

Конспект лекции: Где ждать новых прорывов в области искусственного интеллекта?

Лектор: Гаврилов Д.
Подготовил: Подлегаев А.

1 Введение

Искусственный интеллект (ИИ) продолжает развиваться, и новые прорывы в этой области вызывают ощущение, что все уже достигнуто. Однако это не так. История показывает, что подобные ощущения возникают регулярно, и за ними следуют новые смены парадигм. Например, после выхода ChatGPT многие считали, что ИИ достиг предела, но затем появились новые модели, такие как OpenAI, которые показали, что развитие продолжается.

2 Смена парадигм в ИИ

Ранние подходы к ИИ предполагали решение отдельных задач, таких как машинный перевод или классификация изображений, с использованием специализированных датасетов. Однако современные модели, такие как ChatGPT, обучаются на огромных объемах данных, что позволяет им решать широкий спектр задач без дополнительной настройки.

Ключевые моменты:

- **Ранние модели:** Требовали отдельных датасетов для каждой задачи.
- **Современные модели:** Обучаются на всех доступных данных, что позволяет им обобщать решения.
- **Пример:** ChatGPT демонстрирует способность выполнять задачи, которые ранее требовали сложных эвристик, например, суммаризацию текста.

3 Проблемы и вызовы

Несмотря на успехи, остаются задачи, которые ИИ пока не может решать. Например, доказательство гипотезы Римана или создание новых алгоритмов. Эти задачи требуют не просто обобщения данных, а способности к творчеству и открытиям.

Основные проблемы:

- **Ограничение данных:** Модели работают только с тем, что уже известно.
- **Необходимость верификации:** Для сложных задач важно не только найти решение, но и проверить его правильность.
- **Вычислительные ресурсы:** Обучение на множестве задач требует значительных мощностей.

4 Перспективные направления

Чтобы преодолеть текущие ограничения, можно использовать следующие подходы:

4.1 Обучение с подкреплением

Этот метод позволяет моделям улучшать свои решения через взаимодействие со средой. Примеры успешного применения:

- AlphaGo: Модель научилась играть в го лучше человека, используя самообучение.
- AlphaTensor: Создание новых алгоритмов для матричных операций.

4.2 Верифицируемые задачи

Создание задач, результаты которых можно проверить автоматически, например:

- Математические задачи.
- Программирование: Проверка корректности кода.
- Логические игры: Шахматы, крестики-нолики с варьируемой сложностью.

4.3 Инженерные решения

Разработка распределенных систем для обучения множества задач одновременно. Это требует:

- Эффективного использования вычислительных ресурсов.
- Оптимизации процессов обучения.

5 Заключение

Развитие ИИ продолжается, и следующие прорывы могут быть связаны с обучением на множестве верифицируемых задач и использованием методов обучения с подкреплением. Ключевые направления для исследований:

- Создание алгоритмов для одновременного обучения на разнородных задачах.
- Разработка методов верификации решений.
- Улучшение инженерных решений для масштабирования моделей.

История ИИ показывает, что простые и масштабируемые подходы часто оказываются наиболее эффективными. Поэтому важно сочетать исследования с практической реализацией, чтобы оставаться на переднем крае технологий.