

08 JUILLET 2025

ГДЕ ЖДАТЬ НОВЫХ ПРОРЫВОВ В AI?

Лекция от Head-AI Researcher в Т-Банке

Автор лекции и создатель конспекта
ГАВРИЛОВ Д.Е
БАУЭР Н.В.

MATHEMATICS

SOMMAIRE

КОНСПЕКТ	3
История развития нейросетей (от меня)	3
Почему нас привлекли нейросети в своё время?	3
Проблемы chatGPT	4
Решение проблем chatGPT	4

История развития нейросетей (от меня)

История развития нейросетей за последние годы показывает большой скачок их возможностей:

⚡ 2016 — AlphaGo

Компания DeepMind создала AlphaGo — первую нейросеть, победившую чемпиона мира по игре Го. Это стало символом новой эры ИИ: нейросети впервые начали превосходить человека в сложных стратегических задачах, недоступных классическим алгоритмам.

⚡ 2017–2020 — волна больших языковых моделей

Появляются архитектуры трансформеров (GPT, BERT, T5). Модели начинают понимать, генерировать текст, переводить и даже писать код. Масштабы моделей росли от миллионов до сотен миллиардов параметров.

⚡ 2022–2023 — генерация изображений и общий ИИ

Модели типа DALL-E, Stable Diffusion, Midjourney учатся создавать фотореалистичные и художественные изображения по тексту. Вышел ChatGPT (на GPT-3.5 и GPT-4), который показал высокий уровень диалога и решения практических задач.

⚡ 2024 — состязания на уровне здравого смысла и логики

OpenAI и другие лидеры переходят к качественным бенчмаркам: больше не просто обрабатывать текст или фото, а решать «человеческие задачи», требующие рассуждений.

Ныне языковые модели могут намного лучше работать с информацией, поиском, суммаризацией и иными действиями с текстами, картинками и другими типами представления информации, что двигаться уже просто некуда

Кажется, что AGI уже близок и в AI больше делать нечего, но это неправильный ответ

Почему нас привлекли нейросети в своё время?



Появление ChatGPT стало *настоящим прорывом* в области машинного обучения и удивило многих специалистов. Раньше для решения любой задачи МО требовалось *вручную собирать подходящий датасет* и обучать модель именно на этих данных. Такой подход ограничивал спектр решаемых задач: например, было *практически невозможно собрать хороший датасет* для генерации осмысленных текстов или грамотно выполненной суммаризации. Пример ChatGPT показал, что крупные языковые модели могут эффективно справляться с задачами генерации, суммаризации текстов, перевода и многих других *без необходимости иметь отдельный датасет для каждой задачи*. Это открыло новые горизонты и значительно упростило применение ИИ в самых разных сферах.

«

Проблемы chatGPT



У ChatGPT есть и свои ограничения. Модель учится на огромном количестве текстов из интернета, то есть на информации, уже созданной и опубликованной людьми. Поэтому нейросеть не способна выдавать принципиально новые научные открытия, формулировать ранее неизвестные теоремы или законы природы. Она может лишь обобщать, пересказывать или комбинировать уже существующие знания



Fig. 1 - chatGPT решает только те задачи, которые человек умеет решать

Решение проблем chatGPT

С развитием искусственного интеллекта люди нашли способ применять ИИ для решения задач, которые человеку часто не под силу. Один из ключевых методов — обучение с подкреплением (RL), при котором агент (ИИ) учится выбирать последовательности действий, чтобы добиться наилучшего результата, получая за это «награду» или «штраф».

Яркий пример — компьютерные программы, такие как AlphaGo и современные шахматные движки. AlphaGo с помощью RL и нейронных сетей смог обыграть лучших игроков мира в го — игре, где количество возможных ходов огромно, и традиционные методы перебора не работают. В шахматах ИИ уже давно опережает людей по силе, а новые методы обучения только усилили его возможности. Такие системы помогают решать сложнейшие задачи, где человек не может оптимально просчитать все варианты, и по сути выходят за пределы человеческих стратегий.

Компетенции с RL

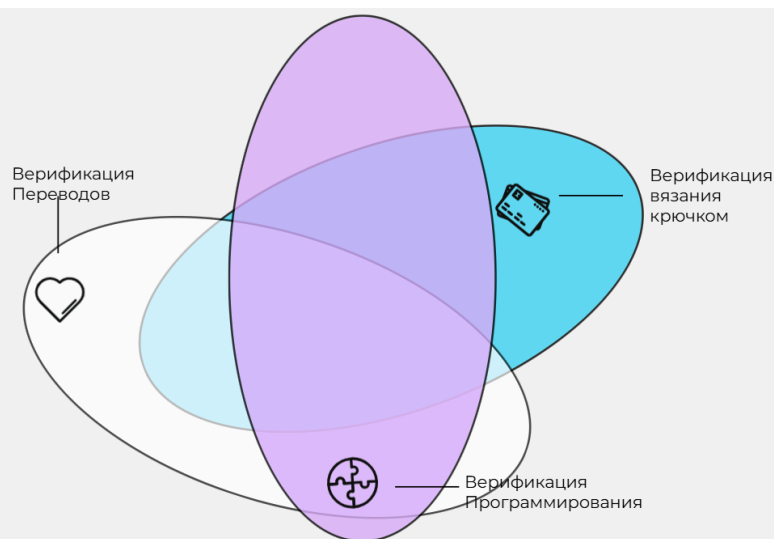


Fig. 2 - Задача для решения + оценка качества + ожидание => обучение



При RL Верификация не ограничена нашими знаниями или текущими решениями. Необходимы лишь ресурсы. Раньше мы обучались для каждой отдельной задачи, а теперь создаём решения, которые можно подстраивать под любые нужды

Сферы, в которые можно вкатиться: Инженерия, Реализация великого множества мат. задач, Аккуратные методы обучения

ГДЕ ЖДАТЬ НОВЫХ ПРОРЫВОВ В AI?

Лекция от Head-AI Researcher в Т-Банке

08 JUILLET 2025

ГАВРИЛОВ Д.Е
БАУЭР Н.В.