberleninka.ru/article/n/nastoyaschee-i-buduschee-neyronnyh-setey... [Электронный ресурс] // cyberleninka.ru – Режим доступа: https://cyberleninka.ru/article/n/nastoyaschee-i-buduschee-neyronnyh-setey/viewer, свободный. – Загл. с экрана

67. Абрагин Артур Викторович Перспективы развития и применения нейронных сетей // Проблемы современной науки и образования. 2015. №12 (42). URL: https://cyberleninka.ru/article/n/perspektivy-razvitiya-i-primeneniya-neyronnyh-setey (19.12.2024).

68. Качагина Кристина Сергеевна, Сафарова А. Д. НЕЙРОННЫЕ СЕТИ – ПЕРСПЕКТИВЫ РАЗВИТИЯ // E-Scio. 2021. №2 (53). URL: https://cyberleninka.ru/article/n/neyronnye-seti-perspektivy-razvitiya (23.02.2025).

69. Ефимова Софья Андреевна РАЗВИТИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА // Цифровая наука. 2020. №6. URL: https://cyberleninka.ru/article/n/razvitie-iskusstvennogo-intellekta (10.12.2024).

70. Новлянский В.В. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В СОВРЕМЕННОЙ НАУКЕ И РОЛЬ В РАЗВИТИИ // Вестник науки. 2024. №4 (73). URL: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-v-sovremennoy-nauke-i-rol-v-razvitii (10.04.2025).

71. Ширин Данияровна Искандерова Влияние искусственного интеллекта на современный мир // Science and Education. 2023. №4. URL: https://cyberleninka.ru/article/n/vliyanie-iskusstvennogo-intellekta-na-sovremennyy-mir (10.12.2024).

72. Вислова А.Д. СОВРЕМЕННЫЕ ТЕНДЕНЦИИ РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА // Известия Кабардино-Балкарского научного центра РАН. 2020. №2 (94). URL: https://cyberleninka.ru/article/n/sovremennye-tendentsii-razvitiya-iskusstvennogo-intellekta (10.12.2024).

73. А.Ю. Анисимов, А.Н. Алексахин, С.А. Алексахина, Е.И. Алёхин ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В СОВРЕМЕННОМ ОБЩЕСТВЕ: ТЕКУЩЕЕ

СОСТОЯНИЕ И ОЦЕНКА ПЕРСПЕКТИВ РАЗВИТИЯ // Естественно-гуманитарные исследования. 2024. №4 (54). URL: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-v-sovremennom-obschestve-tekuschee-sostoyanie-i-otsenka-perspektiv-razvitiya (16.12.2024).

74. Igor Anureev. What Prevents the Use of Formal Methods for Al Verification. [Видео]. 24 марта 2025 г. Видеохостинг RUTUBE, Россия. [Электронный ресурс]. URL: https://rutube.ru/video/ad09740477e5dd2b1df496fa54441746/?ysclid=mdnei2x5sr298173362 (15 июня 2025).

75. 12+ ограничений искусственного интеллекта в 2025 году... [Электронный ресурс] // aimojo.io – Режим доступа: https://aimojo.io/ru/limitations-artificial-intelligence/, свободный. – Загл. с экрана

76. arcsinus | Как устроена нейронная сеть и зачем ей слои [Электронный ресурс] // www.arcsinus.ru – Режим доступа: https://www.arcsinus.ru/blog/neuronet-architecture, свободный. – Загл. с экрана

77. Архитектуры нейросетей: от простых моделей к сложным... [Электронный ресурс] // chataibot.ru – Режим доступа: https://chataibot.ru/blog/arkhitektury-neyrosetey/, свободный. – Загл. с экрана

78. Будущее искусственного интеллекта: Возможности и вызовы [Электронный ресурс] // vc.ru – Режим доступа: https://vc.ru/id5122306/2109039-budushchee-iskusstvennogo-intellekta-vyzovy-i-vozmozhnosti, свободный. – Загл. с экрана

79. Нейросеть, которая может все: какие есть сложности в обучении... [Электронный ресурс] // trends.rbc.ru – Режим доступа: https://trends.rbc.ru/trends/industry/6384647c9a794723d7a12f58, свободный. – Загл. с экрана

80. Понимание моделей ИИ: как выбрать и настроить ИИ... | AppMaster [Электронный ресурс] // appmaster.io – Режим доступа: https://appmaster.io/ru/blog/ponimanie-modelei-ii, свободный. – Загл. с экрана

81. Иллюзия мышления: Почему «думающие» модели на... / Хабр [Электронный ресурс] // habr.com – Режим доступа: https://habr.com/ru/articles/921110/, свободный. – Загл. с экрана

82. Сложность модели и переоснащение в машинном обучении... [Электронный ресурс] // tr-page.yandex.ru – Режим доступа: https://tr-page.yandex.ru/translate?lang=en-ru&url=https://www.geeksforgeeks.org/machine-learning/model-complexity-overfitting-in-machine-learning/, свободный. – Загл. с экрана

83. Управление возможностями ИИ — Википедия [Электронный ресурс] // ru.wikipedia.org – Режим доступа: https://ru.wikipedia.org/wiki/управление_возможностями_ии, свободный. – Загл. с экрана

84. 5 способов, которыми качество данных может повлиять на... | Шаип [Электронный ресурс] // ru.shaip.com – Режим доступа: https://ru.shaip.com/blog/5-ways-data-quality-can-impact-your-ai-solution/, свободный. – Загл. с экрана

85. Прадипта Мишра. Объяснимые модели искусственного интеллекта на Python. Модель искусственного интеллекта. Объяснения с использованием библиотек, расширений и фреймворков на основе языка Python. – М.: ДМК Пресс, 2022. – 298 с.:

86. Рашид, Тарик. Создаем нейронную сеть.:. — СПб.: ООО «Диалектика», 2019. — 272 с.

87. Проблема черного ящика в ИИ: объяснимый искусственный... [Электронный ресурс] // science.mail.ru – Режим доступа: https://science.mail.ru/articles/4129-chernyj-yashik-nauki/, свободный. – Загл. с экрана

88. Ученые все чаще не могут объяснить, как работает ИИ.... / Хабр [Электронный ресурс] // habr.com – Режим доступа: https://habr.com/ru/companies/getmatch/articles/700736/, свободный. – Загл. с экрана

89. Ученые все чаще не могут объяснить, как работает ИИ.... / Хабр [Электронный ресурс] // habr.com – Режим доступа: https://habr.com/ru/companies/getmatch/articles/700736/, свободный. – Загл. с экрана.

90. 10 популярных методов, моделей и алгоритмов в машинном... [Электронный ресурс] // blog.skillfactory.ru – Режим доступа: https://blog.skillfactory.ru/mashinnoe-obuchenie-10-populyarnyh-metodov/, свободный. – Загл. с экрана.

91. 15 примеров применения Natural Language Processing / Хабр [Электронный ресурс] // habr.com – Режим доступа: https://habr.com/ru/companies/otus/articles/930130/, свободный. – Загл. с экрана

92. 5 главных проблем ML и MLOps с данными на практических кейсах [Электронный ресурс] // bigdataschool.ru – Режим доступа: https://bigdataschool.ru/blog/mlops-ml-issues-and-challenges/, свободный. – Загл. с экрана

93. 9 проблем машинного обучения | Блог Касперского [Электронный ресурс] // www.kaspersky.ru – Режим доступа: https://www.kaspersky.ru/blog/machine-learning-ten-challenges/21193/, свободный. – Загл. с экрана

94. KAN: Kolmogorov-Arnold Networks [Электронный ресурс] // doi.org – Режим доступа: https://doi.org/10.48550/arXiv.2404.19756. – Загл. с экрана

95. А. Н. Колмогоров, о представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения, – Докл. АН СССР, 1957, том 114, номер 5, 953–956 [Электронный ресурс] // mathnet.ru – Режим доступа: https://www.mathnet.ru/dan22050

96. Сеть Колмогорова-Арнольда: как теорема 1957 года... [Электронный ресурс] // www.securitylab.ru – Режим доступа: https://www.securitylab.ru/news/551974.php, свободный. – Загл. с экрана

97. Сети Колмогорова-Арнольда (KAN) могут навсегда изменить мир ИИ [Электронный ресурс] // nuancesprog.ru – Режим доступа: https://nuancesprog.ru/p/21925/, свободный. – Загл. с экрана

98. Разбираем KAN по полочкам / Хабр [Электронный ресурс] // habr.com – Режим доступа: https://habr.com/ru/articles/815851/, свободный. – Загл. с экрана

99. Понимание KAN: новейшая альтернатива MLP – Аналитика Vidhya [Электронный ресурс] // tr-page.yandex.ru – Режим доступа: https://tr-page.yandex.ru/translate?lang=en-ru&url=https://www.analyticsvidhya.com/blog/2024/06/kolmogorov-arnold-networks-kan-alternative-to-mlp/, свободный. – Загл. с экрана

100. Революционный подход к нейросетям: рассказываем про KAN... [Электронный ресурс] // telegra.ph – Режим доступа: https://telegra.ph/revolyucionnyj-podhod-k-nejrosetyam-rasskazyvaem-pro-kan-kolmogorov-arnold-networks-06-11, свободный. – Загл. с экрана

101. Разбор статьи про KAN – принципиально новую... | Data Secrets [Электронный ресурс] // datasecrets.ru – Режим доступа: https://datasecrets.ru/articles/9, свободный. – Загл. с экрана

102. Сети Колмогорова-Арнольда: новый «старый» шаг... / Хабр [Электронный ресурс] // habr.com – Режим доступа: https://habr.com/ru/articles/843234/, свободный. – Загл. с экрана

103. Сеть Колмогорова-Арнольда – GeeksforGeeks [Электронный ресурс] // tr-page.yandex.ru – Режим доступа: https://tr-page.yandex.ru/translate?lang=en-ru&url=https://www.geeksforgeeks.org/machine-learning/kolmogorov-arnold-network/, свободный. – Загл. с экрана

104. Адылова Фатима Туйчиевна ИДЕЯ, ОСНОВНЫЕ РАЗРАБОТКИ И ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ. КОЛМОГОРОВА-АРНОЛЬДА: АНАЛИТИЧЕСКИЙ ОБЗОР // Raqamli iqtisodiyot (Цифровая экономика). 2025. №10. URL: https://cyberleninka.ru/article/n/ideya-osnovnye-razrabotki-i-primenenie-neyronnyh-setey-kolmogorova-arnolda-analiticheskiy-obzor (27.07.2025).

105. [2404.19756] KAN: Kolmogorov-Arnold Networks [Электронный ресурс] // arxiv.org – Режим доступа: https://arxiv.org/abs/2404.19756, свободный. – Загл. с экрана

106. Г.М. Мкртчян О ЗАДАЧЕ ПРИМЕНИМОСТИ СЕТЕЙ КОЛМОГОРОВА-АРНОЛЬДА В ЛЕГКОВЕСНОЙ НЕЙРОСЕТЕВОЙ МОДЕЛИ КЛАССИФИКАЦИИ АКУСТИЧЕСКИХ ДАННЫХ // Экономика и качество систем связи. 2024. №4 (34). URL: https://cyberleninka.ru/article/n/o-zadache-primenimosti-setey-kolmogorova-arnolda-v-legkovesnoy-neyrosetevoy-modeli-klassifikatsii-akustich-eskih-dannyh (27.07.2025).

107. Нейросети нового поколения: трансформеры и KAN [Электронный ресурс] // lab.neural-university.ru – Режим доступа: https://lab.neural-university.ru/transformers_kan, свободный. – Загл. с экрана

108. О задаче применимости сетей Колмогорова – Арнольда в [Электронный ресурс] // journal-ekss.ru – Режим доступа: https://journal-ekss.ru/wp-content/uploads/2024/12/152-158.pdf, свободный. – Загл. с экрана

109. KAN: Kolmogorov–Arnold Networks / Хабр [Электронный ресурс] // habr.com – Режим доступа: https://habr.com/ru/articles/856776/, свободный. – Загл. с экрана

110. Новая архитектура нейронных сетей KAN переходит MLP [Электронный ресурс] // anns.ru – Режим доступа: https://anns.ru/articles/news/2024/05/05/novaja_arhitektura_neyronnih_setey_kan_perehodit_mlp, свободный. – Загл. с экрана

111. Исследователи изобрели новый вид нейросетей — KAN [Электронный ресурс] // www.computerra.ru – Режим доступа: https://www.computerra.ru/296831/issledovateli-izobreli-novyj-vid-nejroseti-kan/, свободный. – Загл. с экрана

112. В США создали принципиально новую архитектуру... – CNews [Электронный ресурс] // www.cnews.ru – Режим доступа:

https://www.cnews.ru/news/top/2024-05-02_amerikanskie_uchenye_sozdali, свободный. – Загл. с экрана.

113. Сильный искусственный интеллект: на подступах к сверхразуму / Александр Велихов [и др.]. – М.: Интеллектуальная литература, 2021. – 232 с.

114. JohnR. Searle. Minds, brains, and programs. – The behavioral and brain sciences (1980) 3,417-457.

115. Иванов Александр Иванович, Кубасов Игорь Анатольевич СИЛЬНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ПОВЫШЕНИЕ КАЧЕСТВА НЕЙРОСЕТЕВЫХ РЕШЕНИЙ С ПЕРЕХОДОМ К ОБРАБОТКЕ ВХОДНЫХ ДАННЫХ БОЛЬШОГО ОБЪЕМА // Надежность и качество сложных систем. 2021. №1 (33). URL: https://cyberleninka.ru/article/n/silnyy-iskusstvennyy-intellekt-povyshenie-kachestva-neyrosetevyh-resheniy-s-perehodom-k-obrabotke-vhodnyh-dannyh-bolshogo-obema (23.02.2025).

116. Владимир Николаевич Белов, Сергей Александрович Малофейкин Модели объяснения сознания в контексте проблемы определения сильной формы ИИ // Гуманитарные науки. Вестник Финансового университета. 2024. №4. URL: https://cyberleninka.ru/article/n/modeli-obyasneniya-soznaniya-v-kontekste-problemy-opredeleniya-silnoy-formy-ii (30.07.2025).

117. Константинова Л. В., Ворожихин В. В., Петров А. М., Титова Е. С., Штыхно Д. А. ГЕНЕРАТИВНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ОБРАЗОВАНИИ: ДИСКУССИИ И ПРОГНОЗЫ // Открытое образование. 2023. №2. URL: https://cyberleninka.ru/article/n/generativnyy-iskusstvennyy-intellekt-v-obrazovanii-diskussii-i-prognozy (18.12.2024).

118. И.В. Понкин Применение генеративного искусственного интеллекта в научных исследованиях и в прикладной аналитике в обеспечение государственного управления: позитивные возможности и «подводные камни» // International Journal of Open Information Technologies. 2025. №1. URL: https://cyberleninka.ru/article/n/primenenie-generativnogo-iskusstvennogo-intellekta-v-nauchnyh-issledovaniyah-i-v-prikladnoy-analitike-v-obespechenie (09.03.2025).

119. Гринин Леонид Ефимович, Гринин Антон Леонидович, Гринин Игорь Леонидович ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: РАЗВИТИЕ И ТРЕВОГИ. ВЗГЛЯД В БУДУЩЕЕ. Статья первая. Информационные технологии и искусственный интеллект: прошлое, настоящее и некоторые прогнозы // Философия и общество. 2023. №3 (108). URL: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-razvitie-i-trevogi-vzglyad-v-buduschee-statya-pervaya-informatsionnye-tehnologii-i-iskusstvennyy-intellekt (09.12.2024).

120. Обухова Елена Алексеевна ГЕНЕРАТИВНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ КАК ДРАЙВЕР РАЗВИТИЯ ВЫСОКОТЕХНОЛОГИЧНЫХ СЕКТОРОВ ЭКОНОМИКИ РОССИИ // Экономика и управление инновациями. 2024. №3. URL: https://cyberleninka.ru/article/n/generativnyy-iskusstvennyy-intellekt-kak-drayver-razvitiya-vysokotehnologichnyh-sektorov-ekonomiki-rossii (23.04.2025).

121. Назарова Александра Дмитриевна, Сулимин Владимир Власович ИЗМЕНЕНИЯ НА РЫНКЕ ТРУДА ПОД ВЛИЯНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ПЕРСПЕКТИВЫ БУДУЩЕГО // Международный журнал прикладных наук и технологий «Integral». 2023. №2. URL: https://cyberleninka.ru/article/n/izmeneniya-na-rynke-truda-pod-vliyaniem-iskusstvennogo-intellekta-perspektivy-buduschego (14.12.2024).

122. Новоселов Демид Олегович Искусственный интеллект и изменение трудовых ценностей, влияние автоматизации на восприятие работы // Телескоп: журнал социологических и маркетинговых исследований. 2025. №1. URL: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-i-izmenenie-trudovyh-tsennostey-vliyanie-avtomatizatsii-na-vospriyatie-raboty (20.05.2025).

123. Юрченко В. ВЛИЯНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА РЫНКИ ТРУДА: ПОДГОТОВКА К БУДУЩЕМУ // Вестник науки. 2024. №11 (80). URL: https://cyberleninka.ru/article/n/vliyanie-iskusstvennogo-intellekta-na-rynki-truda-podgotovka-k-buduschemu (18.01.2025).

124. Лялькова Евгения Евгеньевна, Богдашкина Елизавета Андреевна, Лобкова Виктория Эдуардовна ВЛИЯНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА РЫНОК ТРУДА: АНАЛИЗ ИЗМЕНЕНИЙ В СПРОСЕ НА КВАЛИФИКАЦИИ И ОБУЧЕНИИ // E-Scio. 2023. №5 (80). URL: https://cyberleninka.ru/article/n/vliyanie-iskusstvennogo-intellekta-na-rynok-truda-analiz-izmeneniy-v-sprose-na-kvalifikatsii-i-obuchenii (04.01.2025).

125. Шляпников Виктор Валерьевич НЕКОТОРЫЕ ПРОБЛЕМЫ ЭТИКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА // Идеи и идеалы. 2023. №2-2. URL: https://cyberleninka.ru/article/n/nekotorye-problemy-etiki-iskusstvennogo-intellekta (10.12.2024).

126. Песоцкая Кристина Ивановна, Дунаев Роман Алексеевич ЭТИЧЕСКИЕ ПРОБЛЕМЫ В РАЗРАБОТКЕ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА // NOMOTHETIKA: Философия. Социология. Право. 2023. №1. URL: https://cyberleninka.ru/article/n/eticheskie-problemy-v-razrabotke-sistem-iskusstvennogo-intellekta (20.12.2024).

127. Миндигулова Арина Александровна ЭТИКА И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ПРОБЛЕМЫ И ПРОТИВОРЕЧИЯ // Медицина. Социология. Философия. Прикладные исследования. 2022. №3. URL: https://cyberleninka.ru/article/n/etika-i-iskusstvennyy-intellekt-problemy-i-protivorechiya (21.12.2024).

128. Ратнер Н.П. ЭТИКА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ВЫЗОВЫ, РИСКИ И РЕШЕНИЯ // Вестник науки. 2023. №10 (67). URL: https://cyberleninka.ru/article/n/etika-iskusstvennogo-intellekta-vyzovy-riski-i-resheniya (10.12.2024).

129. Серобян Гагик Ашотович, Яковенко Андрей Александрович Этико-правовые проблемы использования систем искусственного интеллекта // Теология. Философия. Право. 2018. №4 (8). URL: https://cyberleninka.ru/article/n/etiko-pravovye-problemy-ispolzovaniya-sistem-iskusstvennogo-intellekta (20.12.2024).

130. Галина Васильевна Ярошенко, Иван Андреевич Савушкин Социальные последствия применения систем искусственного интеллекта в образовании // Государственное и муниципальное управление. Ученые записки . 2023. №3. URL: https://cyberleninka.ru/article/n/sotsialnye-posledstviya-primeneniya-sistem-iskusstvennogo-intellekta-v-obrazovanii (11.12.2024).

131. Цвык Владимир Анатольевич, Цвык Ирина Вячеславовна СОЦИАЛЬНЫЕ ПРОБЛЕМЫ РАЗВИТИЯ И ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА // Вестник Российского университета дружбы народов. Серия: Социология. 2022. №1. URL: https://cyberleninka.ru/article/n/sotsialnye-problemy-razvitiya-i-primeneniya-iskusstvennogo-intellekta (09.12.2024).

132. Вантяева Анастасия Сергеевна СОЦИАЛЬНЫЕ РИСКИ ВНЕДРЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА // Теория и практика общественного развития. 2022. №7 (173). URL: https://cyberleninka.ru/article/n/sotsialnye-riski-vnedreniya-tehnologiy-iskusstvennogo-intellekta (13.12.2024).

133. Городнова Наталья Васильевна МОДЕЛИРОВАНИЕ РАЗВИТИЯ И ВНЕДРЕНИЯ СИСТЕМ "СЛАБОГО" И "СИЛЬНОГО" ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ АСПЕКТЫ // Вопросы инновационной экономики. 2022. №1. URL: https://cyberleninka.ru/article/n/modelirovanie-razvitiya-i-vnedreniya-sistem-slabogo-i-silnogo-iskusstvennogo-intellekta-sotsialno-ekonomicheskie-aspekty (14.02.2025).

139. ГОСТ Р 59898-2021. Оценка качества систем искусственного... [Электронный ресурс] // gostassistent.ru – Режим доступа: https://gostassistent.ru/doc/d26985f3-7280-4a64-a90e-1b8e0e2ed5be, свободный. – Загл. с экрана

135. Исследователи разрабатывают инструмент для оценки качества ИИ [Электронный ресурс] // dzen.ru – Режим доступа: https://dzen.ru/a/zctmu4293d16sgaj, свободный. – Загл. с экрана

138. Методика оценки качества функционирования [Электронный ресурс] // telemil.ru – Режим доступа: https://telemil.ru/pages/archive/magazine1/9_ягович_методика_оценки_ред_3.pdf, свободный. – Загл. с экрана

134. Нейросети и реален ли сильный ИИ: большая проблема... / Хабр [Электронный ресурс] // habr.com – Режим доступа: https://habr.com/ru/companies/ddosguard/articles/892396/, свободный. – Загл. с экрана

136. Скачать ГОСТ Р 59898-2021 Оценка качества систем... [Электронный ресурс] // files.stroyinf.ru – Режим доступа: https://files.stroyinf.ru/data/769/76946.pdf, свободный. – Загл. с экрана

137. Что такое СИИ? Будущее... [Электронный ресурс] // www.ultralytics.com – Режим доступа: https://www.ultralytics.com/ru/blog/what-is-strong-ai-looking-into-the-future-of-ai, свободный. – Загл. с экрана

138. Глуздов Д.В. ФИЛОСОФСКО-АНТРОПОЛОГИЧЕСКИЕ ОСНОВАНИЯ ВЗАИМОДЕЙСТВИЯ ИСКУССТВЕННОГО И ЕСТЕСТВЕННОГО ИНТЕЛЛЕКТА // Вестник Мининского университета. 2022. №4 (41). URL: https://cyberleninka.ru/article/n/filosofsko-antropologicheskie-osnovaniya-vzaimodeystviya-iskusstvennogo-i-estestvennogo-intellekta (22.12.2024).

139. Беликова Евгения Константиновна, Попов Евгений Александрович Современные проблемы соотношения ЕИ и ИИ в парадигме культуры // Социально-гуманитарные знания. 2023. №11. URL: https://cyberleninka.ru/article/n/sovremennye-problemy-sootnosheniya-estestvennogo-i-iskusstvennogo-intellekta-v-paradigme-kultury (22.12.2024).

140. Т.А. Кузовкова, М.М. Шаравова, Д.А. Катунин АНАЛИЗ ПЕРСПЕКТИВ РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА // Экономика и качество систем связи. 2024. №1 (31). URL: https://cyberleninka.ru/article/n/analiz-perspektiv-razvitiya-iskusstvennogo-intellekta (10.12.2024).

141. Седых Н.В., Фоканов И.П. ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ РАЗВИТИЯ ТЕХНОЛОГИИ ИСКУСТВЕННОГО ИНТЕЛЛЕКТА // Естественно-гуманитарные исследования. 2022. №6 (44). URL: https://cyberleninka.ru/article/n/problemy-i-perspektivy-razvitiya-tehnologii-iskustvennogo-intellekta (17.12.2024).

142. Воронов М.П., Усольцев В.А., Часовских В.П. Исследование методов и разработка информационной системы определения и картирования депонируемого лесами углерода в среде Natural: Монография, 2 изд. испр. и доп. – Екатеринбург: Урал. гос. лесотехн. ун-т, 2015. 192 с.

143. Жилин В.В., Сафарьян О.А. ИИ в системах хранения данных // Advanced Engineering Research (Rostov-on-Don). 2020. №2. URL: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-v-sistemah-hraneniya-dannyh (13.01.2025).

144. Сколько данных создается каждый день? (2025). //[Электронный ресурс] // inclient.ru – Режим доступа: https://inclient.ru/data-create-stats//, свободный. – Загл. с экрана

145. Макаров Д.А., Шибанова А.Д. АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ // Теория и практика современной науки. 2018. №6 (36). URL: https

146. Оксенчук Т.А. АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РЕГРЕССИОННОГО АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ЧИСЛОВЫХ ДАННЫХ // Вестник науки. 2025. №6 (87). URL: https://cyberleninka.ru/article/n/algoritmy-mashinnogo-obucheniya-dlya-regressionnogo-analiza-i-prognozirovaniya-chislovyh-dannyh (27.06.2025).

147. Язгельдиев Ш. , Абаев Т., Гурбанмырадов Б. АНАЛИТИКА БОЛЬШИХ ДАННЫХ В УПРАВЛЕНИИ ЦЕПОЧКАМИ ПОСТАВОК: ПРИМЕРЫ УСПЕШНЫХ КЕЙСОВ // Вестник науки. 2024. №11 (80). URL: https://cyberleninka.ru/article/n/analitika-bolshih-dannyh-v-upravlenii-tsepochkami-postavok-primery-uspeshnyh-keysov (10.12.2024).

148. Чехарин Евгений Евгеньевич Большие данные: большие проблемы // Перспективы науки и образования. 2016. №3 (21). URL: https://cyberleninka.ru/article/n/bolshie-dannye-bolshie-problemy (14.12.2024).

149. Горячев А.С. ОБЗОР АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ ПРЕДИКТИВНОГО АНАЛИЗА РАБОТЫ ТЕХНОЛОГИЧЕСКОГО

ОБОРУДОВАНИЯ // Инновационная наука. 2024. №10-1. URL: https://cyberleninka.ru/article/n/obzor-algoritmov-mashinnogo-obucheniya-v-zadachah-prediktivnogo-analiza-raboty-tehnologicheskogo-oborudovaniya (06.04.2025).

150. Бойдаченко Вячеслав Владимирович, Бутин Петр Алексеевич, Семененко Андрей Владимирович, Косникова Оксана Владимировна ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА БОЛЬШИХ ДАННЫХ В ЭКОНОМИКЕ // Индустриальная экономика. 2024. №4. URL: https://cyberleninka.ru/article/n/primenenie-mashinnogo-obucheniya-dlya-analiza-bolshih-dannyh-v-ekonomike (16.03.2025).

151. М. Минский. Вычисления и автоматы. / М. Мир, 1971. – 366 с.

152. М. Минский. Машина Эмоций. / – М. АСТ, 200. – 93 с.

153.Лабунец В.Г., Часовских В.П. Является ли мозг классическим компьютером, работающим в алгебре Клиффорда? Математические основы теории. М-во науки и высш. образования Рос. Федерации, Урал. гос. экон. ун-т. — Екатеринбург: Изд-во ООО «Акдениз», 2022. — 182 с.

154. Воронов М.П., Усольцев В.А., Часовских В.П., Сенчило И.С. Лазарев. Н.В. Автоматизированная система определения и картирования депонируемого лесами углерода в среде СУБД ADABAS // Известия высших учебных заведений. Лесной журнал. 2013. №1 (331). URL: https://cyberleninka.ru/article/n/avtomatizirovannaya- sistema-opredeleniya-i-kartirovaniya-deponiruemogo-lesami-ugleroda-v-srede-subd-adabas (02.04.2025).

155. Деннет, Дэниел К. Виды психики: на пути к пониманию сознания / М., Идея-Пресс, 2004. – 183 с.

156. Гудфеллоу Я. Глубокое обучение / Ян Гидфеллоу, Иошуа Бенджио, Аарон Курвилль. — 2-е цветное издание, исправленное. — Москва : ДМК Пресс, 2018. — 651 с.

157. Роуз, Дэвид. «Будущее вещей: как сказка и фантастика становятся реальностью» (Enchanted objects). — 3-е изд. — Москва : Альпина нон-фикшн, 2017. — 343 с.

Информация об авторах

Часовских Виктор Петрович **—** профессор кафедры шахматного искусства и компьютерной математики Уральского государственного экономического университета, доктор технических наук, профессор
e-mail: u2007u@ya.ru

Усольцев Владимир Андреевич – профессор Уральского государственного лесотехнического университета, доктор сельскохозяйственных наук, профессор
e-mail: usoltsev50@mail.ru

Кох Елена Викторовна **—** доцент кафедры шахматного искусства и компьютерной математики Уральского государственного экономического университета, кандидат сельскохозяйственных наук, доцент
e-mail: elenakox@mail.ru

Оформление и верстка *Ю. Болдырева*

Дата подписания к использованию: 08.08.2025
Объем издания: 2,9 Мб. Комплектация: 1 электрон. опт. диск (CD-R)
Тираж 10 экз.

Издательство АНО ДПО «Межрегиональный центр инновационных технологий в образовании»
610047, г. Киров, ул. Свердлова, 32а, пом. 1003 Тел.: 8-800-222-30-98
https://mcito.ru/publishing; e-mail: book@mcito.ru

ISBN 978-5-907974-77-7