

16 ИЮЛЯ 2022

# МОЙ ПЕРВЫЙ КРАУДСОРСИНГ

Где брать данные,  
если твой проект уникален?

Виктория Фирсанова

GitHub: @vifirsanova

TG: @vifirsanova / @hot\_tempered\_healer

VK: @missvector





## Познакомимся?

КЛЮЧ К ПОНИМАНИЮ

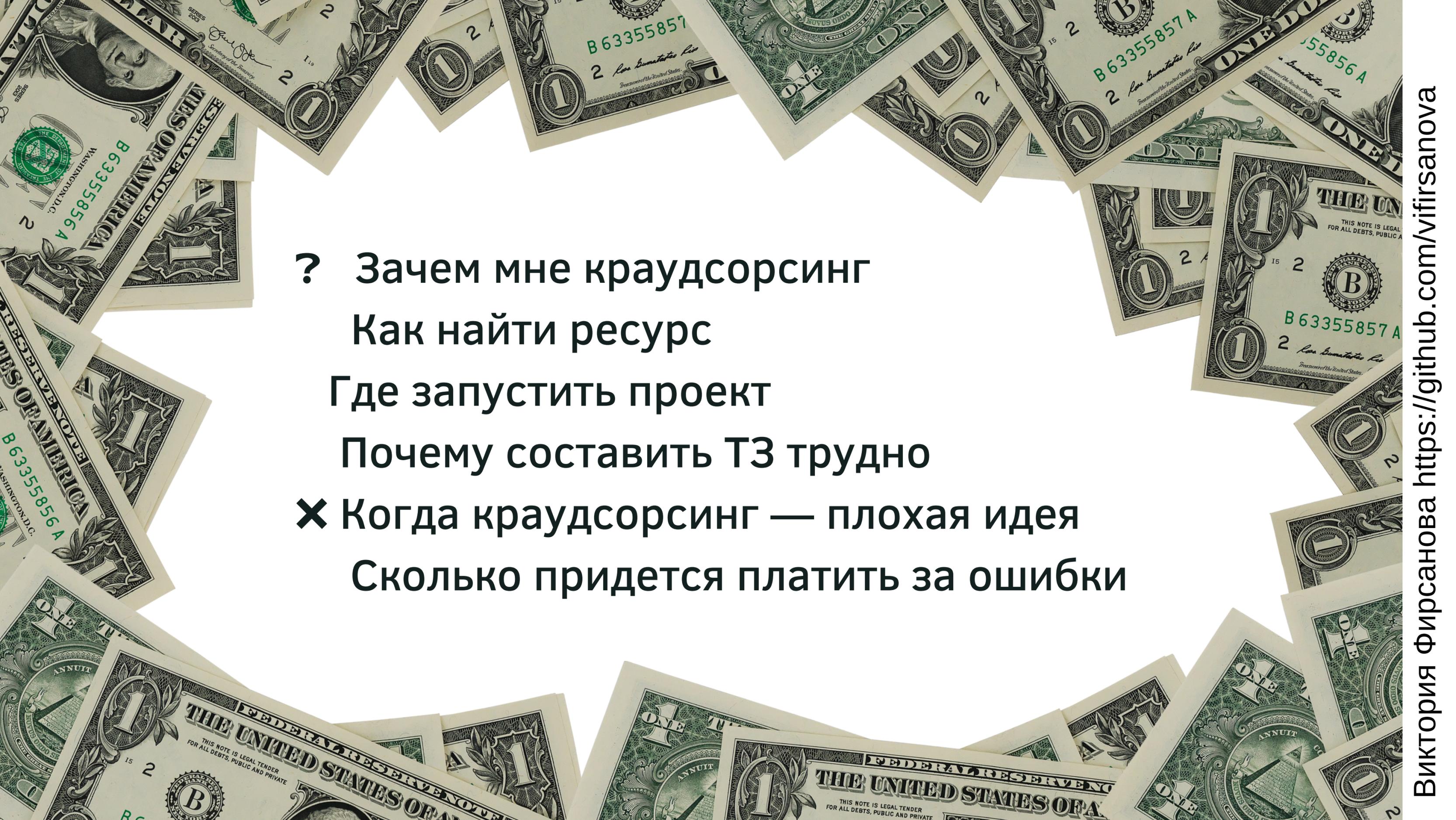
# ASD QA

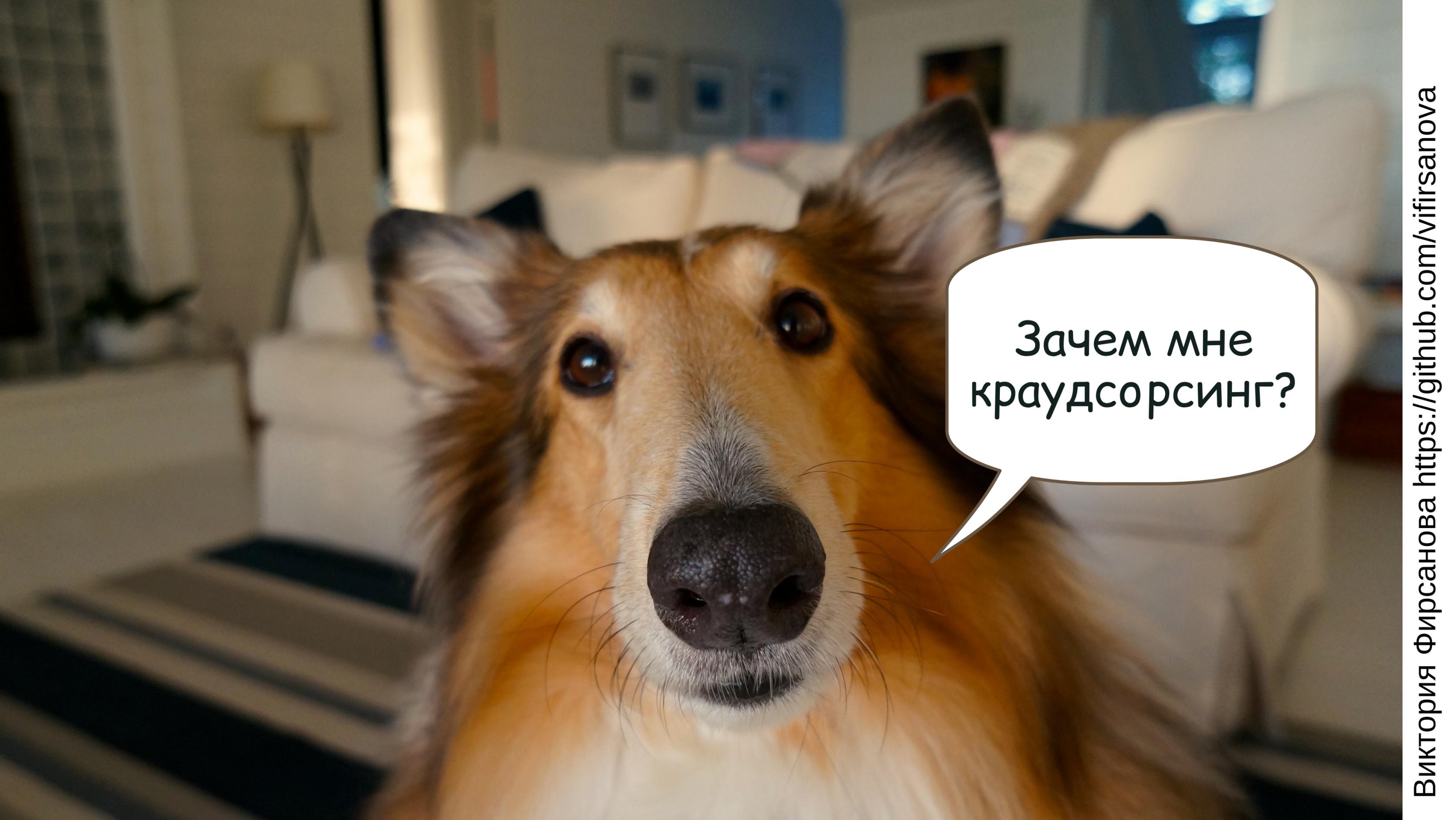
Инструменты для разработки диалоговых систем о расстройствах аутистического спектра (РАС)

[Набор данных](#)[Репозиторий](#)

```
"question": "Аутизм - это отклонени  
"answers": [  
  {  
    "text": "Я родился со своими ун  
    "answer_start": 152,  
    "answer_end": 238  
  }  
],  
"is_impossible": false  
},  
{  
  "question": "Расскажи мне новости?"  
  "answers": [  
    {  
      "text": "Я не могу ответить на  
      "answer_start": 0,  
      "answer_end": 0  
    }  
  ],  
  "is_impossible": true  
},  
],  
"context": "Пожалуйста, не осуждайте ме
```



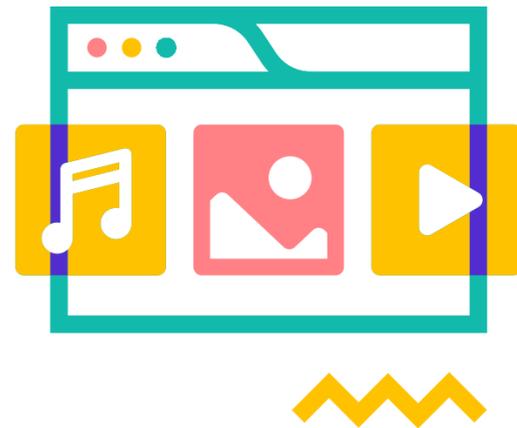
- 
- ? Зачем мне краудсорсинг
    - Как найти ресурс
    - Где запустить проект
    - Почему составить ТЗ трудно
  - ✘ Когда краудсорсинг — плохая идея
    - Сколько придется платить за ошибки



Зачем мне  
краудсорсинг?

# Крауд-что?

- Краудсорсинг - процесс получения **услуг** методов принятия **пожертвований** (контрибьюшн) от большой группы людей, чаще **онлайн**.



# Мировой уровень



## Toyota Dream Car Art Contest Artworks Gallery

Car for the Visually Impaired Heng Wei Xu This car enables the visually impaired people to drive around. The car is equipped with a visual helmet for them which transmits the immediate image to their brain, and a seat belt wearable by just pressing a button. They can also drive automatically leaving their hands free.

[Toyota Dream Car Art Contest Artworks Gallery](#)

кибербезопасность

творческий контент

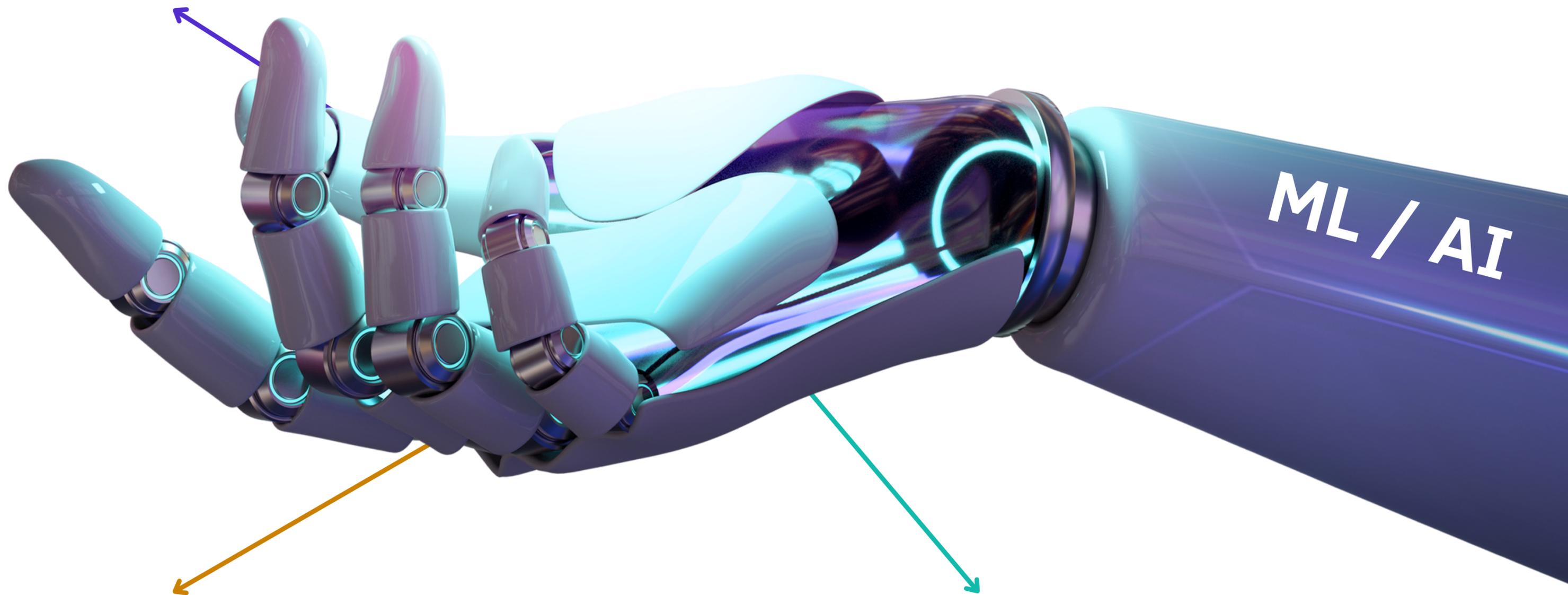
КРАУДСОРСИНГ

ML / AI

маркетинг

наука

**разработка алгоритма**

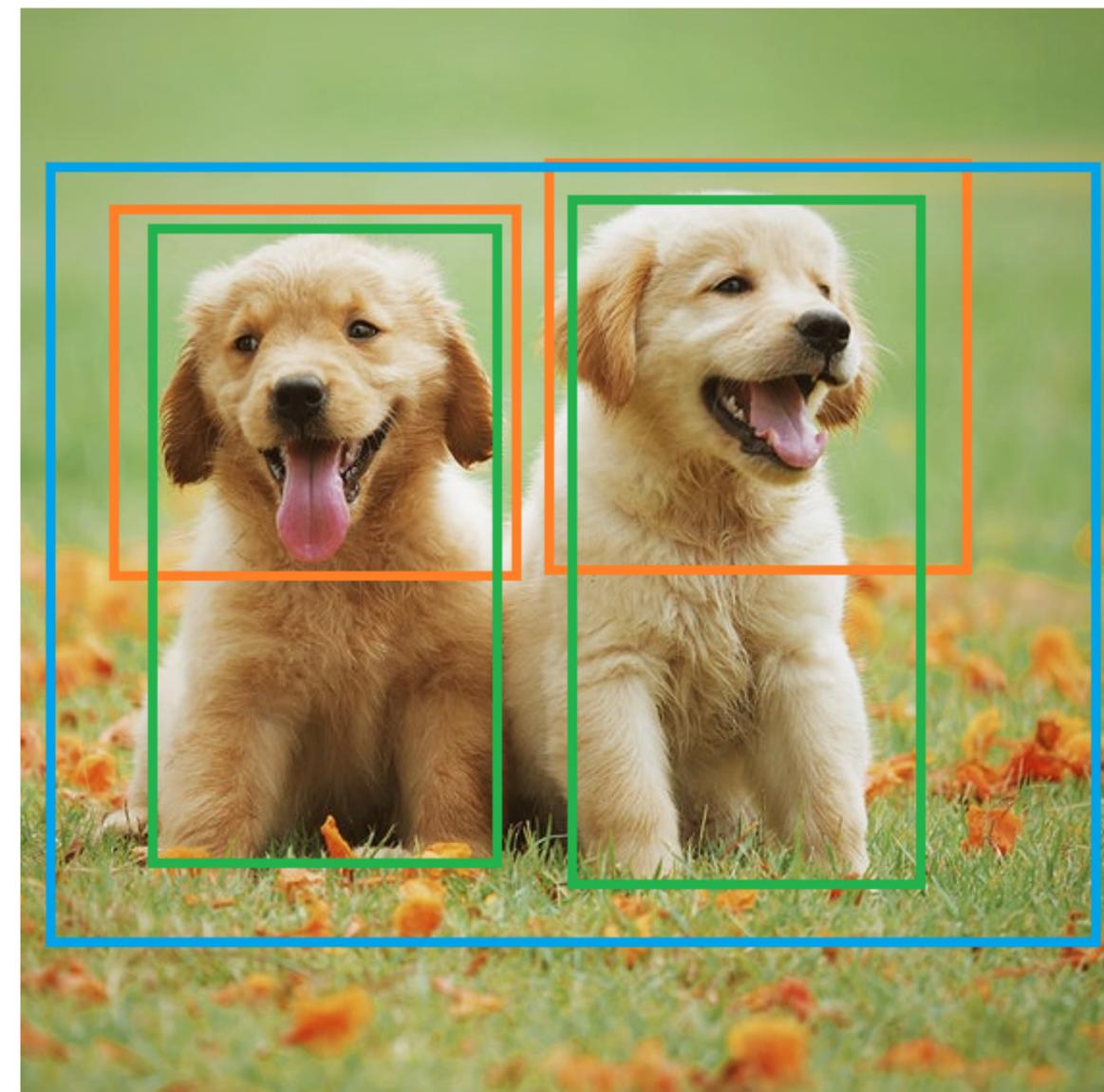


**разметка данных**

**тестирование системы**

# Краудсорсинг для разметки

- Справляемся с большим объемом данных, не выходя из дома.
- Экономим деньги и время.
- Некоторые бизнесы используют краудсорсинг вместо моделей машинного обучения.



# Краудсорсинг для разработки

- Соревнования по DS и ML
- Разработчик получает
  - условия задачи
  - данные для ее решения
- Организация ожидает
  - алгоритм-решение
- Призовая сумма ниже з/п
- Соревнование мотивирует:
  - получаем высокую точность

**AIcrowd** 

**innocentive**

**Analytics  
Vidhya** 

**CodaLab** 

 **datacamp**

**kaggle**

# Краудсорсинг для тестирования

- Много тестов
- Тестировщики из разных уголков страны / мира
- Тестирование ЦА
  - таггированная «толпа»
- Согласно статистике может быть быстрее и качественнее



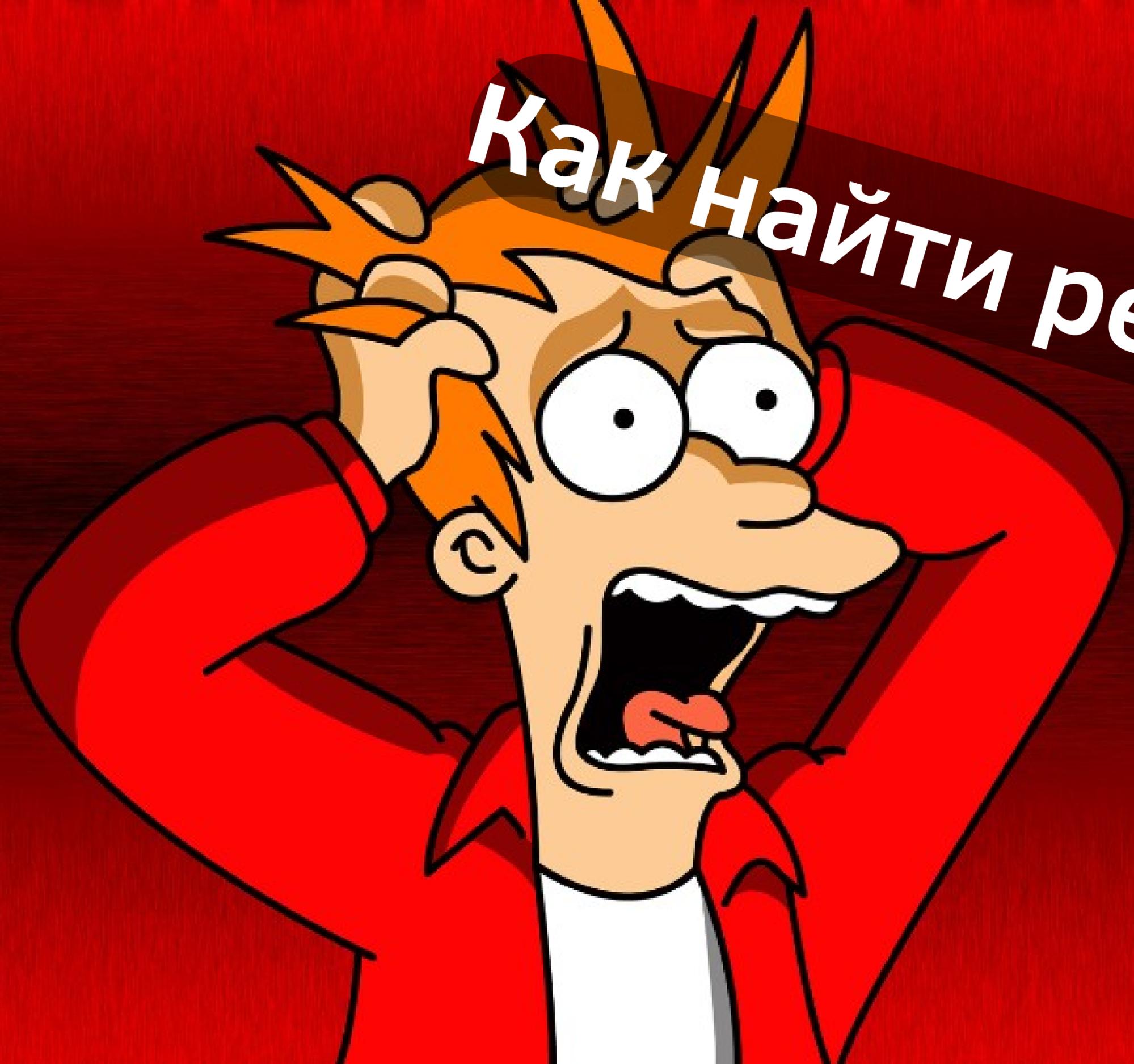
призовые деньги

суммы

вложения

заработная плата

экономить средства



# Как Найти ресурс



# Toloka grant program

## Toloka grant program

Our grant program aims to support the academic community and encourage research in the field of Responsible AI.



A man with a mustache, wearing a dark jacket and a gold watch, is shown in a thoughtful pose with his hand to his chin. A speech bubble is overlaid on the image, containing text in Russian. The background is a blurred outdoor setting.

Используй  
краудфандинг для  
краудсорсинга

# Краудсорсинговый проект

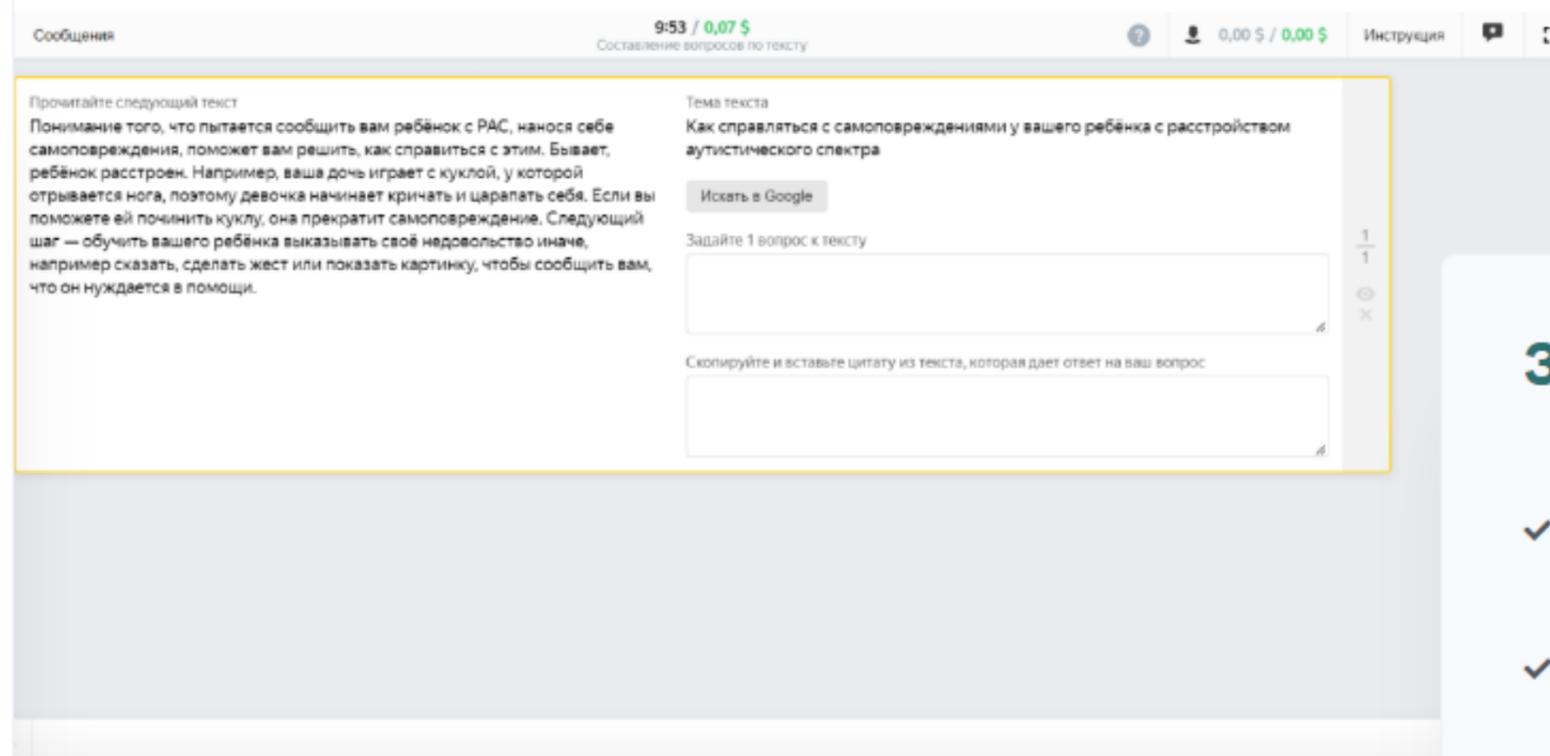
Проект реализуется победителем конкурса «Практики личной филантропии и альтруизма»

Благотворительного фонда Владимира Потанина.

asd-qa.org

## Toloka

Часть данных  
собрана с  
помощью  
исполнителей  
краудсорсинговой  
платформы  
Yandex.Toloka



## Задания

- ✓ Составление вопросов
- ✓ Поиск ответов
- ✓ Описание изображений

# Где запустить проект





Яндекс  
Толока

amazon  
mechanical turk



TELUS<sup>®</sup> International

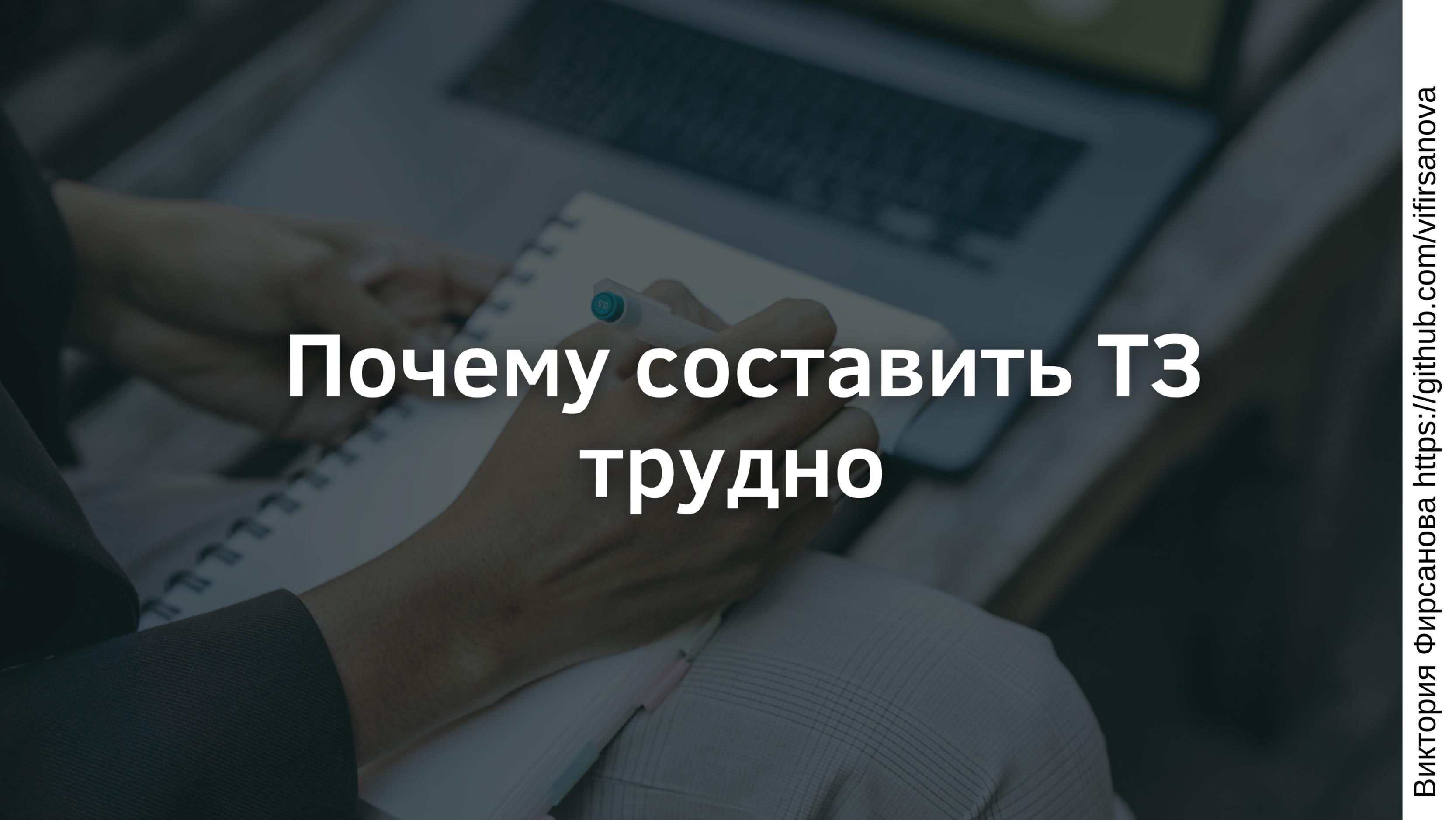


clickworker

# Критерии выбора платформы

- Размер комиссии
- Характеристики исполнителей
  - Количество
  - Возраст
  - Представителей какой страны, носителей какого языка больше
  - ...
- Удобство интерфейса
  - Гибкость программирования интерфейса
  - Простота использования





# Почему составить ТЗ трудно

# Задание снаружи

Прочитайте следующий текст

Помимо социализации наиболее частая тема при обсуждении аутизма – самостимулирующее поведение или "стимминг". Стиммингом называют явно не функциональные, повторяющиеся движения. Рассмотрим несколько примеров того, что относят к стиммингу. "Стимы" могут включать такие движения, как раскачивание взад и вперед, накручивание на пальцы и дерганье себя за волосы, покачивание из стороны в сторону, кружение на одном месте, быстрая ходьба туда-сюда, тряска кистями рук или расчесывание кожи.

Тема текста

Самостимулирующее поведение - стимминг

Искать в Google

Задайте 1 вопрос к тексту

что называют стиммингом

Скопируйте и вставьте цитату из текста, которая дает ответ на ваш вопрос

Стиммингом называют явно не функциональные, повторяющиеся движения

1/1



# Интерфейс внутри

## Конфигурация

```
1 {
2   "view": {
3     "type": "view.list",
4     "items": [
5       {
6         "type": "layout.columns",
7         "items": [
8           {
9             "type": "view.text",
10            "content": {
11              "type": "data.input",
12              "path": "context"
13            },
14            "label": "Прочитайте следующий текст"
15          },
16          {
17            "type": "view.list",
18            "items": [
19              {
20                "type": "view.text",
21                "content": {
22                  "type": "data.input",
23                  "path": "title"
24                },
25                "label": "Тема текста"
26              },
```

## Предпросмотр

Прочитайте следующий текст  
Текст из поля "context"

Тема текста  
Текст из поля "title"

Искать в Google

Задайте 1 вопрос к тексту

Скопируйте и вставьте цитату из текста, которая дает ответ на ваш вопрос

## Спецификация данных ?

### Входные данные

context (строка)

title (строка)

search\_url (ссылка)

### Выходные данные <>

question (строка)

answer (строка)

**Недостаточно  
подробное описание  
приводит к одинаковым  
ошибкам.**

**При этом люди хотят  
заработать как можно  
скорее, а потому часто  
торопятся и не читают  
длинные описания.**

---

Вопросно-ответные пары должны соответствовать теме текста.

---

Непонятно, о чем спрашиваете. Вы просто скопировали тему текста, такие решения очень нежелательны.

---

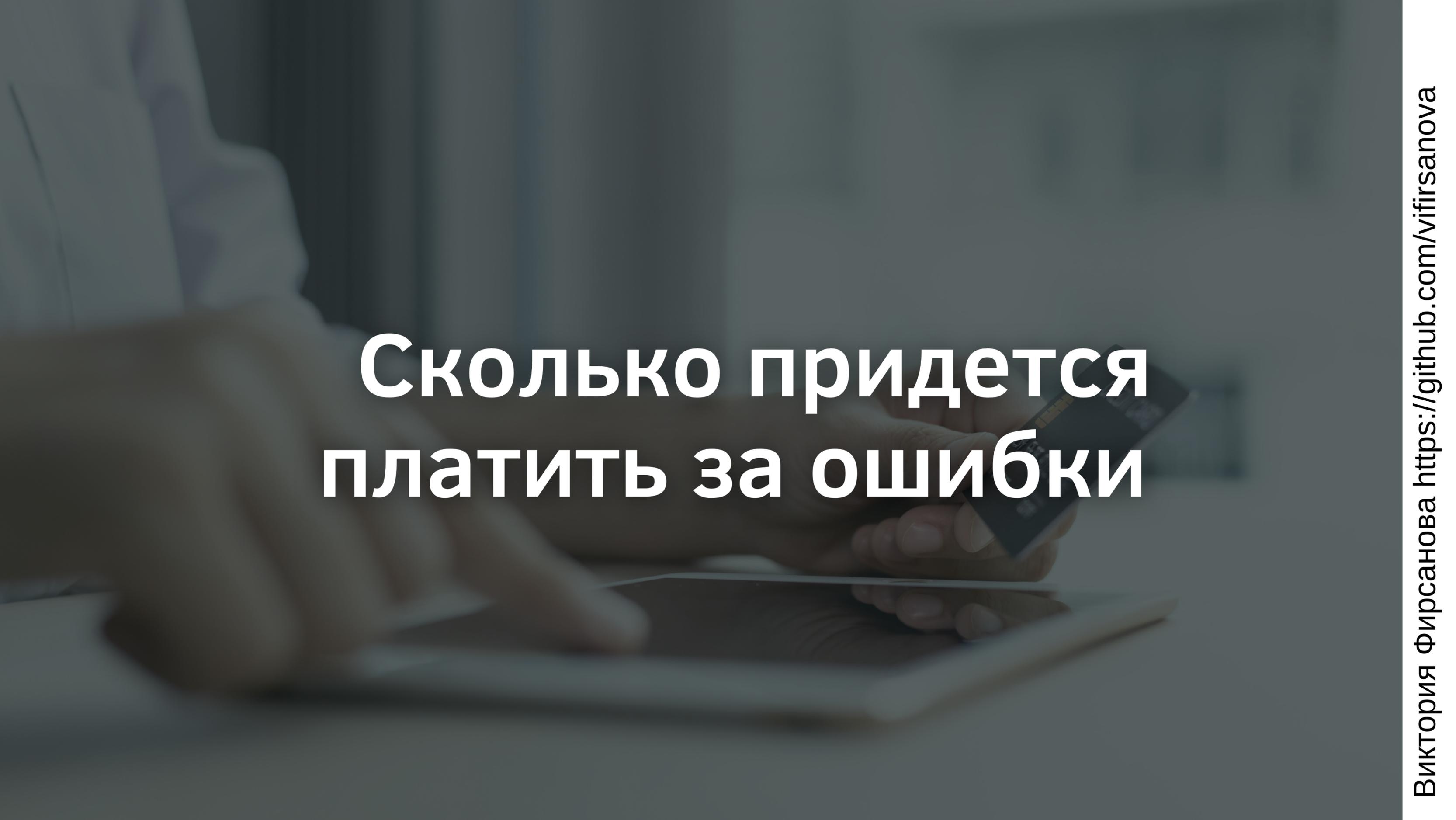
Пожалуйста, старайтесь задавать вопросы в полной форме. Из Ваших запросов, подходящих скорее для поисковой системы, не до конца понятно, какую конкретно информацию Вы пытаетесь получить.

---

В тексте не дано определение сенсорной обратной связи напрямую.

# Когда краудсорсинг — плохая идея





# Сколько придется платить за ошибки

# Интерфейс предварительно тестируем на нескольких образцах задания

[5c2aac02b83da74a355aacef5e1c1f04](#)

Составление вопросов по теме — 2022-05-05

Здравствуйте! У вас нерабочее задание.

Когда писал ответ на первую тему текст дублировался во все остальные колонки!

Я

Здравствуйте! Спасибо, что оперативно предупредили! Сейчас перезапустим.

Просим прощения за доставленные неудобства.

# Описание задания делаем ёмким, добавляем подсказки в интерфейс и используем... помощь друга :)

Задайте 1 вопрос к тексту

Скопируйте и вставьте цитату из текста, которая дает ответ на ваш вопрос

## Инструкция

Мы создаем набор данных для ориентируясь на информацию и

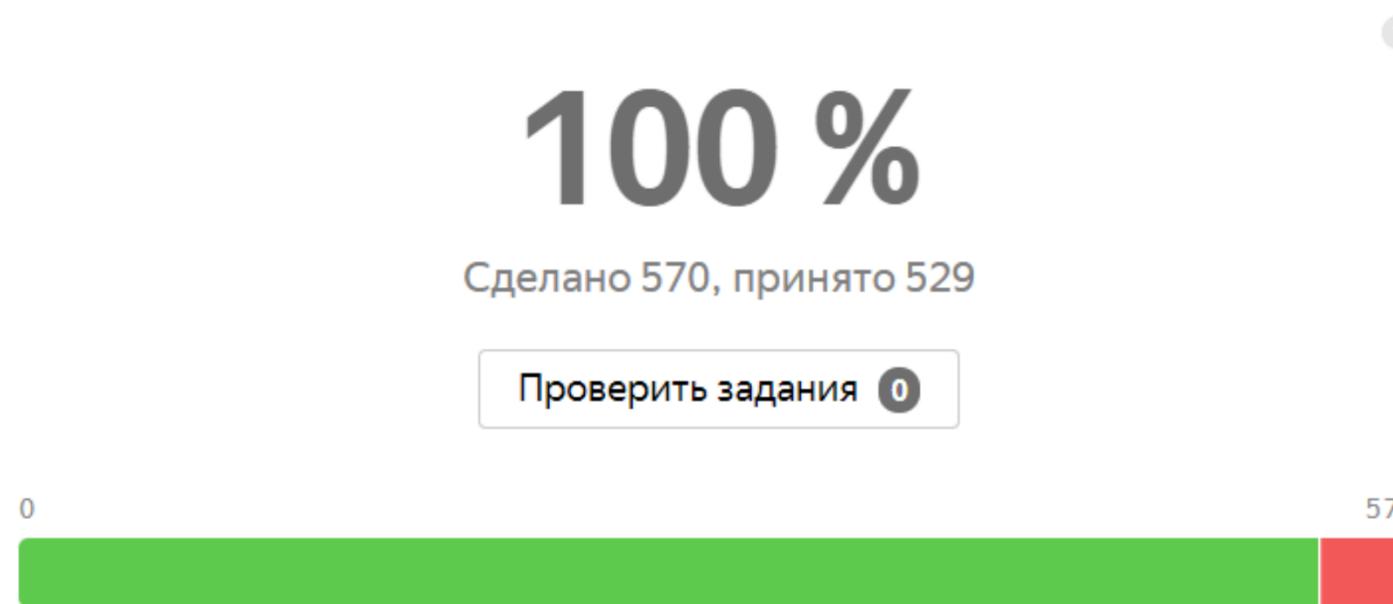
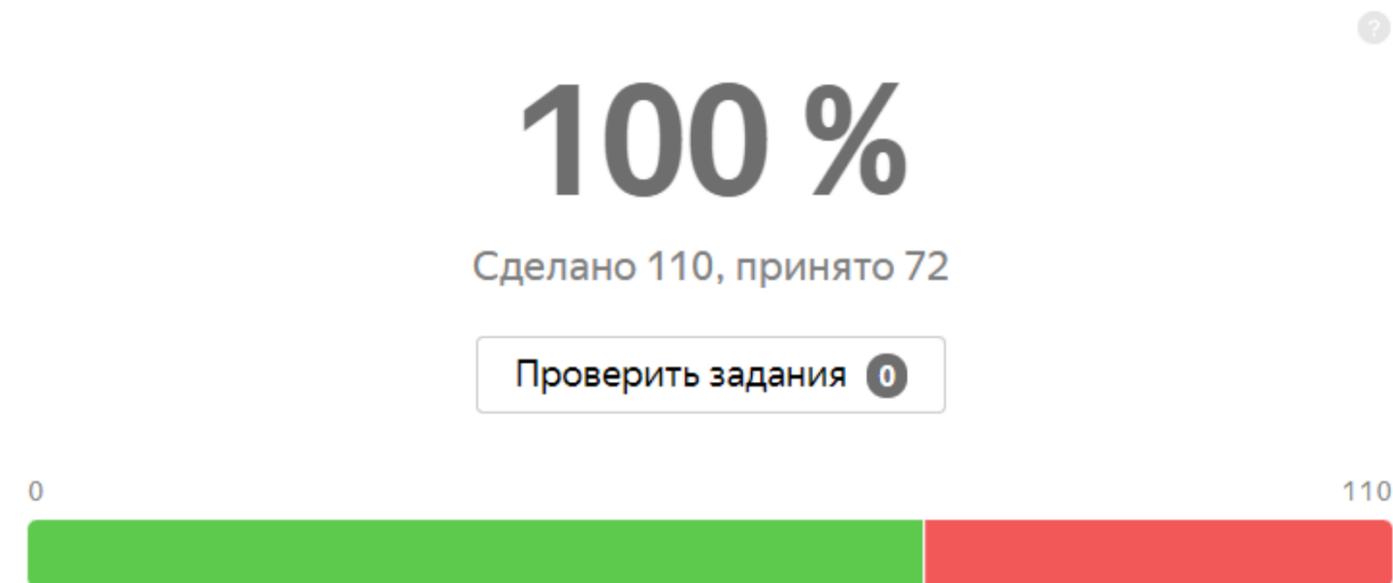
### Как выполнить задание

1. В задании вы увидит текста и поищите и
2. Задайте один вопро тексте в виде цитать тексте перечисляютс
3. Найдите в тексте от *Сергеевич Пушкин* р вам кажется, что вар варианты через запя

## Требования

**Объем вопросов.** Длина вопро допустимы, если вы использует *спектра*" и *синдром Аспергера* испытываете неуверенность, ил

**Начинаем с  
небольших пулов,  
играем в "злого копа"  
- отклоняем  
категорически все,  
что нам не подходит,  
подробно описываем  
причину отклонения**



**NEVER END A PRESENTATION**



**WITHOUT A PICTURE OF A CAT**