Понимание задачи

Чат-бот для генерации текста в инфостиле

16 марта 2023
Ответственный дизайнер: Дмитрий Крюков
Преподаватель: Артём Горбунов

Задача

В качестве дипломного проекта я хочу разработать чат-бот для генерации текста в инфостиле на основе дообученной ГПТ-модели.

Я предполагаю, что проект будет полезен блогерам, эсэмэмщикам, копирайтерам, дизайнерам, менеджерам, редакторам — в общем, всем, кто генерирует информационный контент. Этим людям нужны хорошие тексты, но им может не хватать времени, навыка, денег, возможности оплатить зарубежный сервис из-за санкций.

Вот примеры болей целевой аудитории:

Я дизайнер. Мои макеты в фигме выглядят средненько, потому что в них текст «рыба», а написать хороший текст сразу я не могу.

Я ищу работу. Боюсь отправлять резюме на вакансии, потому что там нужно писать тексты, а я стесняюсь.

Я фрилансер. Мне приходят заказы на фрилансе, а я не берусь за них, потому что не хватает времени вести текущие проекты, было бы хорошо переложить часть работы на сотрудника, но это невыгодно.

Я индивидуальный предприниматель. Только что запустил бизнес по продаже микрозелени, нужно вести социальные сети, но нет денег на контент-менеджера.

Я блогер, мне нужно постоянно генерить контент для удержания пользователей, ну я устаю и не успеваю.

Я услышал о хайпе вокруг Чат-ГПТ, но не смог зарегаться, потому что нет зарубежной карты, а сайт в РФ недоступен. Грущу.

Мой проект не заменит редактора, но подойдёт для некоторых задач, а затрат будет меньше, чем при найме редактора. Доступ к нейросети будет реализован через привычный интерфейс в телеграм-боте без установки дополнительных программ. Бот будет доступен на телефоне, планшете и компьютере, а подписку можно будет оплатить с карты РФ.

Задача: помочь людям, которые пишут информационный контент делать это быстрее, дешевле и лучше.

Решение: модель на основе ГПТ-алгоритма дообученная на текстах в информационном стиле с

доступом через бот в телеграме.

Для обучающей выборки мы будем использовать тексты, написанные в инфостиле. Для начального обучения можно самостоятельно написать несколько абзацев текста под разные запросы. Даже малое количество текста позволяет донастроить модель до целевого стиля. Для расширения выборки можно взять тексты из блога Максима Ильяхова (с его разрешения).

С точки зрения закона обучение искусственного интеллекта на текстах, защищённых авторским правом, не нарушает доктрины правомерного использования. Обученная модель — чёрный ящик. Нельзя установить, на каких исходных материалах она обучалась. Но чтобы обезопасить себя от юридических рисков в будущем, перед коммерциализацией можно провести обучение в два этапа. Сначала модель обучится на исходных текстах, а после обучится повторно на уже сгенерированных текстах. Таким образом в модели не останется материалов третьих лиц.

Для обучения нейросети подготовим набор данных, который будет состоять из текстовых файлов в джейсон-формате. Для предобработки текстов будем использовать токенизацию. Затем полученный датасет мы будем использовать для дообучения ГПТ-модели при помощи Python, Transformers и Pytorch Lightning.

Параметры обучения:
1. Количество эпох обучения: 5.
2. Размер батча: 64.
3. Коэффициент скорости обучения: 0,0001.
4. Функция потерь: перекрёстная энтропия.

После обучения модели мы проведём оценку качества модели на тестовых данных. Для этого будут использованы метрики преплексия и точность с целевыми значениями 13 и 70 % соответственно.

Часть про обучение нейросети сложно выглядит, но фактически всё делается несколькими консольными командами. В разработке поможет мой знакомый программист, у которого уже есть опыт работы с ГПТ-моделями. Также для удобного доступа мы разработаем телеграм-бот, который будет работать, как тонкий клиент — перенаправлять запросы в разные АПИ и отдавать ответ пользователю.

Интересы сторон

1. Для пользователей: получить удобный инструмент редактуры.
2. Для меня: получить удобный инструмент редактуры и попробовать себя в роли стартапера.

Стоимость разработки

1. Домен — 199 ₽.
2. Wildcard SSL — бесплатно.
3. Сервер — 3000 ₽.
4. Аккаунт в «ОпенАИ» для тестирования — 1000 ₽.

Монетизация

Один запрос к «ОпенАИ» стоит примерно 1 ₽. Исходя из этой стоимости я предлагаю примерно такую раскладку тарифов:

1. 20 запросов — 100 ₽ (5 ₽/запрос).
2. 200 запросов — 800 ₽ (4 ₽/запрос).
3. 1000 запросов — 2500 ₽ (2,5 ₽/запрос).

После первого месяца работы мы посчитаем аналитику по запросам и уточним цены.

Элементы дизайна

1. Дообученная модель.
2. Интерфейс бота — тексты для контроллов.
3. Сценарии бота — тексты для приветствия, обучения и других сценариев.
4. Оформление бота — аватарка-логотип.
5. Продвижение проекта — маркетинговые тексты и письма для блогов.

Функциональность бота

1. Авторизация. Она нужна, чтобы корректно привязывать запросы к пользователю.
2. Оплата подписки. Мы интегрируем платежный сервис «Пей-бокс».
3. Запрос. Формулируется на естественном языке, например «напиши понимание задачи на разработку нейросети в школу Бюро». Запрос отправляется в «ОпенАИ» и возвращает ответ от нейросети.
4. Вывод количества оставшихся оплаченных запросов. Если счётчик близок к нулю, отправляем пользователю предупреждение.
5. Логирование запросов и ответов для аналитики и кэширования.

Этапы и сроки

Начало работ: 10 апреля 2023, завершение работ: 19 мая 2023.

Точка невозврата: 5 мая 2023.

Этап, *состав команды*	**Результат**	**Срок, раб. дней**
1. *Вся команда:* понимание задачи, подготовка архитектуры	Понимание задачи и уточнённый план проекта, архитектура проекта	3 дня 10–13 апреля
2. *Редактор, дизайнер и разработчик:* подготовка обучающей выборки и точная настройка ГПТ-модели	Подготовленная обучающая выборка и обученная модель	3 дня 14–18 апреля
3. *Разработчик:* подготовка АПИ для	Готовое и развернутое АПИ	7 дней 19–27 апреля

телеграм-бота		
4. *Разработчик и дизайнер:* разработка телеграм-бота	Работающий телеграм-бот	6 дней 28 апреля – 5 мая
(Точка невозврата)		
5. *Руководитель, дизайнер, редактор, маркетолог:* план продвижения	Разработанные промоматериалы в виде текстов и изображений	3 дня 8–10 мая
6. *Вся команда:* Тестирование и дебаг проекта	Баглист и подфикшенный проект	4 дня 11–16 мая
7. *Руководитель, редактор:* запуск и промо проекта	Опубликованные промоматериалы, запуск рекламы	3 дня 17–19 мая
	Итого	**29**

Технические требования

Формат файлов

Разрабатываемые элементы передаются в виде набора файлов следующих форматов:

1. Сценарии и тексты для бота — ссылка на Гугл-документы.
2. Обучающая выборка — джейсон-файл.
3. Растровые изображения в пнг-формате.

Языковые версии

Бот создается в русской языковой версии. Обучающая выборка также на русском языке.

Программно-аппаратная платформа

Бот написан на платформе Node.js версии 16, язык разработки — Typescript. База данных PostgreSQL, для взаимодействия с базой — ORM Prisma. Для упрощения взаимодействия с АПИ Телеграма будем использовать Node-telegram-bot-api. Веб-сервер на базе Nginx.

Принципы совместной работы

В основе совместной проектной работы лежат принцип fix time and budget, flex scope («ФФФ») и общая ответственность за срок выхода конечного продукта.

В команде предусмотрены несколько ролей, которые делятся на две группы:

1. Ученики Школы — руководитель, редактор и дизайнер.
2. Внешние подрядчики — разработчик, специалист по рекламе.

Команда еженедельно согласовывает дизайн и функциональность с учётом фактического состояния разработки с арт-директором — преподавателем Школы. В процессе принятий решений

команда корректирует приоритет разработки.

Планируемые сроки включают время на согласование результатов.

В плане присутствует точка невозврата. Она разделяет план на реализацию элементов дизайна и доработку, исправление ошибок, наполнение информацией, тестирование и подготовку к запуску. После точки невозврата в продукт не может добавляться новый функционал.