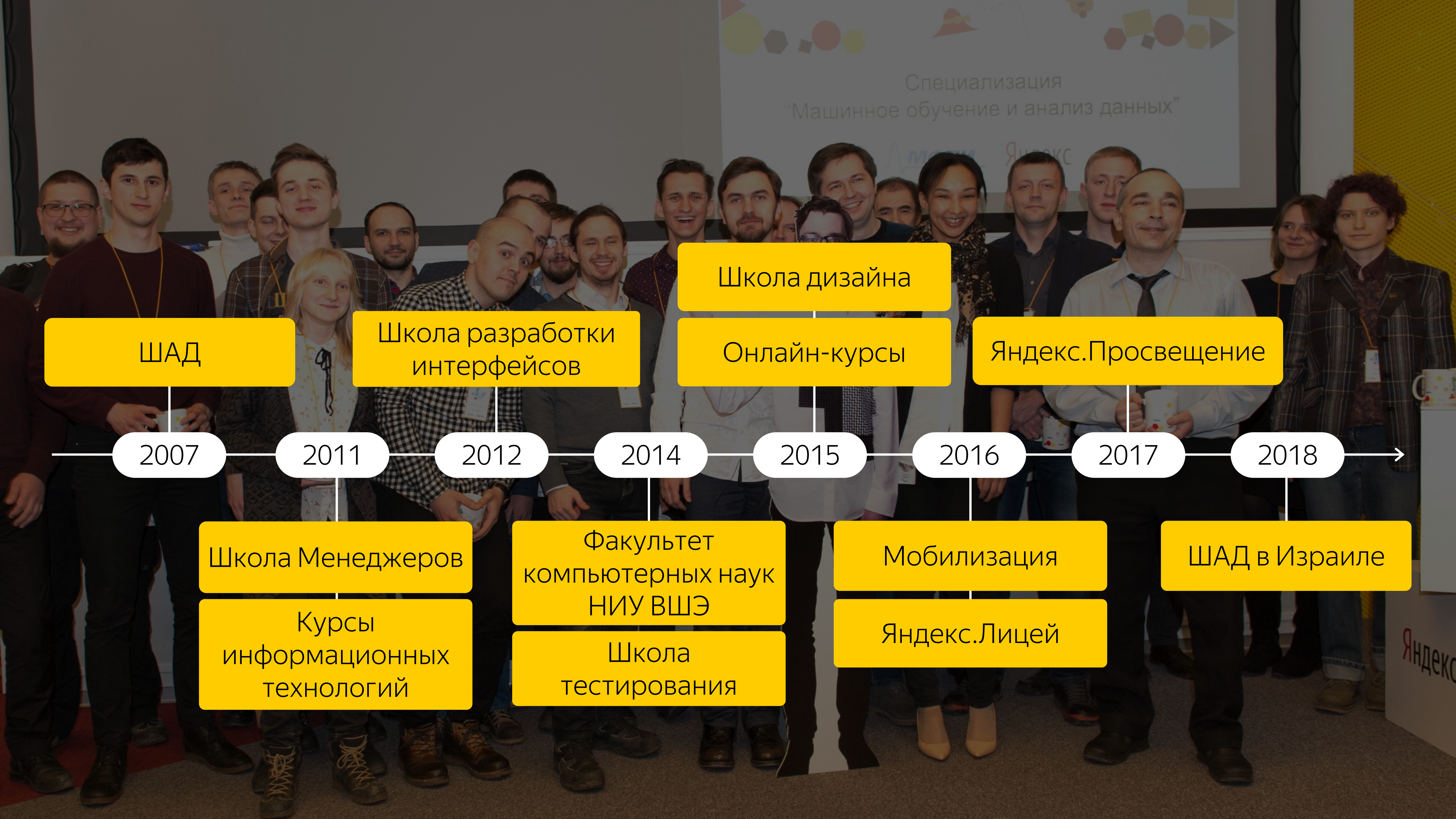


Немного о Яндексе



# Как попасть в Яндекс?

- › Заявка
- › Алгоритмический контеcт
- › Собеседования
- › Подбор команды



Специализация  
"Машинное обучение и анализ данных"

АМБА ЯНДЕКС

ШАД

2007

Школа Менеджеров

Курсы  
информационных  
технологий

Школа разработки  
интерфейсов

2011

2012

Факультет  
компьютерных наук  
НИУ ВШЭ

Школа  
тестирования

2014

Школа дизайна

Онлайн-курсы

2015

Мобилизация

Яндекс.Лицей

2016

Яндекс.Просвещение

2017

ШАД в Израиле

2018

Яндекс

# О Яндексе

## Работа:

- › Встречи с командами и руководителями
- › Стендап команды
- › Командировки
- › Оборудование

# Офис



# О Яндексе

## Отдых:

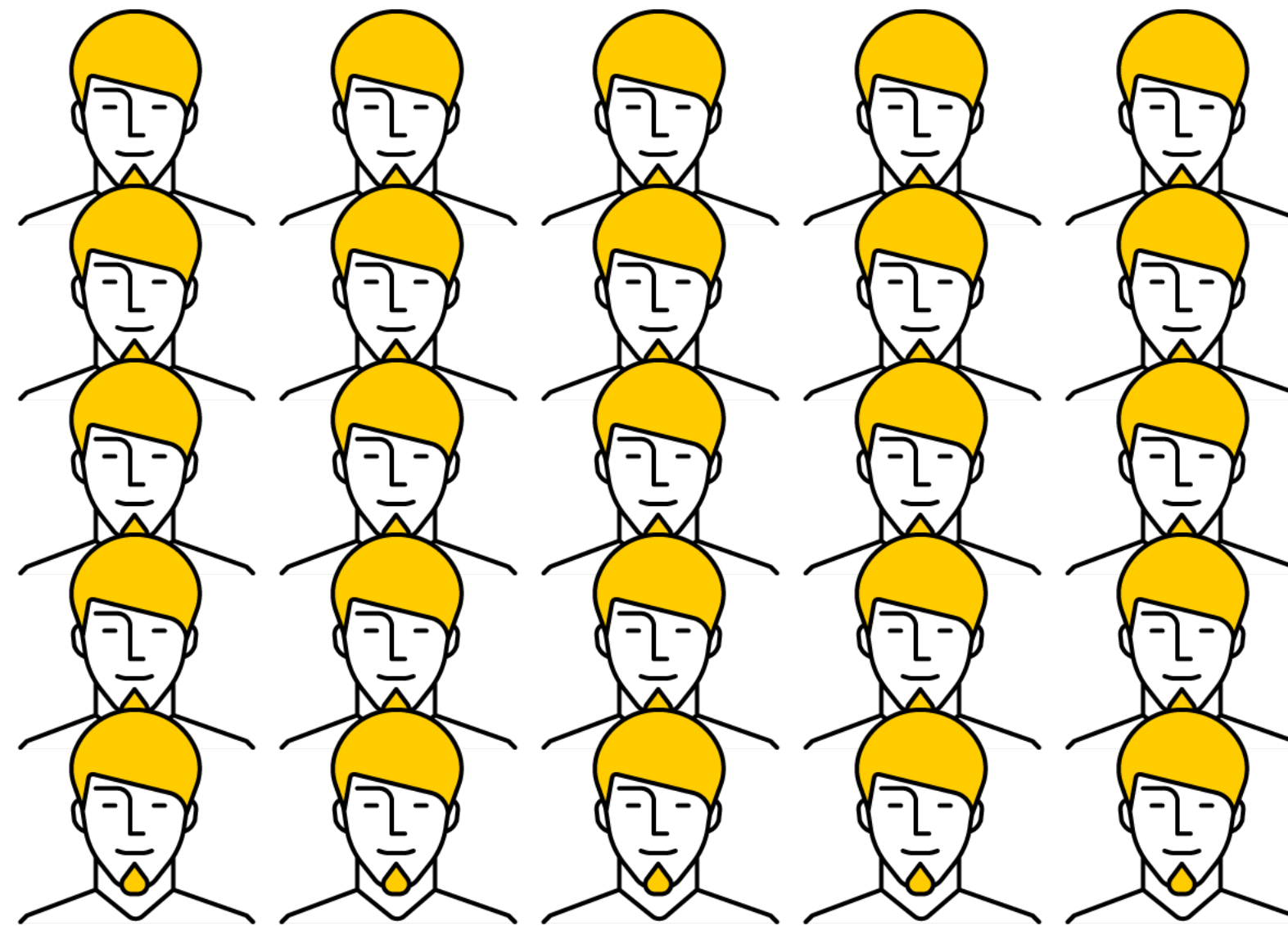
- › Кофепоинты
- › Еда
- › Кикер\теннисный стол
- › Массаж
- › Кружки по интересам



Яндекс.Толока



# Что такое краудсорсинг?



Большая задача

Облако исполнителей

Результат



# Примеры задач




алиса зачем тебя создали

Я стараюсь не задумываться об этом. Вдруг окажется, что два программиста просто понравились друг другу и забавы ради сделали бета-версию.

Ответ показан на экране действие

1  Хороший ответ

2  Плохой ответ

▶ 0:00 / 0:06   

1  Есть речь

2  Нет речи / неразборчиво


Текст аннотации

преобразователь ржавчины три литра петровича ✕

# Примеры задач

Финальный Результат

Горячие клавиши

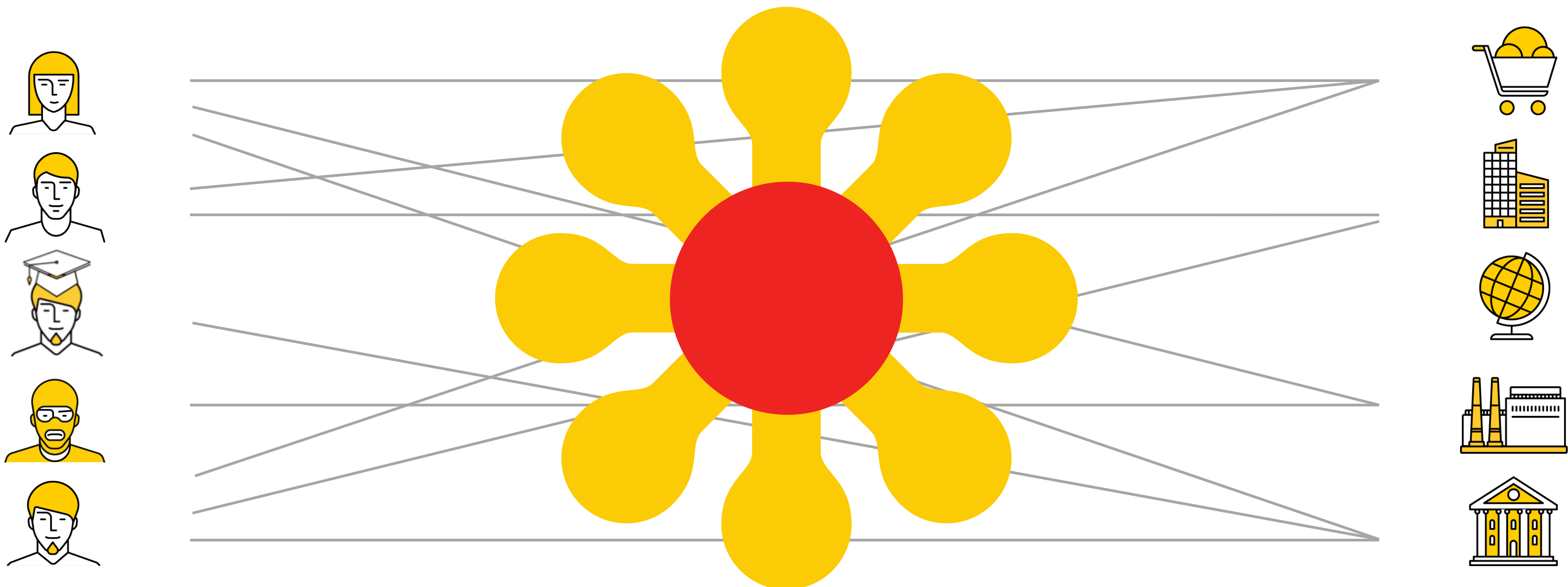


The image displays a computer vision application interface. The main window shows a night-time street scene with various objects highlighted by colored bounding boxes and labels. A person is highlighted in green, and a car is highlighted in orange. The interface includes a toolbar at the top with icons for navigation and a control panel on the right with a list of object classes and their visibility status.

Control Panel Object List:

- другой объект
- другой объект
- другой объект
- светофор
- другой объект
- разметка
- другой объект
- транспорт
- транспорт
- человек
- человек
- светофор
- ✓ человек
- транспорт
- дорога
- бордюр
- тротуар
- разметка
- переход
- другой объект
- Непонятно/грязь
- другой объект
- транспорт
- тротуар
- дорога

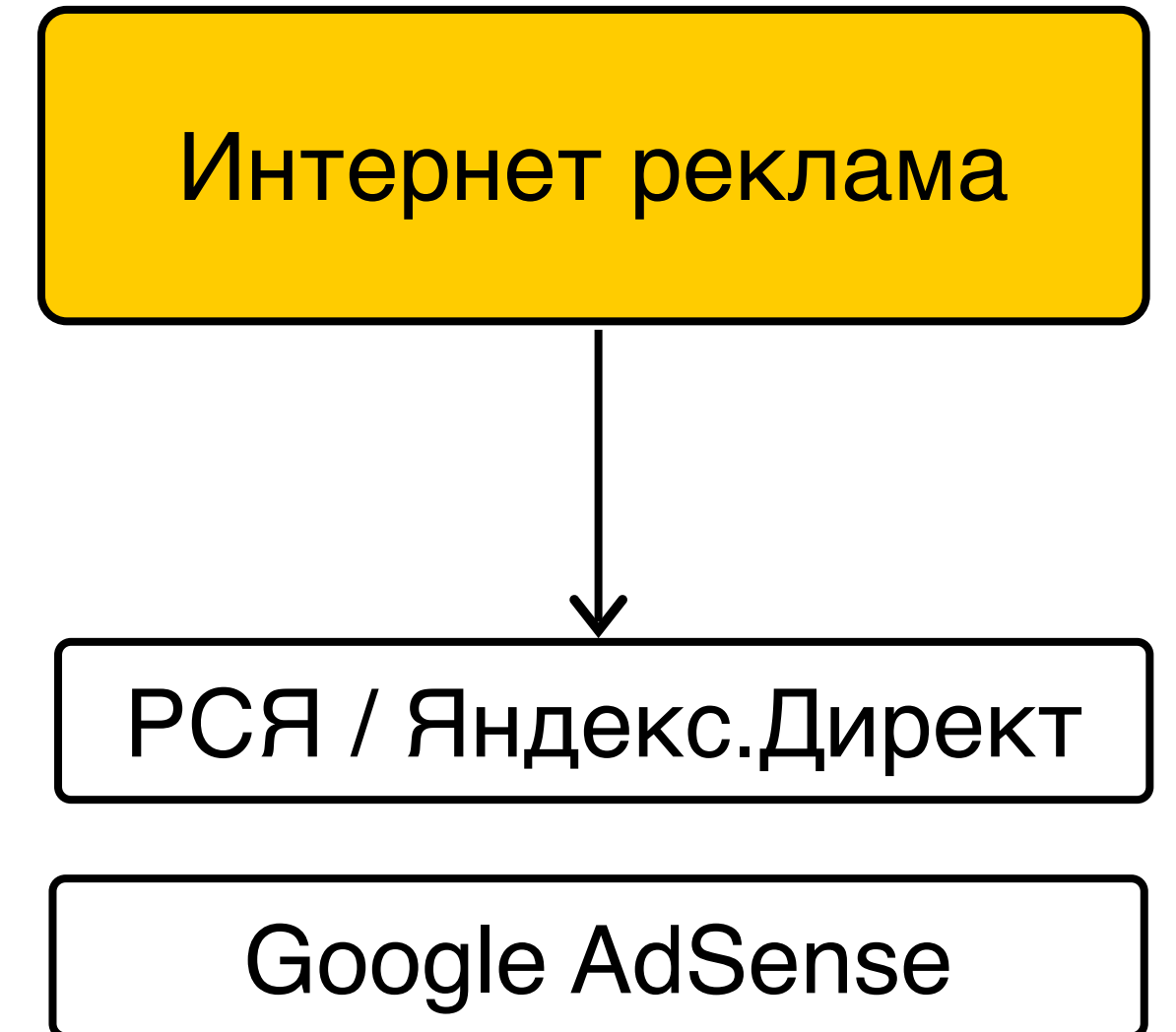
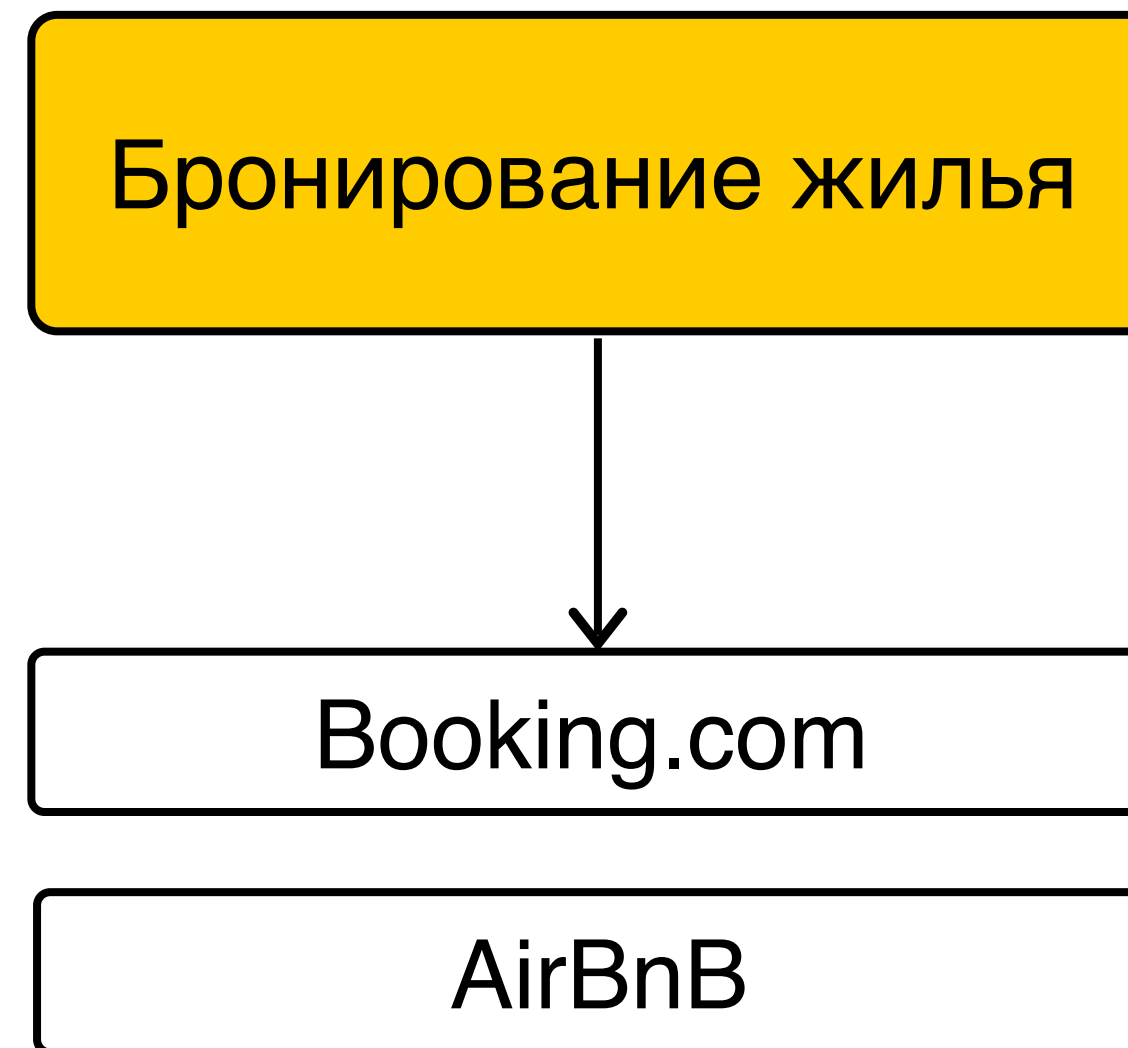
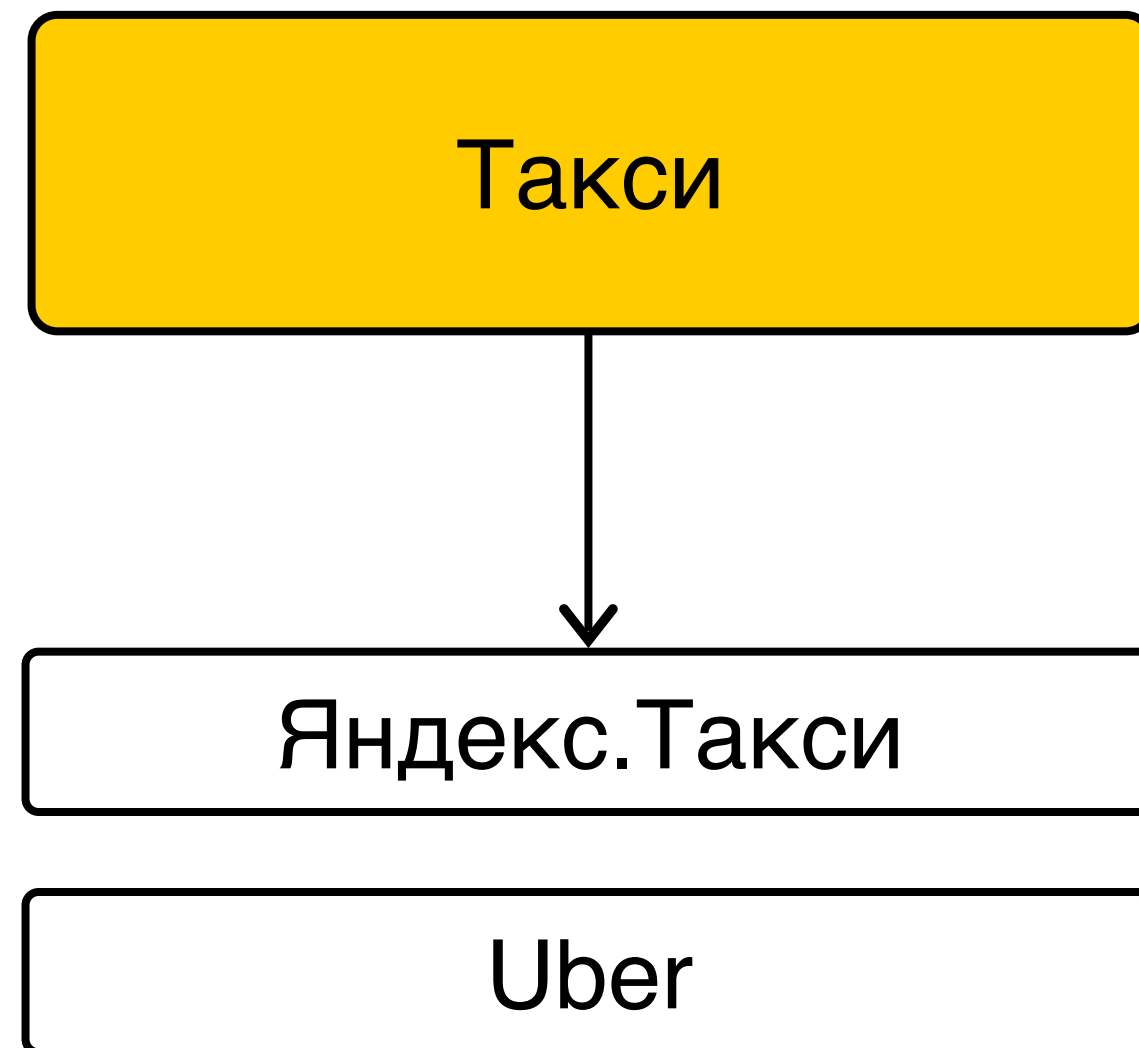
# Платформа «Толока» – среда / экосистема



Потенциальные исполнители  
(рынок исполнителей)

Потенциальные заказчики  
(рынок заданий)

# Аналогии: на что мы похожи



# Задачи

## Исследовательские задачи:

- › Как измерять минимальную ставку в час ?
- › Как измерять ценность задания для заказчика?
- › Как получить качественный результат?

## Продуктовые задачи:

- › Что нужно делать, чтобы эффективность платформы выросла?

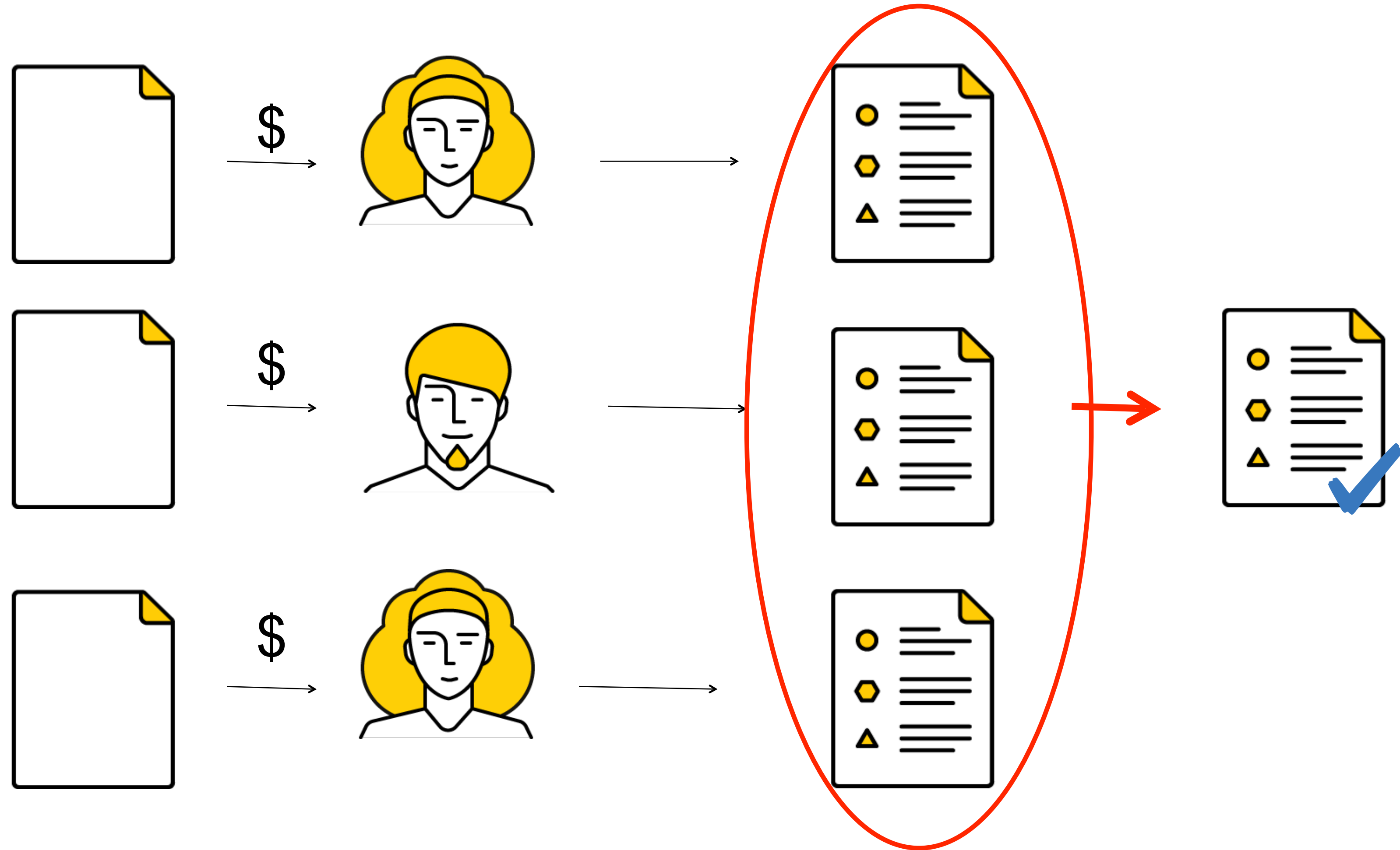
# Пример задачи



# Агрегация ответов

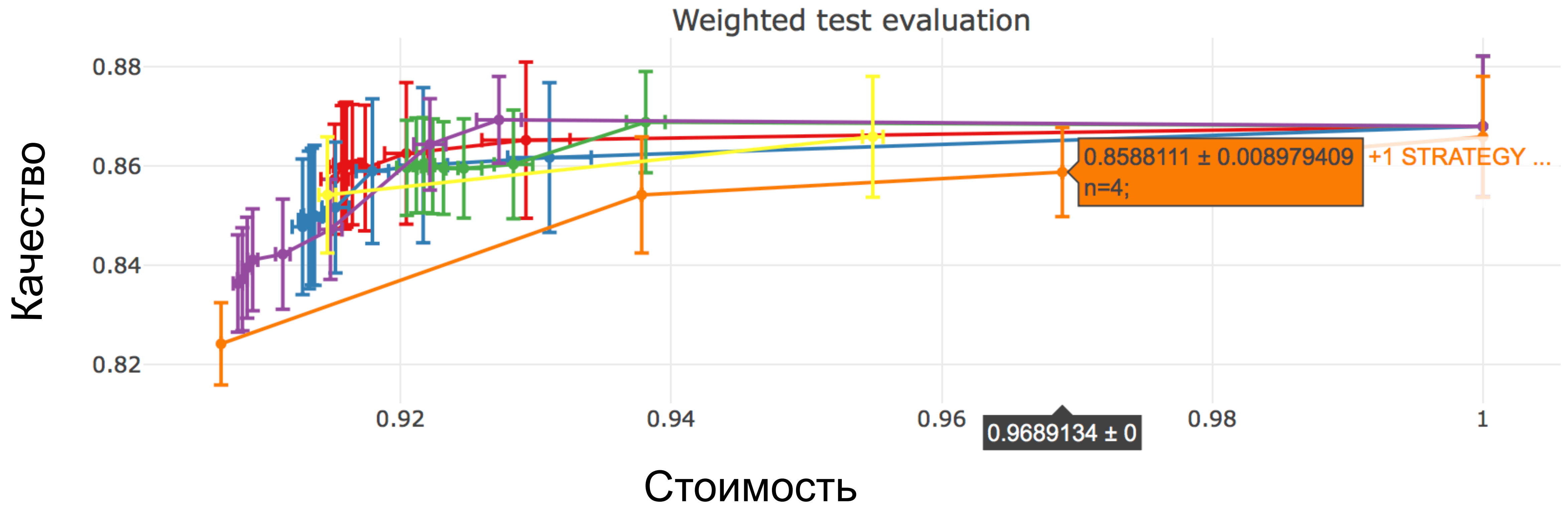
## Толокеры:

- › Не профессионалы
- › Могут ошибаться
- › Боты
- › Злонамеренные



# Моя задача

Баланс между размером перекрытия (ценой) и качеством





# Пример задачи



# Мои задачи

- › Организовать процесс сбора фичей для ML
- › Автоматизировать бан в Толоке

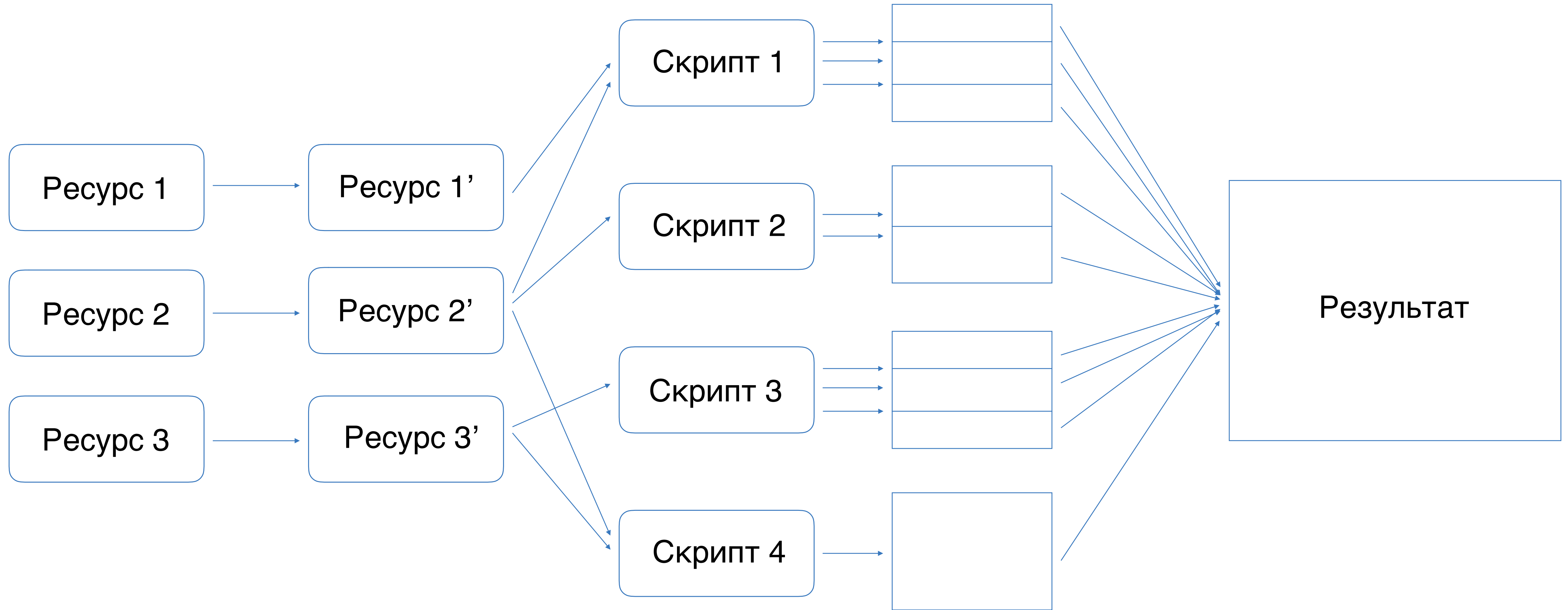
# Сбор фичей

- › Удобство использования
- › Простота понимания
- › Скорость добавления новых фичей
- › Масштабируемость
- › Переносимость на другие задачи

# Внутренние инструменты

- › YT - Yandex Table
- › YQL - Yandex Query Language
- › Nirvana

# Схема сбора



Подготовка

Подсчет фичей

Объединение

# Полученное решение

## Плюсы:

- › Оно работает
- › Масштабируемость
- › Переносимость на другие задачи

## Минусы:

- › Добавить фичу можно быстро, а посчитать за всю историю не всегда
- › Нужно один раз разобраться, как это работает

# Автобан

- › Таргет
- › Сбор датасета
- › CatBoost

# Что получилось

- › Хорошая точность
- › Приемлемая полнота
- › непонимание саппортов



# Вопросы

Надежда Бугакова  
Стажер-аналитик

 [belatriss@yandex-team.ru](mailto:belatriss@yandex-team.ru)

Александр Хворов  
Стажер-аналитик

 [akhvorov@yandex-team.ru](mailto:akhvorov@yandex-team.ru)