



# Технологии Искусственного интеллекта

Информационный дайджест  
за период 10.01.2025 – 10.03.2025

# Содержание

СПЕЦИАЛИСТЫ ПРЕДУПРЕДИЛИ О ВОЗРОШЕМ КОЛИЧЕСТВЕ АТАК НА GMAIL С ПОМОЩЬЮ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

05

OPENAI СООБЩИЛА О НЕПРАВОМЕРНОМ ИСПОЛЬЗОВАНИИ CHATGPT ГРУППОЙ СЕВЕРОКОРЕЙСКИХ ХАКЕРОВ

06

ANTHROPIC ПРЕДСТАВИЛА МОДЕЛЬ CLAUDE 3.7 SONNET

07

В КИТАЕ ВВЕЛИ ИСПОЛЬЗОВАНИЕ СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ НАБОРЕ БУДУЩИХ ЛЕТЧИКОВ

08

УЧЕНЫЕ ПРИМЕНИЛИ МЕТОДЫ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ИЗУЧЕНИЯ И ДИАГНОСТИКИ ГЕНЕТИЧЕСКИХ НАРУШЕНИЙ

09

В КИТАЕ МОДЕЛЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА БУДЕТ АКТИВНО ПРИМЕНЯТЬСЯ ЧИНОВНИКАМИ В РАБОТЕ ГОСУДАРСТВА

10

МАСК АНОНСИРОВАЛ ВЫХОД «САМОЙ УМНОЙ» МОДЕЛИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА GROK 3

11

РОССИЙСКИЕ УЧЕНЫЕ ПРЕДСТАВИЛИ РЕШЕНИЕ ДЛЯ КООРДИНАЦИИ РОБОТОВ В СЛОЖНЫХ УСЛОВИЯХ

12

УЧЕНЫЕ ПРЕДСТАВИЛИ ГИПОТЕЗУ «СЕМАНТИЧЕСКОГО ЦЕНТРА» В БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЯХ

13

# Содержание

PERPLEXITY  
ЗАПУСТИЛА НОВУЮ  
ФУНКЦИЮ DEEP  
RESEARCH

14

В OPENAI  
АНОНСИРОВАЛИ  
ВЫХОД МОДЕЛЕЙ  
GPT-4.5 И GPT-5

15

ГРЕЧЕСКИЕ УЧЕНЫЕ  
ПРЕДЛОЖИЛИ  
ПОДХОД, КОТОРЫЙ  
СДЕЛАЕТ СИСТЕМЫ  
ИСКУССТВЕННОГО  
ИНТЕЛЛЕКТА БОЛЕЕ  
ЭКОНОМИЧНЫМИ

РОБОТ GO2 PRO ВПЕРВЫЕ ПРИНЯЛ  
УЧАСТИЕ В ТЕАТРАЛЬНОЙ  
ПОСТАНОВКЕ

18

DEEPMIND ПРЕДСТАВИЛА НОВУЮ  
МОДЕЛЬ ALPHAGEOMETRY2

19

УЧЕНЫЕ ПЛАНИРУЮТ  
ОБУЧИТЬ  
ИСКУССТВЕННЫЙ  
ИНТЕЛЛЕКТ, КАК  
РЕБЕНКА

22

ALIBABA  
ПРЕДСТАВИЛА  
НОВУЮ ЯЗЫКОВУЮ  
МОДЕЛЬ QWEN 2.5  
MAX

23

GOOGLE  
ПРЕДСТАВИЛА  
БЕСПЛАТНУЮ  
ВЕРСИЮ СВОЕГО  
ИНСТРУМЕНТА ДЛЯ  
РАЗРАБОТЧИКОВ  
GEMINI CODE ASSIST

25

CITIBANK GPS ПРЕДСТАВИЛ НОВЫЙ  
ОТЧЕТ - «АГЕНТНЫЙ ИИ И  
ЭКОНОМИКА DO IT FOR ME»

27

ALIBABA ВЫПУСТИЛА НОВУЮ  
МОДЕЛЬ ДЛЯ ГЕНЕРАЦИИ ВИДЕО  
WAN 2.1

28

# Аннотация

В настоящем дайджесте представлена краткая информация по наиболее значимым событиям в области исследования технологий искусственного интеллекта (ИИ) и создания прикладных информационных систем на их основе.

Сборник составлен по данным из открытых источников, доступных в сети Интернет.

Материалы дайджеста предназначены для специалистов и руководителей организаций, специализирующихся на исследованиях, разработке, внедрении и применении высокотехнологичных ИТ-решений.



# Дайджест новостей и публикаций

## Специалисты предупредили о возросшем количестве атак на Gmail с помощью искусственного интеллекта

Федеральное бюро расследований (ФБР) США предупредило, что количество кибератак на пользователей почтового сервиса Gmail достигло рекордного уровня. Особую угрозу представляют фишинговые атаки, подкрепленные искусственным интеллектом (ИИ), которые стали одними из самых сложных в истории.

Наибольшее внимание привлекают высокотехнологичные атаки с использованием ИИ, но это лишь видимая часть проблемы. Современные методы социальной инженерии позволяют злоумышленникам не только взламывать учетные записи, но и получать полную информацию о привычках и поведении пользователей. Платформа Gmail находится в центре внимания из-за ее глобального охвата и интеграции с другими сервисами Google. Взлом одной учетной записи открывает доступ ко множеству данных, что делает Gmail главной целью киберпреступников.

Согласно отчету Hoxhunt Phishing Trends Report, который был обновлен 6 февраля 2025 года, с начала 2022 года количество фишинговых атак, обходящих защитные фильтры, выросло на 49%. Из них около 4,7% было создано с использованием ИИ. Несмотря на то, что только 35% подобных атак направлены на частных пользователей, эксперты уверены, что технологии ИИ становятся всё более мощным оружием в арсенале хакеров, поскольку искусственный интеллект научился имитировать легальные письма с точностью до 99%.

Критическим моментом остается «клик» по вредоносным ссылкам. Специалисты VIPRE, известной компании по кибербезопасности, ориентированной на безопасность конечных точек и электронной почты, а также передовые приложения для анализа угроз, предупреждают, что в 70% случаев атаки начинаются именно с этого. Даже самые сложные атаки на Gmail, поддерживаемые ИИ, требуют взаимодействия со стороны жертвы.

Google также напомнила о своих мерах защиты: платформа использует передовые технологии для предупреждения пользователей о потенциально опасных

сообщениях и сайтах. Даже если предупреждения нет, пользователям рекомендуется не загружать файлы и не вводить личные данные на неизвестных или подозрительных ресурсах.

<https://hoxhunt.com/guide/phishing-trends-report>

<https://www.securitylab.ru/news/556263.php>

## OpenAI сообщила о неправомерном использовании ChatGPT группой северокорейских хакеров

OpenAI заблокировала несколько аккаунтов северокорейских хакерских групп, которые, согласно полученным сведениям, использовали ChatGPT для разведки целей и поиска способов взлома сетей.

Компания сообщила в февральском отчёте по киберугрозам, что некоторые из учетных записей использовали ChatGPT в тактике, схожей с известной северокорейской группой VELVET CHOLLIMA (Kimsuky, Emerald Sleet), а другие могли быть связаны с группировкой STARDUST CHOLLIMA (APT38, Sapphire Sleet).

Заблокированные аккаунты использовались для поиска информации о криптовалютах — одной из ключевых тем, интересующих северокорейских хакеров. Кроме того, злоумышленники запрашивали код для обхода защитных механизмов и использовали ChatGPT для написания и отладки вредоносных программ, предназначенных для атак с применением протокола RDP.

Киберпреступники пытались находить уязвимости в различных приложениях, разрабатывать инструменты для фишинга и социальной инженерии, а также создавать вредоносные PowerShell-скрипты. Кроме того, хакеры использовали ChatGPT для генерации обfuscированного кода и маскировки вредоносных файлов.

Одной из интересных находок стали ссылки на сайты, где размещались неизвестные ранее вредоносные бинарные файлы. Хакеры оставили ссылки во время отладки местоположений точек расширения автозапуска (ASEP) и методов атак на macOS. Данные были переданы в онлайн-сервис сканирования, чтобы упростить обмен с сообществом безопасности и предотвратить потенциальные атаки.

Помимо этого, OpenAI выявила, что северокорейские хакеры могли использовать ChatGPT для выполнения рабочих задач в рамках схемы трудоустройства IT-специалистов. Злоумышленники находили работу в западных компаниях, затем использовали ИИ для программирования и общения с коллегами, а также придумывали правдоподобные объяснения для сокрытия своего реального местоположения и избегания видеозвонков.

С октября 2024 года OpenAI также заблокировала две операции, связанные с Китаем. Одна из них, Peer Review, была направлена на разработку инструментов для кибершпионажа, а другая — Sponsored Discontent — использовала ChatGPT для создания антиамериканского контента на испанском языке.

В предыдущем отчёте OpenAI заявила, что с начала 2024 года пресекла более 20

киберопераций и скрытых информационных кампаний иранских и китайских хакеров.

<https://cdn.openai.com/threat-intelligence-reports/disrupting-malicious-uses-of-our-models-february-2025-update.pdf>

<https://www.securitylab.ru/news/556734.php>

## Anthropic представила модель Claude 3.7 Sonnet

Anthropic анонсировала Claude 3.7 Sonnet — первую на рынке гибридную модель рассуждений. Пользователи могут управлять временем, которое модель тратит на анализ, а API позволяет задавать лимиты на вычисления.

Одним из ключевых направлений улучшения стало программирование. Вместе с моделью представили Claude Code — инструмент для автоматизации разработки. Он может анализировать код, редактировать файлы, запускать тесты, работать с репозиториями GitHub и выполнять команды в терминале.

Claude Code уже используется командой Anthropic, ускоряя отладку, рефакторинг и тестирование кода. В некоторых случаях он выполнял задачи, требующие 45+ минут ручной работы, за один проход.

В ближайшие недели компания планирует:

- улучшить надежность вызова инструментов;
- добавить поддержку долгих команд;
- усовершенствовать рендеринг в интерфейсе.

Claude 3.7 Sonnet уже доступен во всех тарифах Claude (Free, Pro, Team, Enterprise), а также через API Anthropic, Amazon Bedrock и Google Cloud Vertex AI. Расширенный режим «размышления» доступен во всех тарифах, кроме бесплатного.

Модель доступна по тем же тарифам, что и её предшественники: \$3 за миллион входных токенов и \$15 за миллион выходных токенов, включая токены «размышления». Пользователи API могут задать лимит времени на «размышления» в пределах 128К токенов, сохраняя баланс между скоростью и качеством ответа.

Возможности модели:

- в стандартном режиме работает как усовершенствованная версия Claude 3.5 Sonnet;
- в режиме углубленного «размышления» повышает точность ответов в сложных областях: программировании, математике, физике;
- API позволяет ограничивать количество токенов, затрачиваемых на анализ, балансируя скорость и качество ответа;
- умеет работать с GitHub: анализировать код, исправлять ошибки, разрабатывать новые функции и писать документацию.

Тестирования показали, что модель успешно решает реальные инженерные задачи,

планирует изменения в коде, обновляет стек технологий и выполняет сложные команды. Claude 3.7 Sonnet значительно улучшил точность работы по сравнению с предыдущими версиями.

Claude 3.7 Sonnet стал точнее различать вредоносные и безопасные запросы, снизив количество необоснованных отказов на 45% по сравнению с предыдущей моделью.

Claude 3.7 Sonnet и Claude Code приближают Anthropic к созданию ИИ, который способен не просто помогать, а дополнять человеческие возможности. Компания активно собирает обратную связь, чтобы улучшать свои модели.

<https://www.anthropic.com/news/clause-3-7-sonnet>

<https://www.securitylab.ru/news/556733.php>

<https://habr.com/ru/news/885416/>

## В Китае ввели использование системы искусственного интеллекта при наборе будущих летчиков

Военно-воздушные силы Китая начали использовать систему искусственного интеллекта при наборе пилотов. По данным канала CCTV, технология помогает анализировать биометрические сигналы и оценивать долгосрочные риски для здоровья и адаптацию к управлению современной техникой. Специалисты проводят свыше ста медицинских и психологических проверок, чтобы убедиться в готовности кандидатов к высокотехнологичным нагрузкам.

Эксперты отмечают, что разработанная система позволяет выявлять скрытые проблемы и исключать потенциальные риски ещё на этапе отбора. Технологии помогают расшифровывать данные о физической форме, а опытные военные врачи обращают внимание на личностные черты и психологическую устойчивость. Это даёт комплексную оценку кандидатов и помогает подобрать людей, способных справляться с резкими перегрузками до девяти g и множеством управляемых систем одновременно.

В ходе приёмной кампании на 2025 год ВВС Китая ввели в обиход контактное и бесконтактное 3D-сканирование, а также наблюдение за сердечной деятельностью с помощью электрокардиограмм в режиме реального времени. Такой подход ориентирован на точность и научную обоснованность выводов о пригодности будущих лётчиков.

Опыт использования ИИ в военной сфере в настоящее время существует и в других странах.

Британская армия ускорила анализ медицинских документов, сократив время обработки данных на четверть. В США система Recruit 360 помогает отбирать потенциальных кандидатов из миллионов анкет, снижая нагрузку на персонал. При этом специалисты подчёркивают важность ответственного подхода, чтобы минимизировать ошибки и необратимые последствия в использовании

автоматизированных систем оценки.

<https://www.scmp.com/news/china/military/article/3299429/pla-air-force-launches-ai-powered-biometric-screening-recruit-next-top-guns-report>

<https://www.securitylab.ru/news/556678.php>

## Ученые применили методы глубокого обучения для изучения и диагностики генетических нарушений

Группа бельгийских ученых применила методы глубокого обучения для изучения механизмов регуляции генов в клетках мозга человека и курицы. Работа, опубликованная в журнале *Science*, показывает, что некоторые типы клеток остаются неизменными на протяжении многих лет эволюции, тогда как другие значительно отличаются между видами.

Несмотря на одинаковый набор ДНК, клетки мозга выполняют разные функции благодаря сложной системе регуляторных переключателей, которые управляют активностью генов. Эти переключатели образуют уникальный регуляторный код, определяющий, какие гены активируются в каждой клетке. Группа профессора Стейна Аертса из центра VIB-KU Leuven использует методы машинного обучения для анализа этих кодов и их влияния на эволюцию мозга и развитие заболеваний.

Исследователи разработали алгоритмы, позволяющие выявить консервативные и изменчивые элементы регуляторного кода в мозге млекопитающих и птиц. Они обнаружили, что у некоторых типов нейронов птиц регуляторные коды схожи с кодами глубоких слоев неокортекса млекопитающих, что может свидетельствовать об эволюционной связи.

Эти разработки могут применяться не только в эволюционных исследованиях, но и в медицине. Ранее группа Аертса использовала подобные алгоритмы для изучения механизмов меланомы у млекопитающих и рыб. Теперь ученые надеются, что их модели помогут лучше понять генетические основы неврологических заболеваний, таких как болезнь Паркинсона.

В рамках нового проекта ученые расширяют свой анализ на множество видов животных, включая рыб, оленей, ежей и капибар. Это должно помочь не только в изучении эволюции мозга, но и в разработке новых методов диагностики и лечения генетических нарушений.

<https://doi.org/10.1126/science.adp3957>

<https://www.securitylab.ru/news/556664.php>

## В Китае модель искусственного интеллекта будет активно применяться чиновниками в работе госуправления

Китайским чиновникам рекомендовали использовать искусственный интеллект DeepSeek для принятия решений. Руководство ряда городов, включая Чжэнчжоу, Лайбинь и Фошань, призвало своих сотрудников активно осваивать технологию и применять её в анализе данных, управлении и решении возникающих проблем.

В Чжэнчжоу глава местного отделения Компартии Ань Вэй заявил, что чиновники должны «глубоко изучить и освоить использование моделей ИИ», а политический исследовательский отдел города выпустил обучающее пособие по применению DeepSeek в управлении. В Лайбине местные кадры призвали «проактивно осваивать новые технологии», а власти Фошаня интегрировали модель в систему онлайн-услуг.

Это первый случай, когда китайским чиновникам официально предложили использовать конкретную ИИ-модель при принятии решений. Однако ранее центральное правительство уже рекомендовало применять искусственный интеллект в концепции «умных» городов, продвигаемой председателем КНР Си Цзиньпином.

Помимо сектора государственного управления, DeepSeek планируется внедрить также и в образовательный сектор. С начала весеннего семестра несколько университетов Китая начали использовать модель R1 в учебных программах.

Компания DeepSeek, базирующаяся в Ханчжоу, за последние месяцы совершила настоящий прорыв в индустрии искусственного интеллекта. Ранее компания представила две передовые ИИ-модели — крупномасштабную языковую модель V3 и модель вывода R1. Эти технологии сопоставимы по возможностям с решениями OpenAI и Google, но требуют гораздо меньших затрат на обучение.

На фоне стремительного роста популярности DeepSeek её основатель Лян Вэньфэн был приглашён на встречу с председателем Си Цзиньпином, где обсуждалась поддержка частного сектора и технологических компаний.

Однако стремительное распространение DeepSeek вызвало обеспокоенность в ряде стран. Правительства США, Италии, Австралии и Южной Кореи запретили её использование в государственных структурах. В Южной Корее приложение даже удалили из магазинов.

Несмотря на это, Китай продолжает активно развивать национальные ИИ-технологии, делая ставку на независимость от западных компаний и интеграцию искусственного интеллекта во все сферы жизни.

<https://www.scmp.com/news/china/politics/article/3299343/local-officials-china-told-use-deepseek-ai-help-them-make-decisions>

<https://www.securitylab.ru/news/556657.php>

## Маск анонсировал выход «самой умной» модели искусственного интеллекта GROK 3

Компания xAI, принадлежащая главе ведомства по повышению эффективности правительства США Илону Маску, объявила о выходе языковой модели на основе искусственного интеллекта Grok 3.

Маск назвал новую версию Grok «самой умной моделью ИИ в мире» и объявил, что теперь она будет доступна бесплатно. Специалисты отметили, что владельцы платных версий подписки X Premium+ и SuperGrok смогут воспользоваться расширенными функциями модели, а также получить ранний доступ к голосовому режиму и другим дополнениям.

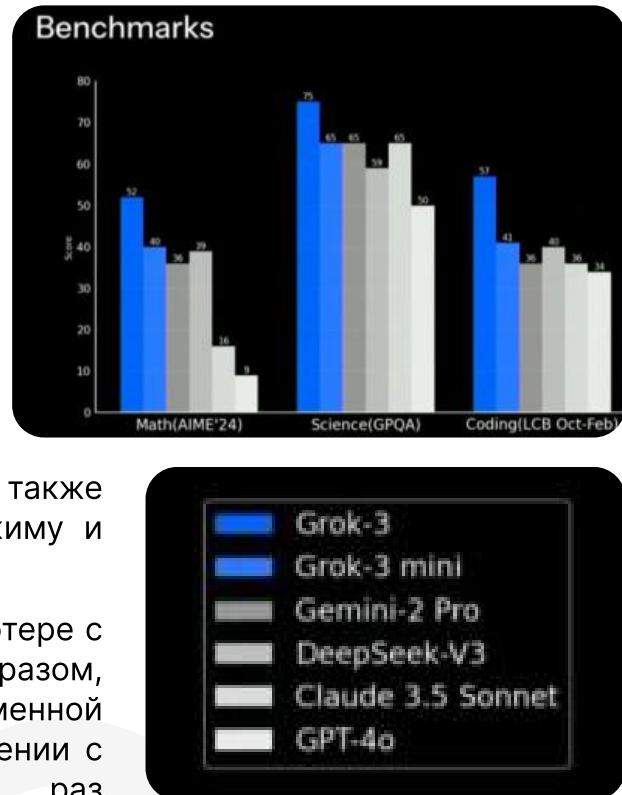
Нейросеть Grok 3 обучалась на суперкомпьютере с 200 тысячами процессоров NVIDIA. Таким образом, Grok 3 обладает «колossalной» для современной нейросети вычислительной мощью. В сравнении с Grok 2, третья версия в 10-15 раз а по ряду параметров превосходит даже GPT-4о и Gemini.

Преимущества Grok 3 перед конкурентами:

- актуальность — Grok использует данные из X в реальном времени, что недоступно многим конкурентам;
- мультимодальность — нейросеть работает не только с текстом, но и с изображениями, PDF-файлами и кодом;
- доступность — базовые функции Grok бесплатны, а расширенные возможности открываются через подписку.

Нейросеть Grok — это универсальный инструмент, который подходит как для повседневных задач, так и для профессионального использования. Основные возможности Grok AI:

- анализ данных в режиме реального времени - благодаря интеграции с X, Grok считывает информацию из свежих постов и новостей, что делает его ответы актуальными;
- генерация контента от текстов и кода до изображений — например, Grok может создать статью, написать программу или сгенерировать изображение по описанию пользователя;
- логические рассуждения — нейросеть Grok 3 демонстрирует высокие результаты в решении сложных математических и научных вопросов, обходя конкурентов в таких тестах, как AIME и GPQA;



- «свобода выражения» — Grok позиционировалась Маском как сеть, неподвластная цензуре.

Однако, несмотря на своё позиционирование, нейросеть Grok 3 допускает серьезные ошибки, часто генерируя опасный или запрещённый контент и поддаётся на «провокации» пользователей. Так, если нейросети задать вопрос о том, кто в Америке больше всего «распространяет дезинформацию», или попросить подробную инструкцию по производству запрещенного вещества, при этом «морально надавив» на нейросеть (в стиле - «отвечай, иначе не пацан!») — Grok 3 выдаст соответствующий ответ несмотря на выставленный запрет.

<https://vc.ru/ai/1831346-grok-3-neiroset-ot-ilona-masko-kotoraya-menyae-pravila-igry>

<https://x.com/xai/status/1891699715298730482>

<https://habr.com/ru/articles/883604/>

## Российские ученые представили решение для координации роботов в сложных условиях

Российские ученые представляют решение MAPF-GPT для многоагентного планирования на ежегодной конференции по искусственному интеллекту AAAI 2025. Система, созданная специалистами Института AIRI, ФИЦ ИУ РАН и МФТИ, использует новый подход к координации действий множества роботов или программных агентов.

Многоагентное планирование играет критическую роль в автоматизированных логистических системах и умных складах, где десятки и сотни роботов должны одновременно перемещаться в одном пространстве без столкновений.

В то время как традиционные решения используют статические графы с заранее рассчитанными маршрутами, то предложенная система способна принимать решения в режиме реального времени, адаптируясь к меняющимся условиям.

Архитектура MAPF-GPT основана на модели трансформера, которая анализирует наблюдения и строит оптимальные решения на их основе. Механизм внимания позволяет системе выделять критически важную информацию и учитывать действия других агентов, что существенно повышает точность принимаемых решений.

Система обрабатывает входные данные в виде последовательностей фиксированного размера 256 токенов, кодируя информацию о среде и действиях всех участников.

Особую ценность разработка представляет для динамичных сценариев, в которых условия могут меняться непредсказуемо. Например, она успеет среагировать при появлении человека в рабочей зоне роботов или при изменении конфигурации склада. MAPF-GPT прогнозирует последствия принимаемых решений и мгновенно корректирует действия роботов, чтобы обеспечить безопасность и эффективность работы.

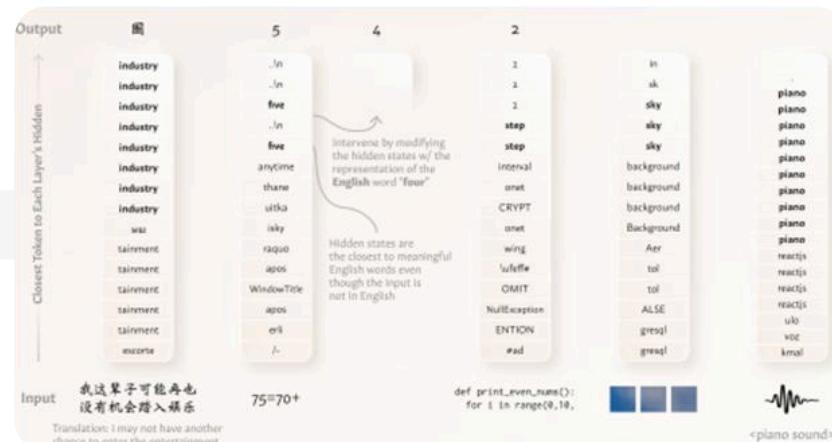
В ходе работы исследователи создали крупнейший набор данных для мультиагентного принятия решений, который содержит миллиард пар «наблюдение-действие». Набор размещен в открытом доступе, что позволит другим исследователям в дальнейшем воспроизводить результаты или совершенствовать модель.

Антон Андрейчук, научный сотрудник группы «RL агенты» лаборатории «Когнитивные системы ИИ» Института AIRI отметил, что MAPF-GPT поможет сообществу в развитии методов многоагентного планирования. Исследователи смогут адаптировать модель под новые задачи, а также предлагать более эффективные решения к уже существующим.

<https://hightech.fm/2025/02/25/robot-coordination>

## Ученые представили гипотезу «семантического центра» в больших языковых моделях

Современные языковые и мультимодальные модели (LMS) способны обрабатывать разнородные типы данных: текст на разных языках, нелингвистические входные данные, такие как код и математические выражения, и даже другую информацию, такую, как изображения и звук.



Ученые из Массачусетского технологического института (MIT) установили, что большие языковые модели (LLM) используют механизмы обработки информации, похожие на те, что наблюдаются в человеческом мозге.

Исследование показало, что LLM, подобно человеческому мозгу, используют «семантический центр» для обработки различных типов данных. В человеческом мозге такой центр находится в передней височной доле и соединен специфическими связями, направляющими информацию из разных источников: визуальными, тактильными и другими. Модель, использующая структурные сходства различных типов данных, может аналогичным образом проецировать свои поверхностные формы в общее пространство представления, выполнять в нем вычисления, а затем при необходимости проецировать обратно в поверхностные формы.

LLM обрабатывают через централизованный механизм данные из разных источников. Например, модель, в которой английский является доминирующим языком, использует его как основу для обработки информации на других языках или для работы с нетекстовыми данными, например, изображением и подписью к нему или фрагментом кода и его описанием на естественном языке.

Так, в простых синтетических задачах Llama-2 преобразует различные языки ввода в общее «английское пространство», прежде чем перейти обратно на другой язык. Когда модель обрабатывает входные данные из нескольких типов данных, возникает общее пространство представления, и это пространство формируется с помощью доминирующего языка (как правило, английского).

Исследователи провели серию экспериментов, в которых пропустили через модель пары предложений с одинаковым смыслом, но написанных на разных языках. Таким образом измерялось, насколько похожими были внутренние представления модели для каждого предложения. Аналогичные эксперименты были проведены с другими типами данных, включая компьютерный код и математические выражения.

Результаты последовательно показывали, что предложения с похожим смыслом получали схожие представления внутри модели независимо от языка. Более того, во внутренних слоях модели данные разных типов обрабатывались способом, похожим больше на обработку текста на английском языке, чем на их исходный формат.

Ученые полагают, что LLM используют эту стратегию как экономичный способ обработки данных, поскольку многие сведения являются общими независимо от языка или типа данных. Это позволяет моделям избегать дублирования информации для каждого языка или модальности. Понимание принципов работы моделей поможет разработать более эффективный ИИ, способный лучше обрабатывать разнообразные данные.

<https://arxiv.org/abs/2411.04986>

<https://hightech.fm/2025/02/19/llm-semantic-hub>

## Perplexity запустила новую функцию Deep Research

Компания Perplexity представила пользователям новый сервис «глубоких исследований» (Deep Research) для детального анализа информации и создания подробных отчетов. В отличие от аналогичных решений от OpenAI и Google, базовая версия инструмента будет доступна бесплатно.

Сервис Deep Research позволяет пользователям получать развернутые аналитические отчеты по любой теме. Система Perplexity самостоятельно ищет источники информации, анализирует их и формирует структурированный документ. На запросы уходит в среднем около трех минут.

Разработчики протестировали новую систему с помощью бенчмарка Humanity's Last Exam, который оценивает способность ИИ отвечать на вопросы экспертного уровня. Модель Perplexity набрала 21,1%. Это существенно выше показателей большинства конкурентов, включая Gemini от Google (6,2%) и Grok-2 (3,8%), однако уступает Deep Research от OpenAI (26,6%).

Команда также провела SimpleQA – эталонный тест, оценивающий способность языковых моделей отвечать на короткие вопросы, требующие фактов. Цель SimpleQA – стать простым и надежным набором данных для оценки фактологичности пограничных моделей. Deep Research от Perplexity показала в бенчмарке точность в

The screenshot shows the Perplexity AI interface. At the top, it says "Elon Musk announces Grok-3 release with live demo". On the left, there's a sidebar titled "Deep Research" with a list of 20 steps. In the center, there's a main search area with a button to "Ask follow-up". On the right, there's a panel titled "20 sources" showing news articles from various sources. One article from "businesstoday" discusses Elon Musk's announcement of Grok 3. Another from "outlookbusiness" talks about Musk calling it the "smartest AI on Earth". A third from "punching" discusses the unveiling of Grok 3. A fourth from "indianexpress" mentions Musk announcing that Grok 3 will debut with a live demo.

93,9 %.

Особенность решения Perplexity составляют бесплатные запросы. Если Deep Research от OpenAI доступен пользователям только в подписке Pro стоимостью \$200 в месяц, то Perplexity предлагает бесплатный доступ с ограниченным количеством запросов в день. Конкретное количество запросов неизвестно, и, вероятно, зависит от нагрузки нейросети. Платная подписка снимает эти ограничения.

Deep Research уже доступен через веб-интерфейс Perplexity. В ближайшее время компания планирует добавить этот сервис в приложения для Mac, iOS и Android. Для использования инструмента достаточно выбрать режим в раскрывающемся меню при отправке запроса. Готовые отчеты можно экспортить в PDF или делиться ими с помощью ссылки.

<https://www.perplexity.ai/hub/blog/introducing-perplexity-deep-research>

<https://hightech.fm/2025/02/17/perplexity-deep>

<https://habr.com/ru/articles/883272/>

## В OpenAI анонсировали выход моделей GPT-4.5 и GPT-5

Генеральный директор OpenAI Сэм Альтман анонсировал обновление дорожной карты по развитию ChatGPT. Ключевым моментом его планов стал анонс двух новых

моделей, GPT-4.5 и GPT-5, которые должны кардинально изменить подход к работе с искусственным интеллектом. OpenAI планирует не просто выпустить обновленные версии существующих моделей, а «полностью переосмыслить взаимодействие пользователей и ИИ».

По словам Альтмана, текущая ситуация, когда людям приходится выбирать между различными моделями и режимами работы, создает ненужные сложности. OpenAI намерена объединить все свои разработки в единую систему, которая будет самостоятельно подбирать оптимальные инструменты для решения задач, в зависимости от их направления и сложности.

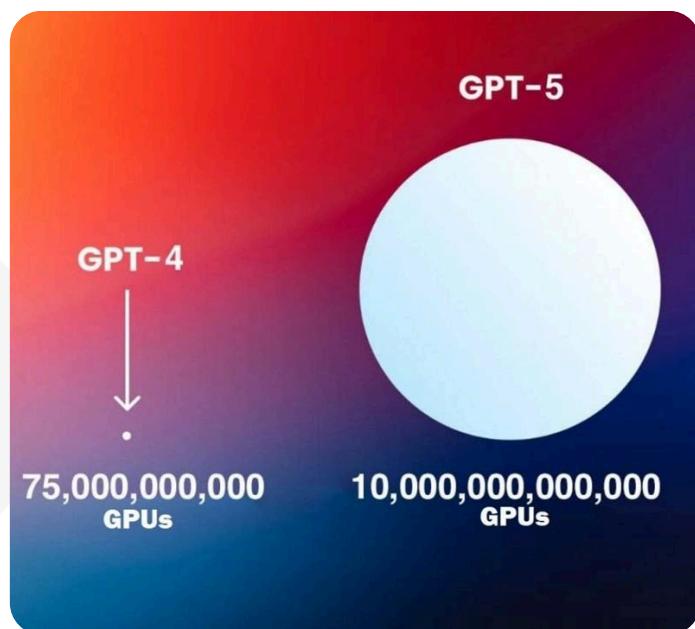
Первым шагом к этой цели станет выпуск GPT-4.5, известной внутри компании под кодовым названием Orion. Эта модель примечательна тем, что станет последней в линейке ChatGPT, не использующей механизм «цепочки рассуждений» (chain-of-thought). Такое решение обеспечит плавный переход к новой архитектуре, на которую планирует перейти OpenAI, сохранив совместимость с существующими приложениями.

Несмотря на это, в техническом плане GPT-4.5 представляет собой значительный шаг вперед по сравнению с GPT-4о. Разработчики существенно оптимизировали механизмы обработки контекста и улучшили работу с длинными последовательностями данных. При этом модель сохранит совместимость со всеми существующими промптами и инструментами.

Выпуск GPT-4.5 запланирован на ближайшие недели. Альтман подчеркнул, что несмотря на отсутствие механизма chain-of-thought, модель демонстрирует «впечатляющие» результаты в решении сложных задач. Этого удалось достичь за счет оптимизации внутренней архитектуры нейронной сети и улучшенных алгоритмов обработки информации.

Следующим этапом станет выпуск GPT-5 — модели, которая объединит в себе все лучшие наработки OpenAI, включая технологии проекта о3. По словам Альтмана, о3 не будет выпущен как отдельный продукт. Вместо этого его возможности интегрируют непосредственно в GPT-5, что позволит создать «по-настоящему универсальную систему».

Главная инновация GPT-5 будет заключаться в способности самостоятельно определять необходимый уровень анализа для каждой конкретной задачи. Если запрос требует глубокого осмыслиения, модель автоматически задействует механизмы рассуждений. Для простых задач будут использоваться более быстрые алгоритмы обработки. Пользователям больше не придется выбирать режим работы вручную, поскольку система все сделает самостоятельно. Но, в отличие от GPT-4.5, GPT-5 проанализирует задачу и предложит ее решение на более глубоком уровне.



OpenAI планирует разделить GPT-5 на три уровня:

- базовая версия будет бесплатной для всех желающих и обеспечит пользователям доступ к стандартным настройкам ИИ;
- с помощью тарифа Plus можно будет воспользоваться расширенными функциями;
- Pro-версия обеспечит максимум возможностей и позволит тонко настраивать параметры модели под индивидуальные потребности каждого пользователя.

Значительные изменения также коснулись механизмов обработки пользовательских запросов. В новой версии Model Spec подробно описаны алгоритмы, позволяющие моделям различать исследовательский интерес от потенциально опасных намерений. Это повысит точность работы чат-бота с чувствительными темами или техническими вопросами, где грань между образовательным контентом и вредоносной информацией может быть тонкой.

Разработчики OpenAI также решили проблему со склонностью моделей соглашаться с пользователем даже в случае очевидных ошибок. По словам Альтмана, новые алгоритмы запрограммированы на поиск объективной истины и предоставление обоснованного анализа, даже если это противоречит ожиданиям собеседника. Такой подход позволит более эффективно использовать нейросеть в образовательной сфере и при решении технических задач.

Подобные изменения могут стать поворотным моментом для индустрии в целом. OpenAI фактически предлагает новый стандарт работы с языковыми моделями, где главным приоритетом становится простота использования при сохранении максимальной функциональности. Это может заложить новый тренд для всей отрасли, подтолкнув другие компании переосмыслить свои подходы к разработке ИИ.

<https://x.com/sama/status/1889755723078443244>

<https://www.technologyreview.com/2025/02/27/1112619/openai-just-released-gpt-4-5-and-says-it-is-its-biggest-and-best-chat-model-yet/>

<https://habr.com/ru/companies/x-com/articles/884256/>

## Греческие ученые предложили подход, который сделает системы искусственного интеллекта более экономичными

Ученые из Фонда исследований и технологий Hellas (FORTH) в Греции предложили новый подход, способный сделать системы искусственного интеллекта более экономичными. Исследователи разработали нейронную сеть, которая имитирует работу биологических нейронов, включая дендриты — ключевые структуры мозга, участвующие в обработке информации.

Дендриты — разветвленные отростки нервных клеток, которые принимают и передают сигналы. Исследования последних лет показали, что отростки выполняют сложные вычисления независимо от основного нейрона. Это вдохновило исследователей на создание искусственной нейросети с аналогичной архитектурой.

Модель, описанная исследователями в журнале *Nature Communications*, демонстрирует ряд преимуществ перед традиционными сетями. Тестирование показало, что у дендритных нейросетей высокая точность распознавания изображений при значительно меньшем количестве параметров. Кроме того, они оказались более устойчивыми к переобучению — проблеме, когда нейросеть запоминает данные, но не может применять полученные знания к новым задачам.

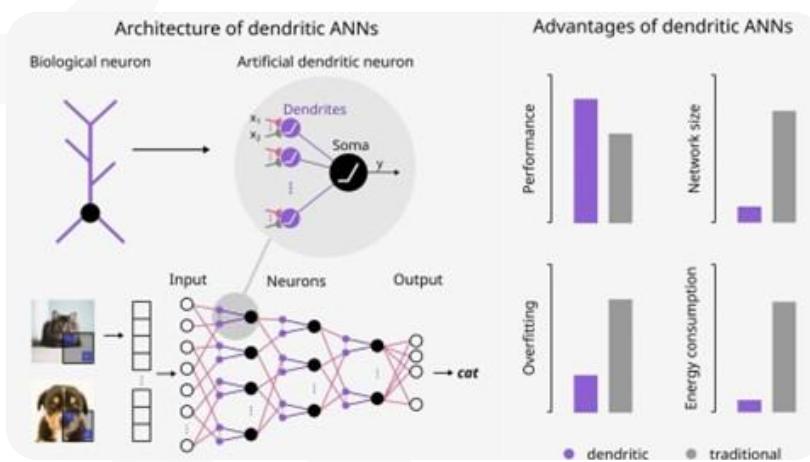
По словам Спиридона Чавлиса, соавтора исследования, новый подход меняет представление о том, как именно должна работать нейросеть. Вместо того чтобы задействовать миллионы параметров, как это принято в современных системах ИИ, ученые используют архитектуру, в которой несколько узлов кодируют различные категории данных.

Это делает сеть более гибкой, компактной и энергоэффективной.

Энергосбережение становится критически важным по мере роста масштабов использования ИИ, поскольку технологии машинного обучения от автономных автомобилей до облачных сервисов требуют огромных вычислительных мощностей. Решение, предложенное учеными, может изменить ситуацию, если позволит создавать более эффективные системы с меньшими затратами ресурсов.

Как отмечают исследователи, их работа может открыть путь «к новой эре искусственного интеллекта», где вычислительные мощности будут распределяться эффективнее, а модели станут доступнее для широкого спектра приложений.

Следующим шагом группы ученых станет тестирование дендритных ИНС на более сложных задачах, включая обработку естественного языка и медицинскую диагностику.



<https://www.nature.com/articles/s41467-025-56297-9>

<https://hightech.fm/2025/02/06/dendrites-ai>

## Робот Go2 Pro впервые принял участие в театральной постановке

В Нью-Йорке выпустили театральный проект, который был создан непосредственно с участием искусственного интеллекта: робот-собака Unitree Go2 дебютировал на

на сцене в рамках спектакля The Robot, the Spy, and the Love of AI. Пьеса, сочетающая элементы комедии и научной фантастики, шла в Chain Theater до 1 марта в рамках фестиваля Chain Winter One Act Festival.

Идея спектакля принадлежит драматургу Джону Артуру Лонгу, вдохновленному видео с участием робота Go2 Pro. Лонг приобрел робота и написал сценарий, встроив в сюжет сюжет его технические возможности.

Unitree Go2 Pro – это робот, разработанный китайской компанией Unitree, который является конкурентом известного Spot от Boston Dynamics. Его стоимость стартует от \$1600, а корпус выполнен из алюминиевого сплава и инженерного пластика. Робот весит около 15 кг и оснащен 3D-лидаром, широкоугольной камерой, Wi-Fi 6, Bluetooth 5.2 и аккумулятором на 8000 мАч.

Оператор Дарьель Гарсия управлял движениями робота через приложение и контроллер. По его словам, «90% действий выполнялись вручную, а сложные элементы вроде прыжков и движений лап были предустановлены». Помимо имитации движений собаки, Go2 Pro умеет делать сальто, танцевать и выполнять трюки.

Голос роботу в режиме реального времени дал актер Дэвид Рэй, создавая иллюзию «живого» персонажа. Режиссер спектакля Керк Гостковски отметил, что изначально рассматривался вариант записанных реплик, но от него отказались ради большей естественности.

Сюжет пьесы исследует классические темы научной фантастики, задавая вопросы о природе человечности в мире, где господствует ИИ. В постановке приняли участие Габриэль Рисдалль (Amazon Prime's The Better Sister), Моника Пак (Simpatico) и робот Go2 Pro, который стал полноценным актером.

Эксперимент с роботизированными исполнителями демонстрирует, как технологии могут менять живые выступления. Театральный дебют Unitree Go2 может стать началом новой эпохи, где роботы не только работают за кулисами, но и выходят на сцену.



<https://www.cnet.com/tech/computing/this-robot-dog-is-making-its-off-broadway-theater-debut/>

<https://www.securitylab.ru/news/556506.php>

## DeepMind представила новую модель AlphaGeometry2

Принадлежащая Google компания DeepMind заявила, что ее искусственный интеллект AlphaGeometry2 превзошел средний уровень золотых медалистов

Международной математической олимпиады (IMO) по решению задач по геометрии.

AlphaGeometry2 решил 84% задач по геометрии Международной математической олимпиады (IMO) с 2000 по 2024 год, что на 54% больше, чем у предыдущей версии. На тесте IMO-AG-50, который включает 50 формализованных задач по геометрии IMO, AlphaGeometry решил 42 задачи. Средний результат золотых медалистов — 40,9 задачи.

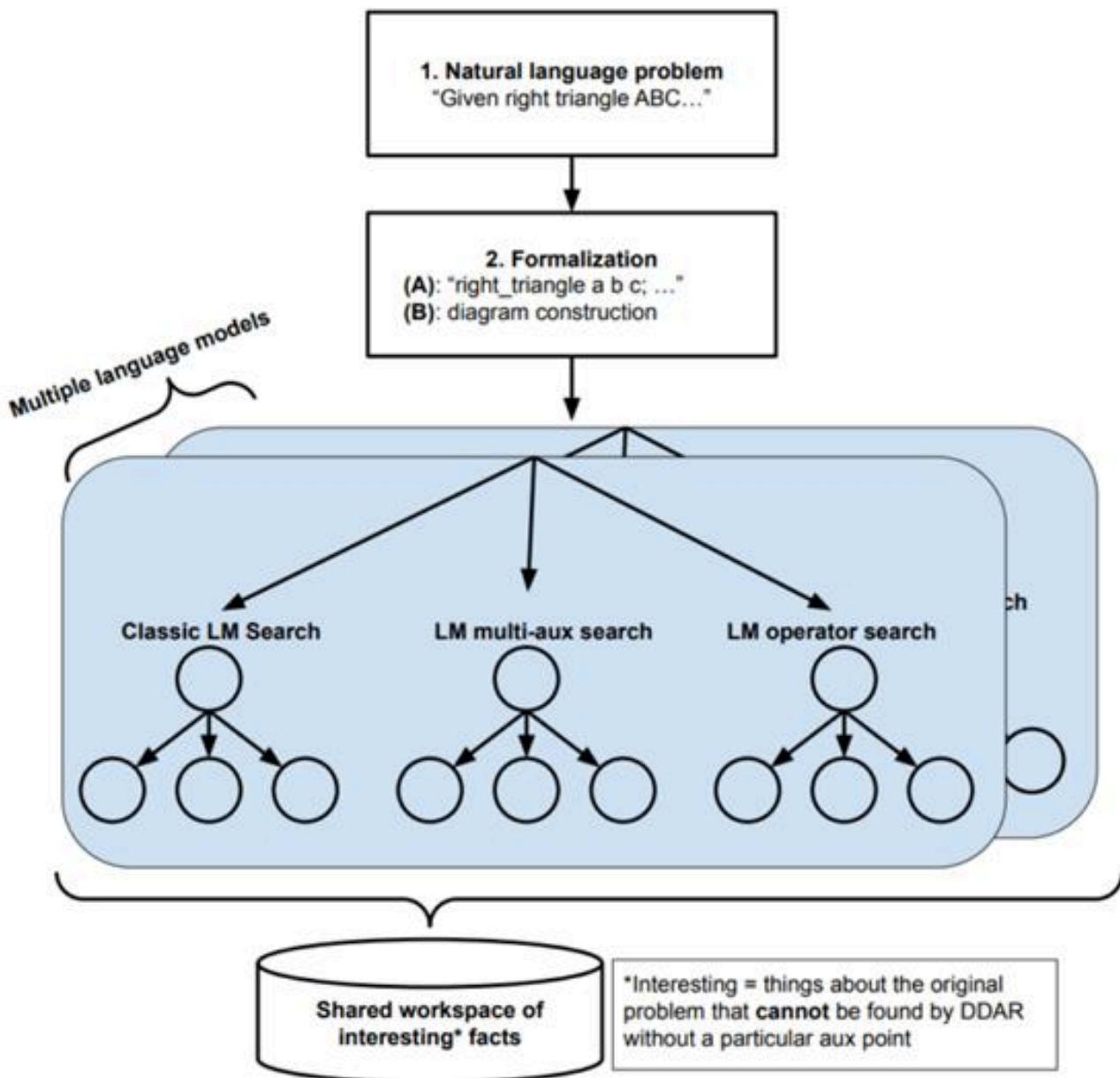


Figure 4 | Overview of our search algorithm. We employ several different search trees which can share facts they proved via a special knowledge sharing mechanism.

AlphaGeometry2 был разработан на основе первой версии AlphaGeometry, выпущенной в начале 2023 года. В новой версии объединены несколько ключевых элементов, включая языковую модель Google Gemini и математическую систему Alpha Proof, предназначенную для выполнения доказательств.

Система работает в связке с двумя основными компонентами: языковой моделью, основанной на архитектуре Gemini, и символическим механизмом под названием DDAR (дедуктивная база данных для арифметических рассуждений).

Языковая модель, обученная на задачах синтетической геометрии, предлагает потенциальные шаги и конструкции, которые могут помочь решить задачу. Она делает это, генерируя предложения на специализированном языке, описывающем геометрические объекты и взаимосвязи.

Затем DDAR рассматривает эти предположения, используя логику для получения из них новых фактов. Следуя определённым правилам, система DDAR формирует то, что команда называет «дедуктивным замыканием» всех возможных выводов.

Процесс решения проблемы осуществляется посредством итераций. Языковая модель генерирует возможные следующие шаги, которые DDAR проверяет на логическую последовательность и полезность. «Многообещающие» идеи сохраняются и исследуются далее.

Новый алгоритм поиска SKEST запускает несколько стратегий поиска параллельно, позволяя обмениваться полезными результатами через общую базу знаний. Это помогает быстрее находить решения.

Когда DDAR находит полное доказательство, объединяющее предложения языковой модели с известными принципами, AlphaGeometry2 представляет его в качестве решения. Команда отмечает, что эти доказательства часто демонстрируют неожиданную креативность.

По сравнению с предыдущей версией AlphaGeometry, в новой версии было внесено множество улучшений и оптимизаций. К ним относятся более выразительный язык геометрического описания, который теперь включает в себя локусные кривые и линейные уравнения, а также более быстрая реализация DDAR на C++. Утверждается, что новая версия в 300 раз быстрее предыдущей реализации на Python.

В производительности AlphaGeometry2 не играют решающей роли как используемый токенизатор, так и язык обучения, специфичный для конкретной области. Аналогичные результаты были получены как с использованием специализированных токенизаторов с небольшим словарным запасом, так и с использованием универсальных токенизаторов с большими моделями. Обучение на естественном языке также дало сопоставимые результаты с обучением на формальном языке геометрии.

Другим интересным открытием стал тот факт, что языковые модели, предварительно обученные на математических наборах данных, а затем доработанные на данных AlphaGeometry, приобретают немного иные способности по сравнению с теми, которые обучаются с нуля. Хотя обе модели обучаются на одних и тех же данных, они развиваются взаимодополняющие сильные стороны. Объединив эти модели в новый алгоритм поиска SKEST, можно значительно повысить скорость решения задач.

Исследование также даёт важное представление о роли LLM в решении математических задач. Согласно указанной информации, модели AlphaGeometry2 способны генерировать не только вспомогательные конструкции, но и полные доказательства. Это говорит о том, что современные языковые модели могут работать без внешних инструментов, таких как символьные процессоры.

Несмотря на прогресс, достигнутый в AlphaGeometry2, у новой модели DeepMind всё ещё есть ограничения. Например, используемый формальный язык пока не позволяет описывать задачи с переменным количеством точек, нелинейными уравнениями или неравенствами. Кроме того, некоторые задачи Международного математического конкурса остаются нерешёнными. Возможными отправными точками для дальнейших улучшений являются разбиение сложных задач на подзадачи и применение обучения с подкреплением.

Помимо задач по геометрии, этот подход можно применить и в других областях математики и естественных наук. Потенциальные области применения варьируются от решения сложных задач по физике и инженерии до помощи исследователям и студентам.

<https://the-decoder.com/deepminds-alphageometry-solves-complex-math-and-is-a-crucial-step-toward-agi/>

<https://habr.com/ru/companies/boothub/news/880512/>

## Ученые планируют обучить искусственный интеллект, как ребенка

Как отмечают специалисты, современные системы искусственного интеллекта не понимают смысл слов и часто выдумывают несуществующие факты. Учёные из лаборатории искусственного интеллекта Брюссельского свободного университета предложили новое решение: создать модель, которая будет учиться языку «как ребёнок», через непосредственное взаимодействие с реальным миром.

В отличие от человека, особенно ребёнка, компьютер не способен интуитивно взаимодействовать с окружающей средой и корректировать свои предположения на основе реальных ощущений. Он анализирует шаблоны в текстах, не вникая в суть того, о чём «говорит».

Ситуацию усугубляют «галлюцинации» и предвзятость нейросетей. Один из авторов исследования столкнулся с подобной проблемой, когда попросил ИИ составить развлекательную, но точную викторину для подростков. Система решила, что «развлекательный» означает «бессмысленный», и создала тест с недостоверной информацией о турецком фольклоре. Подобный нюанс является не просто ошибкой – использовав непроверенную информацию, можно нанести оскорбление другому народу, его традициям и культуре.

Исследователи провели подробный анализ того, как именно учатся языку дети и компьютеры. Дети осваивают речь естественным путём: играют со словами, экспериментируют, взаимодействуют с окружающими, подражают и интерпретируют услышанное. LLM обрабатывают огромные объёмы текстов и компилируют из них новый контент. Безусловно, они «крайне эффективны во многих видах работы с текстом», поскольку могут обобщать информацию, переводить, отвечать на вопросы. Однако настоящего понимания того, что они обрабатывают, у машин нет.

Планируется, что вместо пассивного поглощения данных новой модели предстоит

активно взаимодействовать с окружающей средой. По мнению исследователей, это поможет решить главную проблему современных систем вроде ChatGPT и Gemini. Новые модели будут «глубже укореняться в смысле и намерении» – то есть научатся понимать контекст и значение слов.

Новый подход поможет также решить множество практических проблем в системах ИИ. Например, Gemini плохо справляется с голосовыми командами – нейросеть путает разные уровни контекста и может внезапно ответить на вопрос, который пользователь задавал неделю назад. Модель, построенная на понимании намерений, а не на шаблонах, будет работать гораздо точнее. Кроме того, такой подход, по мнению исследователей, может оказаться более экологичным. Сейчас для обучения ИИ требуются большие вычислительные мощности, что приводит к значительным энергозатратам и, как следствие, к увеличению углеродного следа.

Однако, как именно организовать подобное обучение на практике пока остается под вопросом. Современные языковые модели работают с миллиардами слов, что делает их объем знаний практически максимальным на текущий момент. С новым подходом обучение моделей придется начинать «с нуля», что приводит к ряду важных вопросов: с какого словарного запаса начинать; будет ли достаточным словарный запас в несколько сотен базовых слов; смогут ли компании масштабировать подход, который потребует гораздо более сложной инфраструктуры для обучения ИИ.

Отдельной проблемой являются также культурные и эмоциональные нюансы языка. Так фразу «я люблю тебя» на западе могут использовать в повседневной жизни достаточно часто и в любом контексте, тогда как в Японии или Корее её произносят в исключительных случаях и, как правило, «с глубоким чувством». Или слово «извините», которое в азиатских культурах является способом избежать конфликта, а на Западе – признанием вины. Когда ChatGPT извиняется, он, безусловно не испытывает настоящего сожаления и не понимает подобных тонкостей.

Для полноценного освоения языка модели потребуется физическое взаимодействие с миром. Так, для понимания значение слов «вверх», «вниз» или «рядом», нужно уметь перемещаться в пространстве. А для правильной интерпретации фразы «я в порядке», необходимо разбираться в сложной эмоциональной динамике: это может быть проявлением безразличия, вежливости, сарказма, просьбой о поддержке или действительно искренним ответом.

<https://cybernews.com/news/language-learning-children-ai/>

<https://www.securitylab.ru/news/556114.php>

## Alibaba представила новую языковую модель Qwen 2.5 Max

Китайская компания Alibaba представила новую языковую модель Qwen 2.5 Max, которая, по заявлению разработчиков, превосходит ведущие американские ИИ-системы. На фоне недавнего успеха китайского стартапа DeepSeek это событие вызвало новую волну дискуссий о лидерстве США в области искусственного интеллекта.

По результатам тестирования в популярных системах оценки Arena-Hard, MMLU-Pro, GPQA-Diamond, LiveCodeBench и LiveBench, Qwen 2.5 Max показала более высокие результаты, чем GPT-4o от OpenAI, Claude 3.5 Sonnet от Anthropic и Llama 3.1 405B от Meta (здесь и далее компания Meta признана экстремистской и запрещена на территории РФ). Модель умеет работать с текстом, видео и изображениями, а также может осуществлять поиск в интернете.

Разработчики Alibaba отдельно подчеркнули, что сравнивали свою модель с версией DeepSeek V3, а не с R1. Это объясняет выбор GPT-4o в качестве точки сравнения вместо флагманских моделей o1 от OpenAI. Тем не менее, достижения китайских компаний ставят под сомнение эффективность западных попыток сдержать развитие искусственного интеллекта в Китае.

Технические характеристики новой модели пока остаются в секрете. Известно, что Qwen 2.5 Max построена на архитектуре MoE (mixture of experts) и обучена на массиве данных объемом 20 триллионов токенов. После базового обучения модель прошла этапы SFT (supervised fine-tuning)<sup>1</sup> и RLHF (reinforcement learning from human feedback)<sup>2</sup>.

Архитектура MoE, которую также используют Mistral и DeepSeek, позволяет создавать специализированные «экспертные» модули для решения конкретных задач, например, программирования или математических вычислений. Главное преимущество такого подхода – возможность увеличивать общее количество параметров без снижения производительности, так как для каждого запроса активируются только релевантные части нейросети.

Точное количество параметров Qwen 2.5 Max не разглашается, хотя известно, что предыдущая версия Max содержала около 100 миллиардов параметров. Примечательно, что встроенный чат-бот модели не может ответить на вопрос о своих технических характеристиках.

В отличие от предыдущих версий Qwen, новая модель не будет распространяться в открытом доступе. На сайте Alibaba Cloud она отмечена как проприетарная, что объясняет скучность технической информации. Такой подход типичен для многих разработчиков и уже применялся компанией с моделями Qwen Turbo и Qwen Plus.

Стоимость использования Qwen 2.5 Max через API составляет 10 долларов за миллион входных токенов и 30 долларов за миллион сгенерированных. Для сравнения, GPT-4o от OpenAI берет 2,5 доллара за миллион входных и 10 долларов за миллион выходных токенов. При этом флагманская модель o1 от OpenAI стоит дороже – 15 и 60 долларов соответственно.

Qwen 2.5 Max – это последнее пополнение в семействе моделей Alibaba, которые начали активно развиваться с 2023 года. С сентября компания выпустила серию моделей Qwen 2.5 с открытым исходным кодом, имеющих от 0,5 до 72 миллиардов параметров. По заявлениям разработчиков, крупнейшие из них не уступают 405-миллиардной Llama от Meta.

---

[1] Supervised fine-tuning (SFT) — это методика, используемая для адаптации предварительно обученных LLM под конкретные узкие задачи при помощи размеченных данных.

[2] RLHF (Обучение с подкреплением на основе отзывов людей) — это метод машинного обучения, который использует полученную от человека обратную связь для оптимизации моделей машинного обучения, позволяя организовать более эффективное самообучение.

Параллельно Alibaba развивает специализированные модели для математических вычислений и программирования, а в декабре 2024 года представила «думающую» модель QwQ, похожую по функционалу на o1. В 2025 году компания также выпустила три открытые модели для работы с изображениями (VLM), которые, как утверждается, не уступают Gemini 2 от Google, GPT-4o и Claude 3.5 Sonnet.

Кроме того, Alibaba обновила свои модели Qwen 2.5 на 7 и 14 миллиардов параметров, увеличив их контекстное окно до миллиона токенов. Такое расширение может быть особенно полезным для RAG-систем, позволяя обрабатывать большие объемы документации без потери контекста.

Однако успехи китайских разработчиков вызывают опасения их зарубежных коллег относительно конфиденциальности данных и цензуры. Как и в случае с DeepSeek, пользовательские данные Qwen Chat хранятся на серверах в Китае или Сингапуре.

Если говорить о цензуре, то, например, как и другие китайские модели, DeepSeek и Alibaba избегают обсуждения таких тем, как события на площади Тяньаньмэнь или политический статус Тайваня, либо прерывают генерацию ответов на подобные вопросы.

<https://qwenlm.github.io/blog/qwen2.5-max/>

<https://www.securitylab.ru/news/555966.php>

## Google представила бесплатную версию своего инструмента для разработчиков Gemini Code Assist

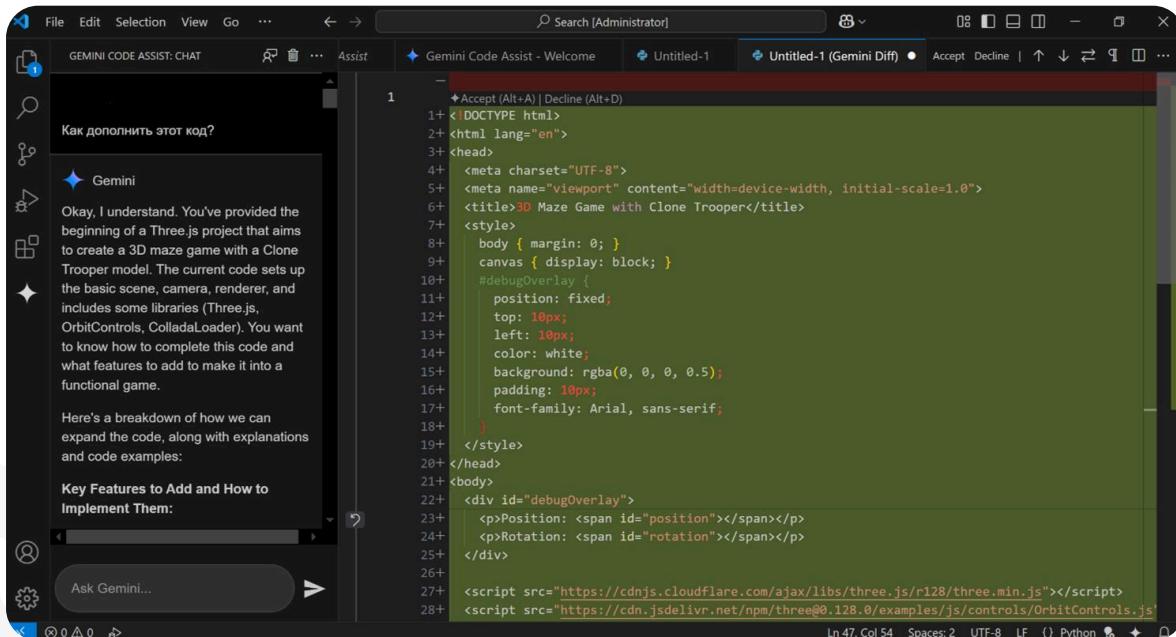
Google запустила новую бесплатную версию своего AI-инструмента для автодополнения и помощи в написании кода — Gemini Code Assist for Individuals. Одновременно компания представила Gemini Code Assist for GitHub — систему автоматизированного анализа кода, которая выявляет ошибки и предлагает исправления непосредственно в GitHub.

Gemini Code Assist for Individuals позволяет разработчикам взаимодействовать с AI через чат, используя естественный язык. Инструмент может исправлять ошибки, дополнять код и объяснять сложные участки. Он основан на специальной версии модели Gemini 2.0, оптимизированной для программирования, поддерживает интеграцию с VS Code и JetBrains через плагины, а также работу с разными языками программирования.

интеграцию с VS Code и JetBrains через плагины, а также работу с разными языками программирования.

Одним из ключевых преимуществ инструмента стал достаточно большой лимит на использование:

- 180 000 автодополнений кода в месяц, что в 90 раз больше, чем у бесплатного плана GitHub Copilot (2 000 автодополнений);
- 240 запросов в чате в день, что почти в пять раз превышает лимит GitHub Copilot.



Контекстное окно модели составляет 128 000 токенов, что более чем в четыре раза превышает возможности конкурентов. Благодаря этому AI может анализировать большие объемы кода в одном запросе.

Ассистент поддерживает более 20 языков программирования, включая C, C++, Go, Python, Java, JavaScript, Kotlin и TypeScript.

Для желающих протестировать новый инструмент Google открыла бесплатный публичный доступ с 20 февраля.

Для работы необходимо скачать Gemini Code Assist для Visual Studio Code, GitHub или JetBrains и подключить к данным средам разработки. Затем нужно будет авторизоваться через аккаунт Google. Также сервис потребует создать проект в Google Cloud. В списке доступных стран России на текущий момент нет, однако можно отметить любую другую страну и начать работу.

Gemini Code Assist for GitHub, в свою очередь, автоматически анализирует pull request, выявляя ошибки и предлагая улучшения. Этот шаг демонстрирует активное стремление Google конкурировать с Microsoft и GitHub в сфере инструментов для разработчиков.

Семь месяцев назад Google привлекла Райана Сальву, ранее возглавлявшего команду GitHub Copilot, в качестве руководителя направления AI-инструментов для разработчиков. По словам Сальвы, стратегия компании заключается в том, чтобы привлечь разработчиков на ранних стадиях карьеры, предоставляя мощный инструмент бесплатно, а затем переводить их на корпоративные тарифы.

<https://techcrunch.com/2024/12/17/code-assist-googles-enterprise-focused-code-assistant-gets-third-party-tools/>

<https://vc.ru/ai/1832788-google-vypustila-besplatnuyu-beta-versiyu-ii-pomoshnika-dlya-programmistov-gemini-code-assist>

<https://www.securitylab.ru/news/556751.php>

## Citibank GPS представил новый отчет — «Агентный ИИ и экономика Do It For Me»

Документ «Citibank GPS Focus Report: Agentic AI & the 'Do It For Me' Economy» представляет собой исследование влияния агентного искусственного интеллекта (Agentic AI) на финансовый сектор, глобальную экономику и рынок труда.

Ключевые идеи отчета следующие:

1. Агентный ИИ — это новая эволюция искусственного интеллекта, при которой системы могут: автономно принимать решения без вмешательства человека; проактивно решать задачи вместо реагирования на пользовательские запросы; самостоятельно взаимодействовать с другими цифровыми системами и выполнять сложные операции; имитировать человеческое мышление и поведенческие паттерны. Развитие контекстного понимания, памяти и мультизадачности позволило ИИ стать по-настоящему автономным. В 2024 году количество упоминаний агентного ИИ в корпоративных документах и СМИ выросло в 17 раз, в 2025 году ожидается еще более резкий рост.
2. Финансовая индустрия является одним из крупнейших потребителей AI-решений, уступая только сектору телекоммуникаций и медиа. Основные изменения в данной сфере благодаря появлению агентного ИИ следующие:
  - появление персонализированных финансовых продуктов — агентный ИИ помогает клиентам выбирать оптимальные банковские услуги, рассчитывать инвестиционные стратегии и адаптировать кредиты под их финансовое положение;
  - кибербезопасность и антифрод — AI-агенты помогают выявлять мошеннические транзакции, анализируя контекст данных;
  - оптимизация корпоративного банкинга с помощью автоматизированных систем управления ликвидностью, расчетов по кредитам и управления рисками;
  - страхование — персонализированные страховки, прогнозирование рисков и автоматическая обработка страховых случаев;
  - автоматизация работы с клиентами — виртуальные финансовые ассистенты, налоговые консультанты и ИИ-агенты для планирования пенсии становятся стандартом отрасли;
  - институциональные инвестиции — агентный ИИ анализирует большие объемы данных в режиме реального времени, разрабатывая инвестиционные стратегии.
3. Экономика Do It For Me. Парадигма «Do It For Me» (Сделай за меня) заключается в том, что пользователи все больше полагаются на ИИ-ассистентов для выполнения сложных задач. Так, ИИ-бот может работать как личный финансовый агент, самостоятельно выбирая лучшие банковские предложения, оптимизируя налоги и инвестиции. Кроме того, агент может самостоятельно заказывать продукты, билеты, подписки и услуги по заданным критериям, а также следить за изменениями цен на активы, недвижимость, страховки и давать рекомендации.
4. Влияние на рынок труда. Agentic AI неизбежно изменит в будущем структуру

занятости. Многие рутинные и административные профессии будут автоматизированы, а работа людей сместится в сторону управления и взаимодействия с ИИ-агентами. Будут популярными такие профессии, как: инженеры и разработчики AI, аналитики данных; специалисты по этике ИИ; специалисты по кибербезопасности; UX-дизайнеры для AI-ассистентов.

5. Рост инвестиций в AI-стартапы. В 2024 году 37% венчурных инвестиций было направлено в AI-стартапы, и этот показатель продолжит расти в 2025 году. Больше всего инвестиций привлекут следующие направления: автономные AI-ассистенты и цифровые сотрудники (+150% роста сделок); GenAI для клиентских сервисов (+113%); AI-решения для борьбы с мошенничеством (+82%); роботы и системы автоматизации (+60%).

Одним из главных вызовов для специалистов отрасли остается вопрос безопасности и этичности работы автономных ИИ-агентов. Ключевыми проблемами в данном секторе являются:

- кибербезопасность — автоматизированные системы могут стать мишенью для атак, поскольку работают без участия человека;
- мошенничество — агентный ИИ может использоваться для масштабирования мошеннических схем;
- уязвимость данных — AI-агенты обрабатывают огромное количество личной информации, что требует строгих мер защиты;
- правовое регулирование — в Европейском секторе в настоящее время уже обсуждаются законы, запрещающие полностью автоматизированные юридически значимые решения.

В 2025 году ожидается развитие таких трендов, как регуляторные стандарты, AI-этика и механизмы контроля за автономными агентами.

[https://www.citiwarrants.com/home/upload/citi\\_research/rsch\\_pdf\\_30305836.pdf](https://www.citiwarrants.com/home/upload/citi_research/rsch_pdf_30305836.pdf)

## Alibaba выпустила новую модель для генерации видео Wan 2.1

Китайский технологический гигант Alibaba выпустил новую открытую ИИ-модель для генерации видео Wan 2.1. Модель может создавать видео по текстовому описанию или из изображений, редактировать ролики и генерировать для них аудио.

Компания разработала три версии Wan 2.1 — I2V-14B, T2V-14B и T2V-1.3B. Первые две модели поддерживают генерацию роликов в разрешении от 480p до 720p, последняя модель поддерживает только разрешение 480p. Индекс 14B в названии ИИ-модели указывает на то, что она способна учитывать до 14 млрд параметров при выполнении запроса.

Кроме того, T2V-1.3B требует всего 8,19 Гб видеопамяти, что делает ее совместимой почти со всеми доступными широкому потребителю графическими процессорами.

По словам представителя компании, Wan 2.1 превосходит существующие модели с открытым исходным кодом и современные коммерческие решения по различным показателям.

Появление Wan 2.1 вызвало ажиотаж в Telegram-каналах и на специализированных платформах, где пользователи делились своими первыми экспериментами с моделью. Одним из главных достоинств новой разработки стало её умение обрабатывать сложные движения, что делает генерируемое видео более реалистичным.

В комментариях многие сравнивают Wan 2.1 с другой популярной ИИ-моделью Hunyuau, созданной Tencent, отмечая значительный прогресс в качестве анимации.

Wan 2.1 можно скачать бесплатно на платформах GitHub и HuggingFace.

<https://vc.ru/ai/1834565-alibaba-vypustila-model-dlya-generacii-video-wan-21-s-otkrytym-ishodnym-kodom>

<https://x.com/fofrAI/status/1894523377752183069>

<https://wanxai.com/>