

# Лекция «Где ждать новых прорывов в области искусственного интеллекта?»

## Конспект

**Автор конспекта:** Феофанов Илья Сергеевич. Н07-3

**Спикер:** Гаврилов Даниил Евгеньевич, Руководитель лаборатории научных исследований искусственного интеллекта T-Bank AI Research

---

### Актуальность

Почему возникает вопрос: "Где ждать прорывов в AI"? OpenAI показали, что могут решать очень сложные задачи. И появилось ощущение, что в AI больше делать нечего.

OpenAI и Anthropic, Которые пользовались методами, казущимися странными, сейчас впереди.

Нужно понимать, куда "дует ветер", чтобы понимать, куда двигаться дальше)

Ощущение, что всё в AI всё достигнуто возникает не впервые.

$\text{data} \rightarrow p(\text{data}|\theta)$  – так работает ИИ

$\text{source text, translation} \rightarrow p(\text{translation}|\text{source}, \theta)$  – так работает машинный перевод

$\text{long text, short text} \rightarrow p(\text{short text}|\text{long text}, \theta)$  – похожий метод для создания суммаризатора

Но для обучения таких моделей требовался большой датасет, а следовательно, большое количество разметчиков

Потом вышел ChatGPT и решил задачу проще. Помогла смена парадигм. Модели стали обучать на данных из всего интернета.

### Основная часть

**Обучение с подкреплением:**

**В чём же проблема ChatGPT?** ChatGPT может работать с языками программирования, может подсказать, как взять крючком и т.д. Но кто будет решать проблемы человечества?

**Следующий шаг AI** – научить делать его то, что мы до сих пор не умеем. Как можно добиться этого? – вопрос остаётся открытым

Пример: AlphaGo – модель, которая смогла переиграть человека в игре Go. В создании таких моделей помогает обучение с подкреплением.

Задачи для решения + возможности оценить качество + подождать ⇒ модель, способная решать эти задачи.

### **Хронология:**

- Обучали каждой задаче по отдельности
- Обучали большую модель которую можно дообучить на нужную задачу
- Обучали LLM, которая может хорошо решать многие известные задачи
- Дообучали LLM решать каждую задачу по отдельности

– СЕЙЧАС МЫ ЗДЕСЬ! Что дальше?

**В чём сейчас проблема? Чем можно заниматься, чтобы оказаться во фронтире того, что сейчас происходит?**

- Инженерия
- Реализация великого множества верифицируемых задач
- Аккуратные методы обучения на этом

### **Нет принципиально нерешаемых задач!**

Ранше LLM называли просто брутфорсом. Но сейчас это работает. Таких примеров много. Нужно уметь принимать непонятные, неожиданные решения. **Это и есть путь к прорыву.**