**Введение**

**Что такое LLM**

Большая языковая модель (A large language model (**LLM**)) — это тип модели машинного обучения предназначенной для задач обработки естественного языка, таких как генерация текста. LLM — это языковые модели с большим количеством параметров, которые обучаются с помощью самоконтролируемого обучения на огромном количестве текстов. [(Википедия)](https://en.wikipedia.org/wiki/Large_language_model)

Недавние достижения в области машинного обучения принесли пользу ряду задач, связанных с кодом, таких как перевод кода, суммирование кода и синтез кода.

Задачи генерации включают перевод кода, суммирование кода (подходит ли) и синтез кода.

* **Перевод кода** (код-в-код): Перевод кода — это проблема преобразования исходного кода с одного языка программирования на другой. Эффективный и точный перевод кода имеет ценность в таких сценариях, как миграция устаревшего кода, адаптация программной платформы и т. д.
* **Обобщение кода** (преобразование кода в текст): Цель задачи Code Summarization — генерировать описания кода на естественном языке, который предоставляется в качестве входных данных. Мы выполняем эту задачу в двух вариантах: генерируя сводку на уровне фрагмента, используя пары комментарий-фрагмент, и генерируя сводку на уровне проблемы, используя пары описание проблемы и программный код. Приложения этой задачи включают повышение понятности не прокомментированного или незнакомого кода для впервые просматривающих и упрощение совместной работы, а также обучения.
* **Синтез кода** (текст-код): Задача синтеза кода фокусируется на генерации исходного кода из текстовых входов. Она включает синтез фрагментов и синтез программ. [(статья)](https://arxiv.org/pdf/2206.08474)

**Методы оценки качества моделей**

**BLEU** (Papineni et al., 2002) для оценки задач генерации кода в текст;

**BLEU и CodeBLEU4** (Ren et al., 2020) для оценки задач генерации кода в код и текста в код.