

*Второе демо проекта*

---

iPavlov

Евгений Егоров  
Илья Игашов  
Арсений Крохалев  
Евгений Рубаненко

---

Какую проблему  
пользователя будем решать?

---

# Проблема пользователя

---

- Тратится много времени и сил на создание собственных моделей и алгоритмов работы с русским текстом
- Отсутствие библиотеки и инструментов для работы с русскими текстами



Scope

---

# Scope

---

- **Сбор данных**  
XPath, BeautifulSoup  
Яндекс.Толока
- **Коммерческие проекты с применением собственных технологий**  
Онлайн-консультанты  
Чат-боты
- **Библиотека с инструментами для обработки данных**

В чем заключается  
решение?



---

# 1. Data scientist

---

## **Проблема:**

нет библиотеки, которая одновременно предоставляет необходимые данные, а также удобные инструменты для работы с ними

## **Предлагаемое решение:**

библиотека, в которой будут собраны данные, модели и средства для работы с ними

## **Сроки:**

планируется закончить разработку к концу 2018 года

---

# 1. Data scientist

---

## Рыночные факторы:

### Спрос:

- все респонденты, которым в исследованиях необходимо работать с данными на русском языке (57% от всех опрошенных), считают, что такая библиотека необходима
- ученые, работающие с более распространенными языками (напр., английский) (32% от всех опрошенных), не видят особой нужды в такой библиотеке, но будут рады новым удобным функциям

### Существующие предложения:

Open-source библиотеки: nltk, TensorFlow, Keras, и др.



---

## 2. Стартап

---

### Проблема:

за короткий период времени необходимо сделать рабочий продукт или его прототип

### Предлагаемое решение:

обеспечить разработчиков всеми необходимыми инструментами для быстрого развертывания / разработки своего продукта (удобные функции для обработки данных, возможность легко и быстро сделать чат-бота, использующего ML & DA и др.)

### Сроки:

планируется начать разработку в ближайшее время (как будет разработан MVP (см п.1) и получено дополнительное финансирование);  
на реализацию закладывается 1 год

---

# 2. Стартап

---

## Рыночные факторы:

### **Спрос:**

все 100% опрошенных считают, что данные инструменты необходимы, а те, которые есть, — написаны недостаточно качественно; респондентам не хватает существующего функционал

### **Существующие предложения:**

под каждую из задач есть разные решения, но все они состоят из множества трудоемких шагов: Heroku, Dialogflow

---

# 3. Большой/малый бизнес

---

## Проблема:

неэффективность существующей системы поддержки клиентов

## Предлагаемое решение:

бот-помощник / сайт / десктопное приложение / фреймворк, который автоматически подсказывает человеку из команды поддержки варианты ответа на вопрос пользователя

## Сроки:

данный инструмент находится в финальной стадии разработки  
оценочное время завершения разработки — лето 2018



---

# 3. Большой/малый бизнес

---

## Рыночные факторы:

Основным фактором стал “заказ” Сбербанка

## Спрос:

- многие из опрошенных (63% от всех опрошенных) высказались, что используют ботов-помощников
- так же было замечено, что компании работают в этом направлении (напр., Тинькофф).

## Существующие предложения:

- сейчас все разработки в данной области являются результатом работы отдельной команды разработчиков; многие до сих пор пользуются текстовыми скриптами для решения данной проблемы (напр., колл-центры);
- какого-то инструмента для создания такой системы еще нет

# Метрики

# Метрики

Целевое действие	Целевая метрика
Распознавание речи	Процент правильно распознанных слов
Разбивка слов по группам частей речи, на смысловые группы.	Процент слов, у которых корректно определен смысл
“Понимание” текста	Процент предложений, в которых верно определена тема
Скачать библиотеку	Количество скачиваний библиотеки
Оставить отзыв, создать продукт на основе нашей библиотеки	Количество довольных пользователей



Что нового узнали?

---

# Что нового узнали?

---

1. Яндекс.Толока — сервис для сбора размеченных данных. Очень полезный инструмент, помогающий размечать новые данные. Загружаешь свой датасет в систему, а люди размечают его за небольшое вознаграждение
2. Javascript (jQuery)
3. Работа на сервере (компьютеры с мощными видеокартами, подключение по ssh), нейронные сети (PyTorch)
4. Работа с Web: скачивание текстов с сайтов, работа с Vk API