

академия
больших
данных



mail.ru
group

Нейро телеграмм-канал

Автор идеи: Сергей Парамонов

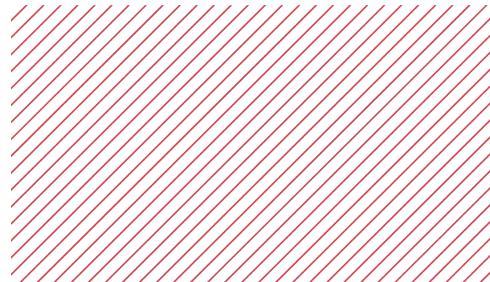
Ментор: Сергей Парамонов

Команда:

Дмитрий Клебанов – подготовка данных,
бэкенд и деплой

Дмитрий Сермягин – разработка бейзлайна,
построение пайплайна обучения

Владислав Астрашаб – обучение модели,
менеджмент проекта



Описание проекта



Описание проекта



Пирожки

в купе заходит ума турман
все ой какая ми ми ми
а та давай рубить им бошки
дверьми



14.6K 10:15 AM





Описание контента для канала

Пирожки – четверостишья

Особенности:

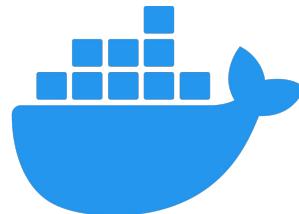
- по форме:
 - без пунктуации, без больших букв
 - белый ямб
- по содержанию:
 - нестандартное использование языка
 - языковые игры, неожиданные смысловые повороты



Где применять и кто пользователи?

- Применение проекта – развлечения и популяризация машинного обучения.
- Целевая аудитория – любые пользователи телеграмма, заинтересованные в результатах работы нейронных сетей.
- Монетизация – реклама в тг-канале. Возможна генерация пирожков по заданным первым словам, например с названием рекламируемого бренда.

Стек технологий



docker





Данные

- Примеры пирожков для обучения собирали сами с канала @pirozhki_page.
- Мы отфильтровали их от рекламных и прочих постов и получили 2968 примеров для обучения.
- Мы разделили их на множества для тренировки и валидации, чтобы с помощью метрики perplexity следить за тем, что модель не переобучается.

жизнь загружалась постепенно
сначала появился свет
за ним последовал поджопник
и возглас мальчик три семьсот

из множества различных мнений
что существуют обо мне
я выбрал самое плохое
и строго следую ему



Модель и обучение

- Использовали предобученную Russian GPT-2 и инфраструктуру Google Colab.
- Тренировали модель 200 эпох на GPU.

СЕРЕБРЯ живёт
в очень маленьком
городе. Там всегда мало
людей.

По-взрослому
одна только обкололась
и сразу стала
действовать

Пойду я и достану
глеба
и сделаю из него
лифчик

Демонстрация проекта

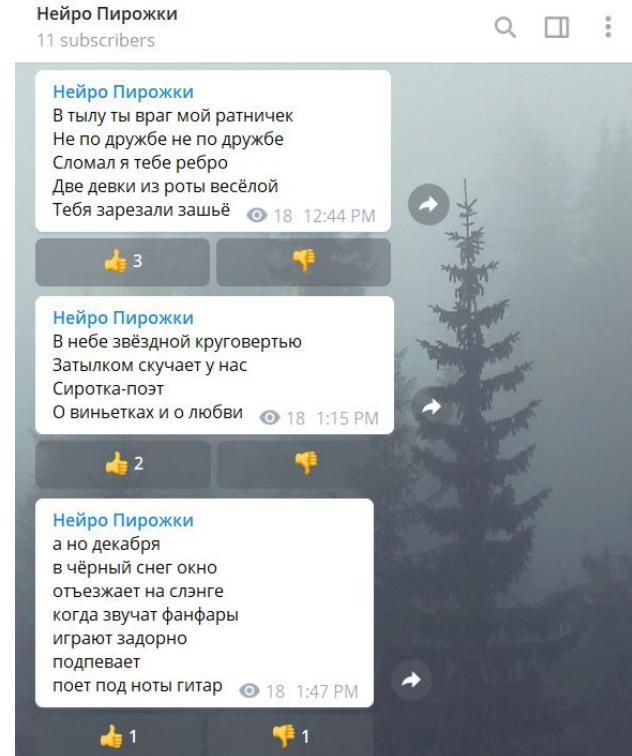
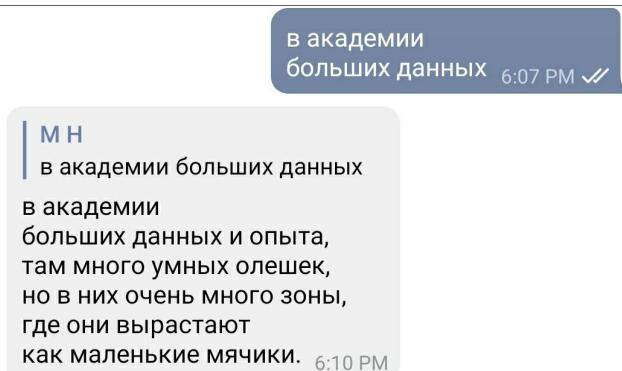
262

263

Демонстрация проекта

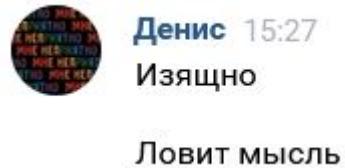
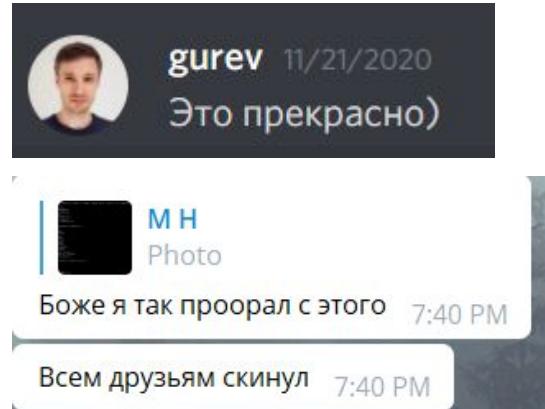
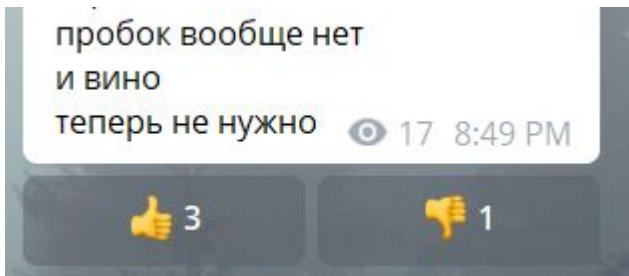
Посмотреть на результаты работы можно в телеграм-канале https://t.me/pirozhki_neuro.

Попробовать сделать “рекламный” пирожок можно отправив боту @pirozhki_neuro_bot слова, с которых он будет начинаться (например название того, что рекламируем)



Оценка работы

- Исходя из развлекательных целей проекта, мы собрали субъективные оценки сгенерированных результатов от потенциальных представителей целевой аудитории.
- Мы предоставили пользователям возможность оценивать каждый пост с помощью реакций @ReactBot.



Запуск бота на сервере

- Бота можно легко запустить на удалённом сервере или локально, для этого применяется docker. Можно настраивать частоту постинга и температуру для генерации. Инструкция к запуску и настройке в репозитории проекта:
https://github.com/dmitriyklebanov/made_2020_project
- Стоит предусмотреть чтобы сервер обладал достаточным количеством ресурсов для запуска модели. Образ занимает 5.25 Gb на диске. Модель может быть запущена как на сри, так и на гри.

3.2 GB

MEMORY USAGE

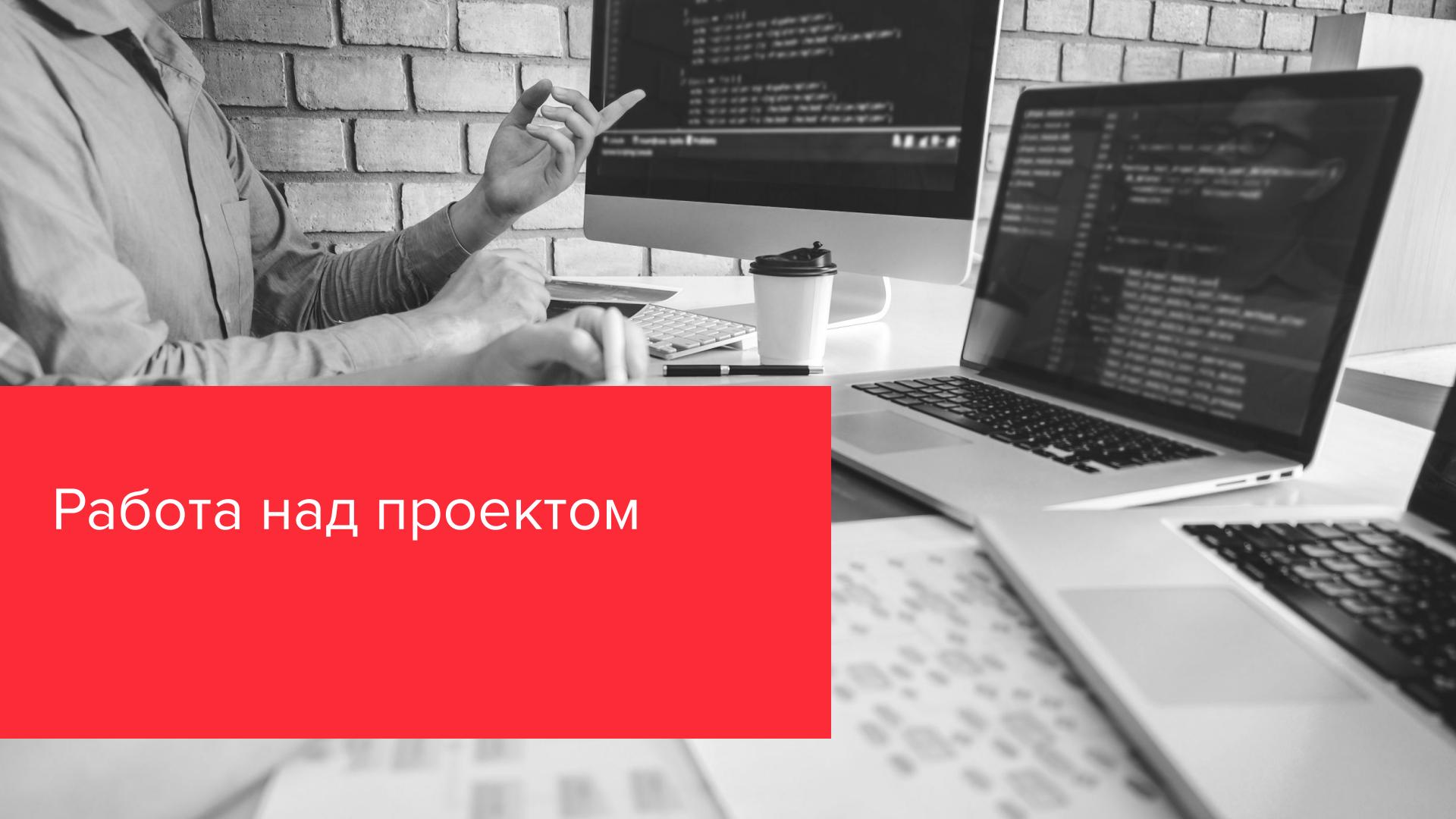
9.5 kB / 2.3 kB

NETWORK I/O



Дальнейшее развитие проекта

- Сбор и хранение данных о реакциях для построения модели, оценивающей успешность пирожка перед постингом.
- Разделение бота для постинга и бота для создания рекламных пирожков. Необходимо предусмотреть чтобы пользователи не могли перегрузить бота работой, и он в первую очередь отвечал администраторам канала.
- Расширение датасета и эксперименты с моделями с целью улучшения качества контента.



Работа над проектом

Обновленный Roadmap

Product Plan

Aa Step	✔ Status	📅 Deadline	☰ Owner
Come up with an idea	Done	Oct 31, 2020	Uladzislau Astraszab Dmitry Klebanov
Collect data	Done	Nov 7, 2020	Dmitry Klebanov
📄 Create a model pipeline, make a baseline	Done	Nov 14, 2020	Dmitry Sermyagin
Train model on GPU	Done	Nov 21, 2020	Uladzislau Astraszab
Deploy the bot	Done	Nov 28, 2020	Dmitry Klebanov



Команда



Команда

Ментор: Сергей Парамонов

Дмитрий Клебанов – сбор и подготовка данных, бэкенд и деплой

Дмитрий Сермягин – разработка базлайна, построение пайплайна обучения модели, презентация

Владислав Астрашаб – обучение модели, менеджмент проекта, подготовка презентации



Спасибо!

Контакты участников:

dmitriy.klebanov@gmail.com

sermyagin@gmail.com

u.astraszab@gmail.com