



УРАЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ  
ЭКОНОМИЧЕСКИЙ  
УНИВЕРСИТЕТ



КАФЕДРА  
ШАХМАТНОГО  
ИСКУССТВА  
И КОМПЬЮТЕРНОЙ  
МАТЕМАТИКИ

<https://vikchas.ru>

**XIX сессия «Синтез права и технологий: на пути к технологическому лидерству»  
4-5 июня 2026 года**

**«Основы будущего: фундаментальные проблемы генеративного  
ИИ»**

**«Правовые границы интеллекта регулирования генеративного ИИ  
в системе обеспечения национальной безопасности России»**

**Часовских Виктор Петрович**  
Профессор, д.т.н., Академик АВН и РАН

Екатеринбург 2026

ИИ не следует определять только через функции, подобные человеческим: обучение, рассуждение, решение задач, распознавание, генерацию.

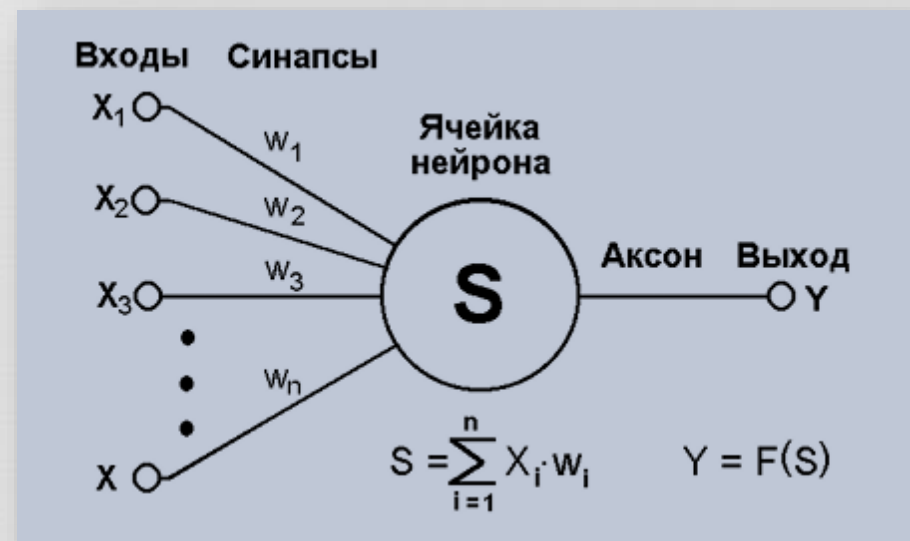
Такие определения описывают эффект, но не сущность технологии.

В реальности современный ИИ существует только как вычислительная система, включающая **ЭВМ, программное обеспечение и данные**.

Именно эти три компонента образуют его алгоритмическую основу, реализующую математические модели и, прежде всего, математическую модель нейрона.



**В**  
**электрохимических**  
**процессах**  
**НЕТ чисел**



**Всё сведено к числам и операциям**

**Электрохимические процессы**  
**(Na<sup>+</sup>, K<sup>+</sup>, Ca<sup>2+</sup> каналы)**

Выдающийся математик **Алан Мэтисон Тьюринг** в 1936 году создал **математическую модель** вычислительной машины, определил понятие **алгоритма**.

Используя модель Тьюринга, в 1945 году Джон фон Нейман, предложил архитектуру ЭВМ. Эта архитектура до сих пор лежит в основе большинства современных ЭВМ.

Тьюринг доказал, что существуют задачи, которые **нельзя решить алгоритмически**.

В настоящее время в Мире эксплуатируется более 22 миллиардов ЭВМ архитектуры фон Неймана, и на этих ЭВМ можно решать только алгоритмические задачи.

Результаты Тьюринга важны, поскольку они гарантированно определяют: будет ли любой алгоритм корректен; не возникнет ли ошибка; не выдаст ли система неправильный результат во всех возможных случаях, а это все усложняет ИИ – машинное обучение, нейронные сети, проекты ИИ.

Важным в проектах ИИ является математическая модель нейрона, а точнее **веса** входных переменных.

Веса формируются в процессе машинного обучения и они являются по сути экспериментальными данными контента машинного обучения.

Выдающиеся математик **В.Н. Вапник** создал фундаментальную часть теоретической базы машинного обучения. Основным положением является **невозможность восстановления исходной функции по экспериментальным данным**, а, как следствие, модель весов нейронов определяет достоверность результатов и практически 100% не достижимо.

Определена фундаментальная теорема **К.Ф. Гёделя**.

**«Формальная система не может быть полной и одновременно доказывать всю собственную корректность»** - алгоритм - это формальная система.

Теорема Гёделя означает, что проект ИИ не должен строиться на предположении о полной самодостаточности, полной доказуемости и абсолютной безошибочности системы.

# Математические ограничения алгоритмических программ

**Тьюринг** показывает пределы вычислений: не существует универсального алгоритма, который заранее гарантированно определит поведение любой программы на любом входе.

Для генеративного ИИ это значит: нельзя создать общий автоматический проверяющий механизм, который всегда заранее скажет, выдаст ли модель правильный, безопасный или бессмысленный ответ.

**Поведение сложной модели не полностью предсказуемо простыми средствами.**

**Вапник** описывает пределы обучения по данным. Модель учится на примерах, но хорошее качество на обучающей или тестовой выборке не гарантирует правильность в новых ситуациях. Для генеративного ИИ это особенно важно: он может уверенно отвечать там, где данные были похожи на обучающие, но ошибаться при новых, редких, неоднозначных или плохо представленных случаях.

**Поэтому качество ИИ — вероятностное, а не абсолютное.**

## Математические ограничения алгоритмических программ

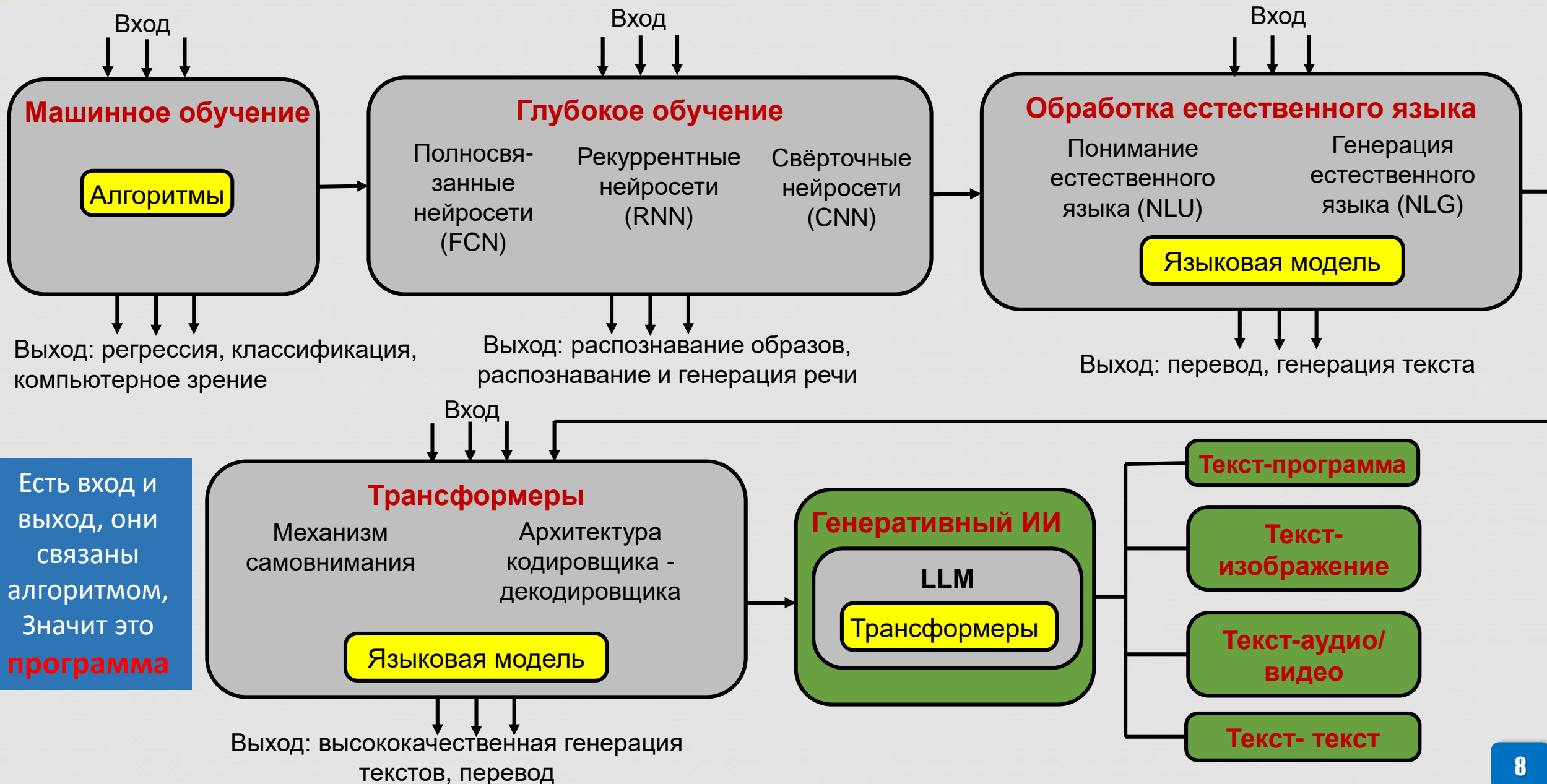
**Гёдель** показывает пределы формальных систем: в достаточно сложной системе есть истинные утверждения, которые нельзя доказать средствами самой системы.

Для генеративного ИИ это означает, что нельзя ожидать от него универсальной способности обосновать любую истину или полностью доказать собственную корректность. Даже если модель строит убедительные рассуждения, это не делает их автоматически истинными.

Поэтому достоверность ответов ИИ должна проверяться внешними способами: фактчекингом(проверка фактов), экспертной оценкой, тестированием, формальной верификацией для критичных задач и сравнением с надёжными источниками.

**Генеративный ИИ полезен как инструмент генерации, анализа и гипотез, но не как окончательный арбитр истины.**

# Появление генеративный ИИ



Есть вход и выход, они связаны алгоритмом, значит это программа

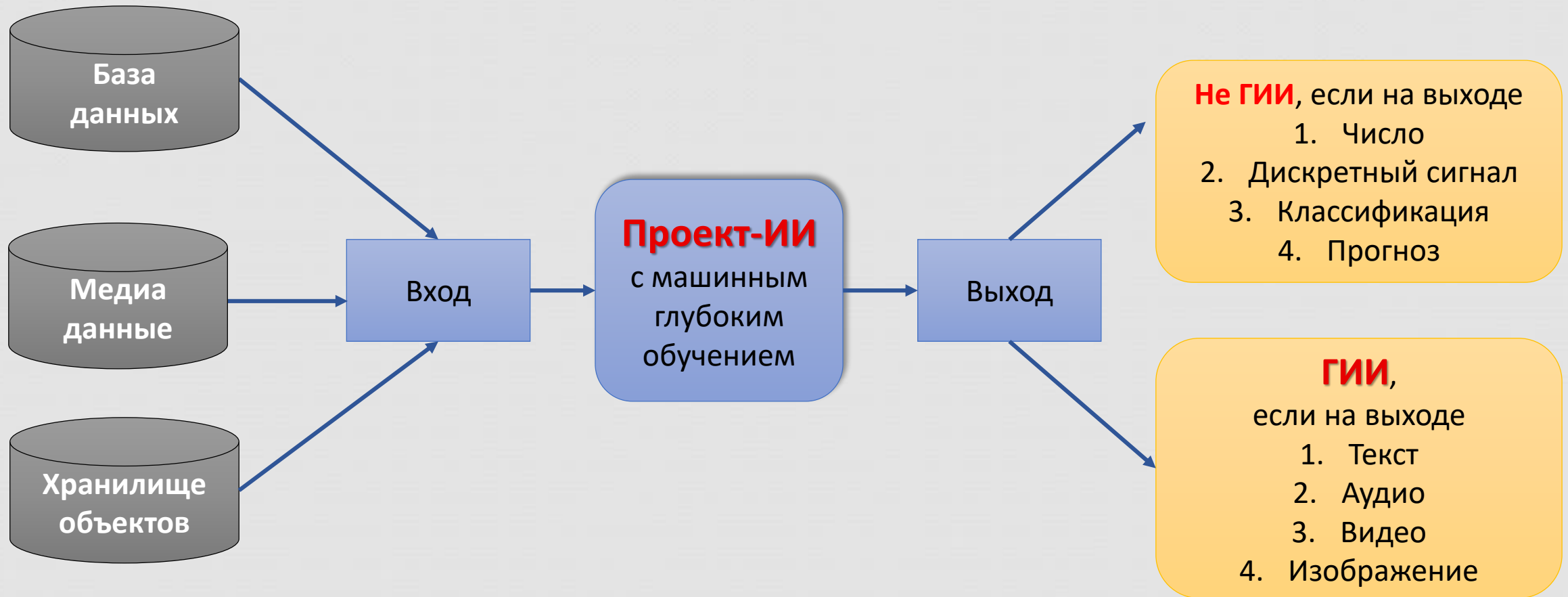
## ПОЯВЛЕНИЕ

Направление	Год появления	Комментарий
Машинное обучение	1959	Термин <i>Machine Learning</i> ввёл Артур Самуэль.
Глубокое обучение	1986	Идеи многослойных нейросетей активно развились после работ по backpropagation; термин <i>Deep Learning</i> стал широко использоваться позже, особенно с 2000-х.
Языковая модель	1948	Основы связаны с работами Клода Шеннона по статистическому моделированию языка и теории информации.
Трансформеры	2017	Архитектура Transformer описана в статье " <i>Attention Is All You Need</i> ".
Генеративный ИИ	1960-е	Ранние генеративные системы и эксперименты с ИИ появились уже в 1960-х, но термин <i>Generative AI</i> стал широко популярным в 2020-х.

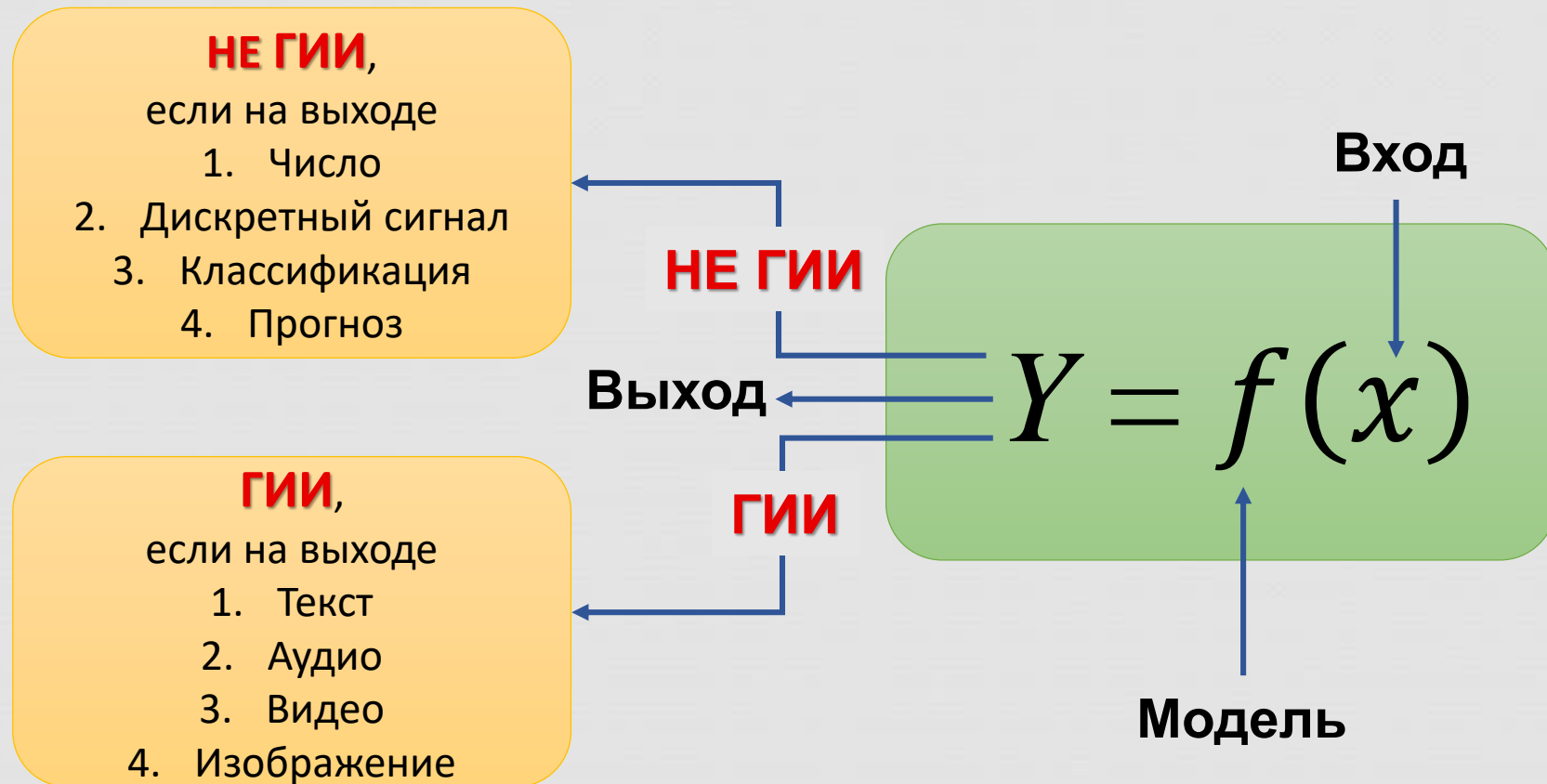
## ПРИМЕНЕНИЕ

Направление	Год применения	Комментарий
Машинное обучение	2010-е	Стало массово применяться в рекомендациях, рекламе, поиске, скоринге, аналитике данных.
Глубокое обучение	2012	Переломный момент — успех AlexNet в ImageNet; после этого deep learning резко распространился в компьютерном зрении, речи и NLP.
Языковые модели	2018–2019	Широкое применение после появления BERT, GPT и других крупных нейросетевых языковых моделей.
Трансформеры	2018–2020	После BERT и GPT трансформеры стали стандартом в NLP, затем распространились в компьютерное зрение и мультимодальные модели.
Генеративный ИИ	2022	Массовая популяризация после ChatGPT, Stable Diffusion, Midjourney и других генеративных систем.

# Что относится к ГИИ?



# Математическое представление



# Достоверность нейронных сетей после машинного обучения

После завершения обучения нейронных сетей их достоверность (точность, надежность) зависит от нескольких ключевых факторов:

## 1. Качество обучающих данных:

- ✓ репрезентативность выборки;
- ✓ отсутствие или минимизация шума и выбросов;
- ✓ корректная разметка данных.

## 2. Архитектура модели:

- ✓ сложность модели должна соответствовать сложности решаемой задачи;
- ✓ правильно подобранные гиперпараметры.

## 3. Метрики эффективности:

- ✓ на тренировочном наборе: обычно 95-99% для хорошо обученных моделей;
- ✓ на валидационном наборе: обычно 85-95%;
- ✓ на **тестовом наборе: обычно 80-95%** для хорошо обобщающих моделей.

## 4. Проблемы обучения:

- ✓ переобучение (overfitting): модель “запоминает” тренировочные данные вместо обобщения;
- ✓ недообучение (underfitting): модель слишком проста для задачи.

# Достоверность нейронных сетей при работе с реальными данными после обучения

Работа с реальными данными после обучения существенно отличается от процесса обучения:

## 1. Распределение данных:

- ✓ реальные данные могут отличаться от тренировочных (проблема смещения данных);
- ✓ достоверность снижается при работе с данными из “других распределений”.

## 2. Типичные показатели в реальных условиях:

- ✓ **как правило, 60-80% от достоверности на тестовом наборе;**
- ✓ может значительно снижаться со временем из-за изменения характеристик данных.

## 3. Факторы, влияющие на достоверность при инференсе (ранее обученную модель для принятия решений на основе новых данных.):

- ✓ доверительный уровень предсказаний;
- ✓ калибровка вероятностей;
- ✓ робастность к шуму и вариациям входных данных.

# Оценка достоверности ГИИ по данным опросов пользователей в 2025 г.

## *Результаты опросов ВШЭ (Россия)*

**Общая достоверность:** 70-75% пользователей отмечают приемлемый уровень достоверности ответов ГИИ.

**Научная информация:** достоверность оценивается в 65-70%.

**Фактическая информация:** около 60% респондентов сталкивались с фактическими ошибками.

**Доверие к источнику:** 55% пользователей проверяют информацию из ГИИ через дополнительные источники.

## *Результаты опросов ВВС (Великобритания)*

**Общественное восприятие:** 62% опрошенных британцев считают ответы ГИИ «в целом достоверными».

**Критическое отношение:** 78% признают необходимость проверки информации.

**Различие по возрастным группам:** молодежь (18-24 года) оценивает достоверность выше (около 70%), люди старше 55 лет — ниже (около 45%).

**Профессиональное использование:** 55% специалистов используют ГИИ в работе, но только 40% полностью доверяют получаемой информации.

## ОТНОШЕНИЕ УЧЕНЫХ К ИИ

Типы ученых в зависимости от отношения к ИИ (в процентах от численности респондентов)

Убежденные сторонники	17%	полностью принимают ИИ
Осознанные прагматики	16%	не слишком сильно доверяют ИИ, но считают, что он экономит время
Нейтральные	13%	не до конца определились с позицией, взвешивают положительные и отрицательные эффекты применения ИИ
Настороженные наблюдатели	13%	негативно оценивают влияние ИИ на научную деятельность, но считают, что их эта проблема не затрагивает
Скептики	11%	не доверяют ИИ
Противники	11%	выступают категорически против ИИ
Осторожные оптимисты	10%	одновременно воодушевлены потенциалом ИИ, но опасаются его применения
Обеспокоенные пессимисты	10%	негативно оценивают влияние ИИ на научную деятельность

## Что такое ГИИ на практике сейчас

Генеративный ИИ (GenAI) — это **вероятностная машина**, решающая задачи оптимизации:

1. Она ищет наиболее вероятный следующий токен (слово/пиксель) в последовательности.
2. Её «творчество» — это стохастический выбор менее вероятных вариантов (регулируется параметром «температуры» генерации).
3. «Знания» модели — это сжатое статистическое представление обучающих данных, а не семантическая сеть понятий.

Генеративный ИИ не является интеллектом в строгом смысле.

Это алгоритмическая технология обработки и генерации информации, ограниченная фундаментальными пределами вычислимости, доказуемости и статистического обобщения.

Поэтому его будущее — не замена человеческого интеллекта, а автоматизация отдельных видов человеческой деятельности.

**Это не делает ГИИ бесполезным или слабым инструментом.**

Напротив, он может быть очень мощным.

**Но его мощьность — это мощьность технологии, а не наличие собственного разума.**

На бытовом уровне говорят: «ИИ знает», «ИИ думает», «ИИ объясняет». Но это метафоры. Так же можно сказать «компьютер не хочет открывать файл» или «поисковик нашёл ответ». На строгом уровне компьютер ничего не хочет и не знает. Он выполняет операции

И снова. Искусственный интеллект — это не интеллект в человеческом смысле, а вычислительная технология. Он состоит из машины, программы и данных. Машина реализует физические процессы, программа задаёт алгоритм, данные служат материалом обработки.

Но ни один из этих элементов **не создаёт субъекта знания**. Поэтому ИИ может имитировать отдельные результаты интеллектуальной деятельности, но не обладает интеллектом, пониманием и априорными формами познания, присущими человеческому разуму.

Следовательно, ИИ может быть учебным инструментом, справочником, тренажёром, генератором заданий, помощником преподавателя. Но он **не может заменить учителя, потому что учитель — это не алгоритм, а мыслящий и ответственный субъект педагогической деятельности**.

## Преподаватель вуза и ИИ

Если ИИ **помогает** преподавателю, результат может быть **полезным**.

Если ИИ **заменяет** преподавателя, результат будет **опасным**.

Оптимальная модель такая:

ИИ должен быть **инструментом** в руках преподавателя, **а не заменой** преподавателя.

Потому что преподаватель — это субъект образования, а ИИ — средство обработки информации.

При замене преподавателя ИИ вуз может сохранить оболочку обучения, но потерять его человеческое, научное и воспитательное содержание.

Студент получит больше информации, но меньше школы мышления и компетенций.

# Проблемы менталитета и этики при внедрении генеративного ИИ

## Менталитет

**Некритичное доверие к ГИИ** - пользователи воспринимают ответы ГИИ как **истину**, не проверяя их. Это ведёт к атрофии критического мышления, особенно у молодёжи. Уверенный стиль генерации создаёт иллюзию компетентности там, где модель просто «галлюцинирует».

**Разрыв** между скоростью технологии и адаптацией людей  
Технология обновляется ежемесячно — менталитет меняется десятилетиями.  
Управленцы внедряют ИИ ради тренда, не понимая последствий.

**Недооценка** природы инструмента  
В отличие от обычных инструментов, ГИИ **генерирует** смыслы, имитирует суждение и влияет на убеждения.

**Пассивность** и позиция «нас не касается»  
Руководство перекладывает ответственность на IT, юристы ждут законов, общество не участвует в обсуждении.

Возникает ситуация – **ГИИ принимает решение.**

# Проблемы менталитета и этики при внедрении генеративного ИИ

## Этика

### **Авторство и интеллектуальная собственность**

ГИИ обучается на текстах, изображениях и коде без согласия авторов. Возникают острые вопросы: кто является автором сгенерированного произведения, имеют ли правообладатели право на компенсацию.

### **Дипфейки и подрыв доверия к реальности**

Технология позволяет создавать неотличимый фейковый контент — видео, голос, «доказательства».

### **Встроенная предвзятость**

ГИИ обучается на данных, созданных людьми, и воспроизводит исторические предрассудки — расовые, гендерные, классовые — в масштабируемом виде. Предвзятость становится системной.

### **Размытие ответственности**

При ошибке ГИИ неясно, кто виноват: разработчик модели, компания-интегратор, пользователь или организация. Традиционные правовые механизмы не рассчитаны на такой сценарий.

### **Манипуляция и информационная среда**

ГИИ позволяет генерировать персонализированный убеждающий контент в промышленных масштабах. Это создаёт инструмент для массового политического и коммерческого влияния, распространения дезинформации и разрушения общего информационного пространства.

## **Проблемы менталитета и этики при внедрении генеративного ИИ**

**Технические возможности ГИИ опережают как готовность общества их осмыслить, так и готовность институтов их регулировать. Без целенаправленной работы с менталитетом и этическими рамками внедрение ГИИ воспроизводит и усиливает уже существующие социальные проблемы.**

**Этика внедрения ИИ — это не технический вопрос. Это вопрос о том, какое общество мы строим и готовы ли мы брать за это ответственность.**

# Ментальная война

## Ментальная (когнитивная) война

это целенаправленное воздействие на сознание, убеждения, эмоции и поведение людей без применения физической силы.

Цель не уничтожить противника, а изменить то, как он думает, что считает правдой и как принимает решения.

Площадь 200 га

Охлаждение 2 мл.  
литров

Электроэнергия  
0,005 ГВт

Область  
0,000113279 ГВт



Называть ГИИ «средством ментальной войны» — не преувеличение, а точная характеристика одного из его применений. Технология нейтральна, но её возможности делают её идеальным инструментом влияния на массовое сознание.

Главная линия фронта сегодня проходит не по территории, а через сознание людей — и ГИИ является на этом фронте оружием с беспрецедентными характеристиками.

## Правовые границы интеллекта регулирования генеративного ИИ в системе обеспечения национальной безопасности России

Правовые границы регулирования генеративного ИИ в системе обеспечения национальной безопасности России должны определяться балансом между:

- ✓ технологическим развитием;
- ✓ защитой прав граждан;
- ✓ предотвращением угроз государству.

Генеративный ИИ способен создавать тексты, изображения, аудио и видео, что открывает как значительные возможности, так и риски: распространение дезинформации, дипфейки, кибератаки, манипулирование общественным мнением, утечки персональных данных и использование ИИ в противоправных целях.

# Правовое регулирование должно включать:

- 1. Определение правового статуса генеративного ИИ и ответственности за результаты его использования;**
- 2. Защиту персональных данных и государственной тайны при обучении и применении ИИ-систем;**
- 3. Требования к прозрачности и маркировке ИИ-контента, особенно в политической, информационной и общественно значимой сфере;**
- 4. Ограничения на использование ИИ в целях, угрожающих национальной безопасности, включая создание фейков, вредоносного кода и информационно-психологическое воздействие;**
- 5. Установление ответственности разработчиков, владельцев и пользователей ИИ, если применение технологии повлекло вред;**
- 6. Государственный контроль и экспертную оценку ИИ-систем, применяемых в критической информационной инфраструктуре и сфере безопасности.**

Правовые границы не должны быть чрезмерно жёсткими, чтобы не тормозить развитие отечественных технологий. Оптимальная модель — **риск-ориентированное регулирование**, при котором наиболее строгие требования применяются к тем ИИ-системам, которые способны причинить существенный вред личности, обществу или государству.

Таким образом, регулирование генеративного ИИ в сфере национальной безопасности России должно сочетать:

- ✓ правовые запреты;
- ✓ технические стандарты;
- ✓ контроль над оборотом данных;
- ✓ механизмы ответственности.

Особое значение имеет закрепление в правовом регулировании положения о том, что результаты работы генеративного ИИ не обладают гарантированной стопроцентной достоверностью.

Генеративные модели формируют выводы на основе вероятностных алгоритмов и обучающих данных, поэтому могут допускать ошибки, искажения, “галлюцинации”, воспроизводить недостоверную информацию или скрытые предвзятости.

В связи с этим, решения, сгенерированные ИИ, особенно в сфере обеспечения национальной безопасности, не должны использоваться как единственное и окончательное основание для принятия юридически значимых или управленческих решений без проверки человеком и независимой экспертной оценки.

В целях минимизации рисков использования генеративного ИИ в системе обеспечения национальной безопасности целесообразно предусмотреть обязательную сертификацию таких систем.

При этом речь должна идти не о подтверждении абсолютной достоверности решений ГИИ, а о выдаче сертификата соответствия установленным требованиям к надёжности, информационной безопасности, качеству обучающих данных, устойчивости к манипуляциям, уровню ошибок и наличию механизмов человеческого контроля. Такой сертификат должен подтверждать, что система прошла независимую проверку и может применяться только в определённых сферах и при соблюдении установленных ограничений.

## Предложение к структуре сертификата. Он должен включать:

1. Оценку качества обучающих данных.
2. Проверку уровня ошибок и “галлюцинаций”.
3. Тестирование устойчивости к внешним воздействиям и манипуляциям.
4. Проверку кибербезопасности модели.
5. Оценку рисков утечки персональных данных, гостайны и служебной информации.
6. Наличие журнала действий и возможности аудита.
7. Обязательность человеческого контроля за итоговым решением.

# Действующие правовые основы регулирования ИИ в России

1. Указ Президента Российской Федерации от 10.10.2019 № 490 «О развитии 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. ФЗ Об основах государственного регулирования сфер применения технологий искусственного интеллекта в Российской Федерации

# Благодарю за внимание!



# Приложения

В Большой российской энциклопедии (БРЭ, издание 2004–2017 годов) информатика определяется как «наука об извлечении информации из сообщений, создании информационных ресурсов, программировании поведения машин, о других сущностях, связанных с построением и применением человеко-машинной среды решения задач моделирования, проектирования, взаимодействия, обучения и др.».

## Научные специальности ВАК:

1. Естественные науки

1.2. Компьютерные науки и информатика – группа специальностей

**1.2.1. Искусственный интеллект и машинное обучение** - специальность

1.2.2. Математическое моделирование, численные методы и комплексы программ

1.2.3. Теоретическая информатика

1.2.4. Кибербезопасность

**ИИ – это специальность** группы специальности «Компьютерные науки и информатика», области науки «Естественные науки».

## Указ Президента РФ от 10.10.2019 г. № 490

**Искусственный интеллект** - комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая поиск решений без заранее заданного алгоритма), и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их. 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;

### **Текущее состояние**

На сегодняшний день **сильный ИИ** остаётся теоретической концепцией. Современные системы (включая продвинутые нейросети вроде GPT) относятся к слабому ИИ: они хорошо справляются с отдельными задачами, но не обладают универсальностью, самосознанием и способностью к подлинному обобщённому мышлению. Учёные предполагают, что создание AGI может занять десятки или даже сотни лет — либо оказаться принципиально недостижимым.

## **ФЕДЕРАЛЬНЫЙ ЗАКОН**

# **Об основах государственного регулирования сфер применения технологий искусственного интеллекта в Российской Федерации (26 страниц)**

**Настоящий Федеральный закон вступает в силу с 1 сентября 2027.**