

## Конспект лекции «Развитие больших языковых моделей, от малых компаний до крупных ИТ-гигантов. Достигнули максимум»

А какие вообще бывают задачи?

- Many-to-many: одинаковая длина начального и конечного текста.

Зачем нужна:

- Исправление ошибок в тексте
- Синонимы



- Many-to-many: разная длина начального и конечного текста

Зачем нужна

- Машинный перевод
- суммаризация
- Перенос стиля



- Many-to-one: большой текст в короткий

Зачем нужна:

- Классификация
- Регрессия

LLM повсюду, среди таких задач как:

- Исправление грамматических ошибок
- Лексическая семантика (значение слов)
- Знания о мире (география)
- Классификация тональности
- Перевод
- Пространственное мышление (выбор места например)
- математика (вычисления)

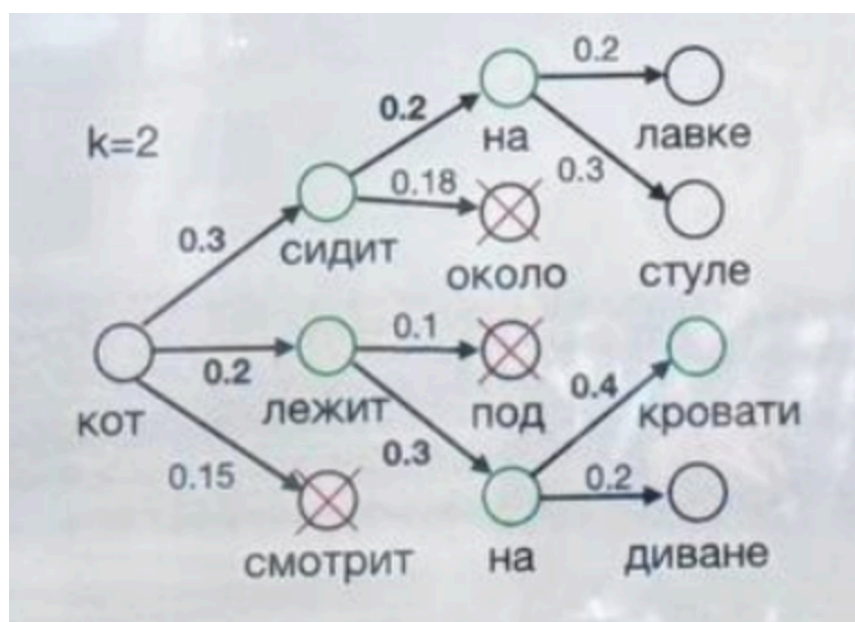
Задачи от бизнеса:

- Боты-помощники
- Разметка данных
- Принятие решений
- Перефраз
- Доменные области (узкая специализация)

Количество параметров или размер LLM:

- 3B (billions) - супер легкие, можно запустить оффлайн на телефоне
- 8B - Для малого бизнеса и тестирование гипотез (чуть помощнее короче, можно доводить качество до 70B)
- 70B - крупный бизнес, можно применять для принятия решений
- > - для доменных областей

Схема Beam Search



Приколы с промптами:

Совет	В чем суть
Сформулировать задачу явно	Назовите конечную цель и нужный формат ответа
Указать контекст	Дать модели нужные данные
Писать пошагово	Попросите думать вслух или решать поэтапно
Задать ограничения	Лимиты на объем, стиль
Показывать пример	демонстрация желаемого формата
Уточнять аудиторию	Уровень знаний, роль читателя
Разбивать запросы	Делить длинные задачи на задачи поменьше