

## Дополнительные задания по разработке персонального ассистента

Дмитрий Калашников, 02.03.2024

Перед вами - дополнительные задания повышенной сложности, в рамках которых вам предстоит разрабатывать телеграм-бота с генеративными моделями.

### Зачем

В каком случае вам стоит выполнять эти задания:

- Вы хотите глубже разобраться в теме генерации текста и больших языковых моделей
- Хотите руками на серьезной практике прощупать, как сделать сгенерированный текст качественным и правдоподобным
- Хотите иметь по итогам курса пет-проект
- Хотите добрать баллы для получения более высокой оценки

### Вопросы

Сразу ответу на возможные организационные вопросы:

- «Обязательно ли выполнять данные задания для успешного прохождения курса?»
  - Нет, баллы за задания по ассистенту дополнительные. Если максимальное суммарное количество баллов за основные тесты равно 100, вы их набрали и сделали дополнительные задания на 60 баллов, то у вас будет 160 баллов, и вы 160% баллов за курс (напомню, на оценку «отлично» нужно набрать 80%)
- «Попробовал запустить ваш ноутбук, ничего не работает, что делать?»
  - Стоит разобраться в чем ошибка, погуглить, поэкспериментировать и решить эту проблему самостоятельно - всё как в реальных проектах : )
- «В вашем шаблоне очень неудобно делать некоторые вещи, что делать?»
  - Можно переписать всё полностью или воспользоваться собственным шаблоном
- «Обязательно ли пользоваться вашей структурой проекта, вашими реализациями модели, вашими датасетами и тд?»
  - Совершенно необязательно, любую часть можно менять под ваши нужды, можно оптимизировать, добавлять методы генерации, использовать любые данные, обучать сколь угодно "большие" модели. Единственное разумное ограничение – реализовать тот подход, который указан в задании (если вы хотите реализовать RNN-сеть вместо модели на основе n-грамм – это не ок).
- «Как и когда задания будут оцениваться»
  - Смотри «Критерии оценивания» внизу
- «Когда дедлайны?»
  - Дедлайн будет одновременно для всех заданий за 1-2 недели до окончания курса, об этом отдельно оповестим в чате
- «Сколько заданий на ассистента будет?»
  - Планируется 3 задания:
    - LM на основе n-грамм

- LM из семейства gpt
- Современная большая языковая модель (>1B параметров), предположительно llama
- «На каких мощностях выполнять задания?»
  - На любых, можете воспользоваться облачными решениями либо собственными.

Относитесь к данным заданиям максимально творчески - любую часть кода можно менять под ваши нужды и желания, можно оптимизировать, добавлять методы генерации, использовать любые данные, обучать сколь угодно "большие" модели.

### **В чем состоят задания**

Проект состоит из двух частей.

- Первая - телеграмм-бот, которого нужно создать, запустить и убедиться в его работоспособности.
- Вторая - обучение моделей в каждом из заданий, сохранение их в подходящем для вас формате и реализация логики в боте, которая данную модель сможет подгружать, запускать и генерировать с помощью нее текст по заданному входному тексту

### **Как устроен телеграм-бот**

Команды:

- [/start](#) старт бота. В целом можно игнорировать эту команду
- [/model](#) выбор модели. Нажимая на эту кнопку, предоставляется выбор модели, которую стоит подгрузить. После нажатия на название модели, должна подгрузиться соответствующая модель с параметрами генерации, после чего текст будет генерироваться с помощью нее.
- [/checkmodel](#) посмотреть название модели, какая модель сейчас загружена
- [/generate](#) сгенерировать текст по контексту (можно использовать без введения команды, просто писать текст, он будет подаваться на вход модели, модель будет генерировать выходной текст)
- [/help](#) – вывести список доступных команд

Чтобы более подробно разобраться в происходящем, необходимо будет погрузиться в код проекта и разобраться что за что отвечает - всё как в реальных проектах : )

### **Небольшие инструкции**

- 1) Как создать телеграм бота.

Это занимает минут 10, можно сделать это по гайду ниже:

<https://habr.com/ru/articles/697052/>

Главное – получить токен вашего бота, который потом вы подставите в переменную TOKEN в проекте. Важно – если вы будете выкладывать свой код куда-то (например, гитхаб, либо отсылать нам), то обязательно удалите свой токен – он персональный.

## 2) Как запустить телеграм-бота

Для начала нужно установить пару библиотек. Чтобы не засорять свое окружение (и не было потенциального конфликта зависимостей), создаем отдельное с помощью miniconda (если не знаете как ее установить и зачем – можете почитать тут: <https://docs.anaconda.com/free/miniconda/index.html>) и устанавливаем туда python 3.9.0. Затем активируем данное окружение и устанавливаем библиотеки. Команды для этого:

```
conda create -n assistant python==3.9.0 -y
conda activate assistant
pip3 install pytelegrambotapi
pip3 install numpy==1.23.5
python3 llm_assistant_bot.py
```

Скрипт llm\_assistant\_bot.py – наш скрипт из репозитория, который отвечает за основную логику бота и обрабатывает все команды.

## 3) Как пользоваться ноутбуками с кодом для каждого из заданий

Как источником для вдохновения. В первом ноутбуке дан код для обучения очень примитивной n-граммной модели, вам предстоит ее улучшить (как минимум, за счет данных и гиперпараметров, максимум – что-нибудь дописать).

### Критерии оценивания:

Главный критерий выполненности данных заданий - телеграм-бот, генерирующий текст и использующий обозначенный в каждом задании подход.

Каждое задание будет оцениваться ближе к концу курса по двум составляющим:

- 1) Наличие и адекватность кода для обучения модели (минимальное соответствие ожидаемому в задании методу. Просто брать из интернета готовую модель – не пойдет) – максимум 10 баллов. В случае обнаружения грубых ошибок баллы будут вычитаться из этих 10ти. Списывание будет караться : )
- 2) Рабочий телеграм-бот – к концу курса анонсируем проверку в удобном для всех формате. Для каждой модели будет проверять субъективное качество генерации в рамках каждого из методов. Максимальное количество баллов – 10. Не стоит в бота подкладывать модель, которую вы не обучали сами, либо которая не соответствует названию – это тоже будет караться : )

### Детали:

- За каждое задание баллы набираются отдельно
  - При желании, можно выполнять не все задания.
- Каждое задание оценивается в одинаковое количество баллов
  - нет более весомых заданий.
- Идеальное выполнение каждого задания эквивалентно идеальному выполнению теоретического + практического тестов по одной лекции
  - то есть, 20 баллов.
- Для сдачи кода и моделей будет предусмотрен отдельный тест в Технотесте.

Таким образом, если вы потеряли 20 баллов на тестах, то можете добрать их с помощью заданий по ассистенту.

### **Итого**

Относитесь к данным заданиям как к дополнительным возможностям для себя и получайте удовольствие : )