



Toloka

# Improving Web Ranking with Human-in-the-Loop: Methodology, Scalability, Evaluation

Alexey Drutsa, Dmitry Ustalov, Nikita Popov,  
Daria Baidakova



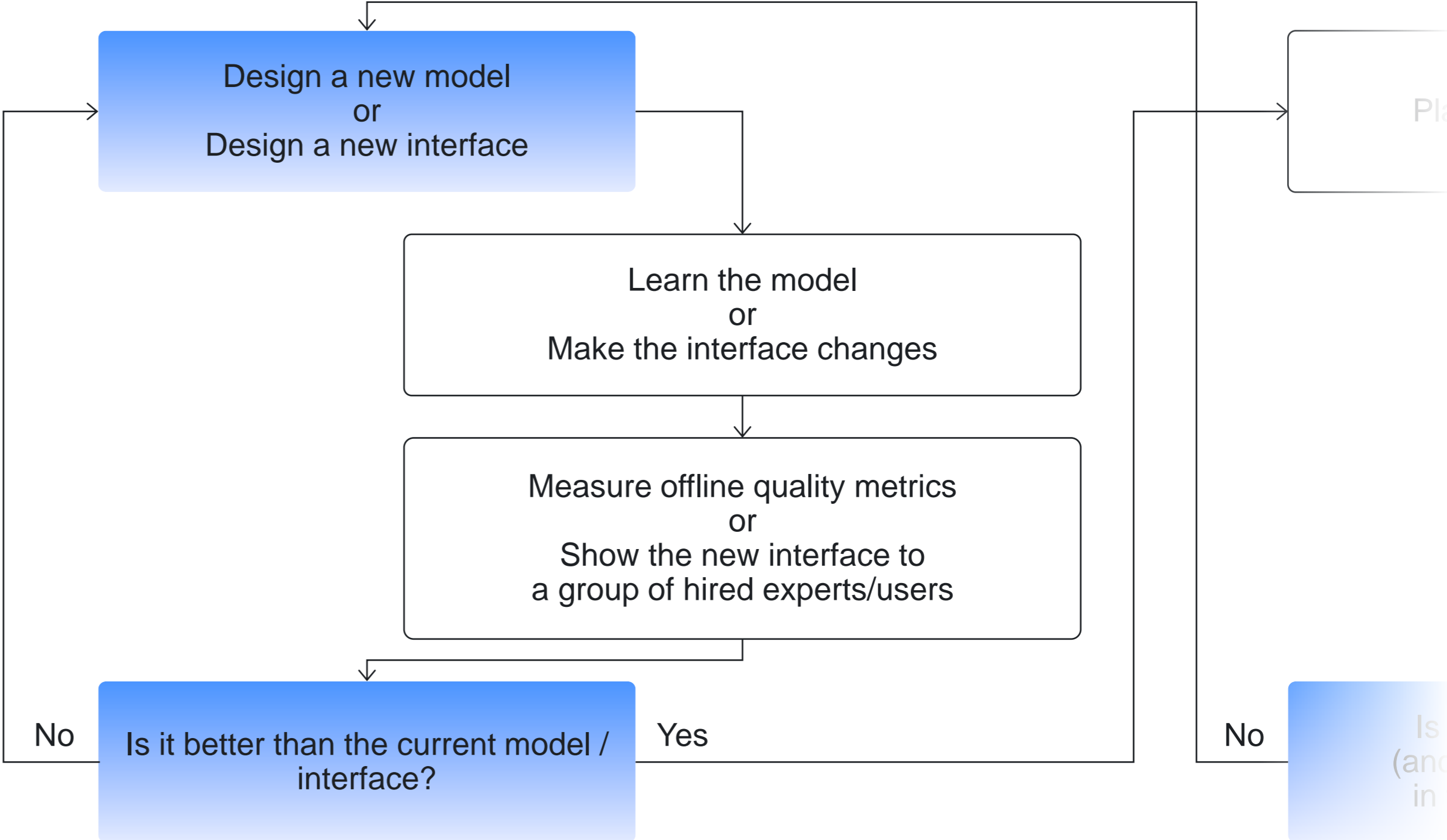
# Introduction

Alexey Drutsa,  
Head of Efficiency and Growth Division at Toloka

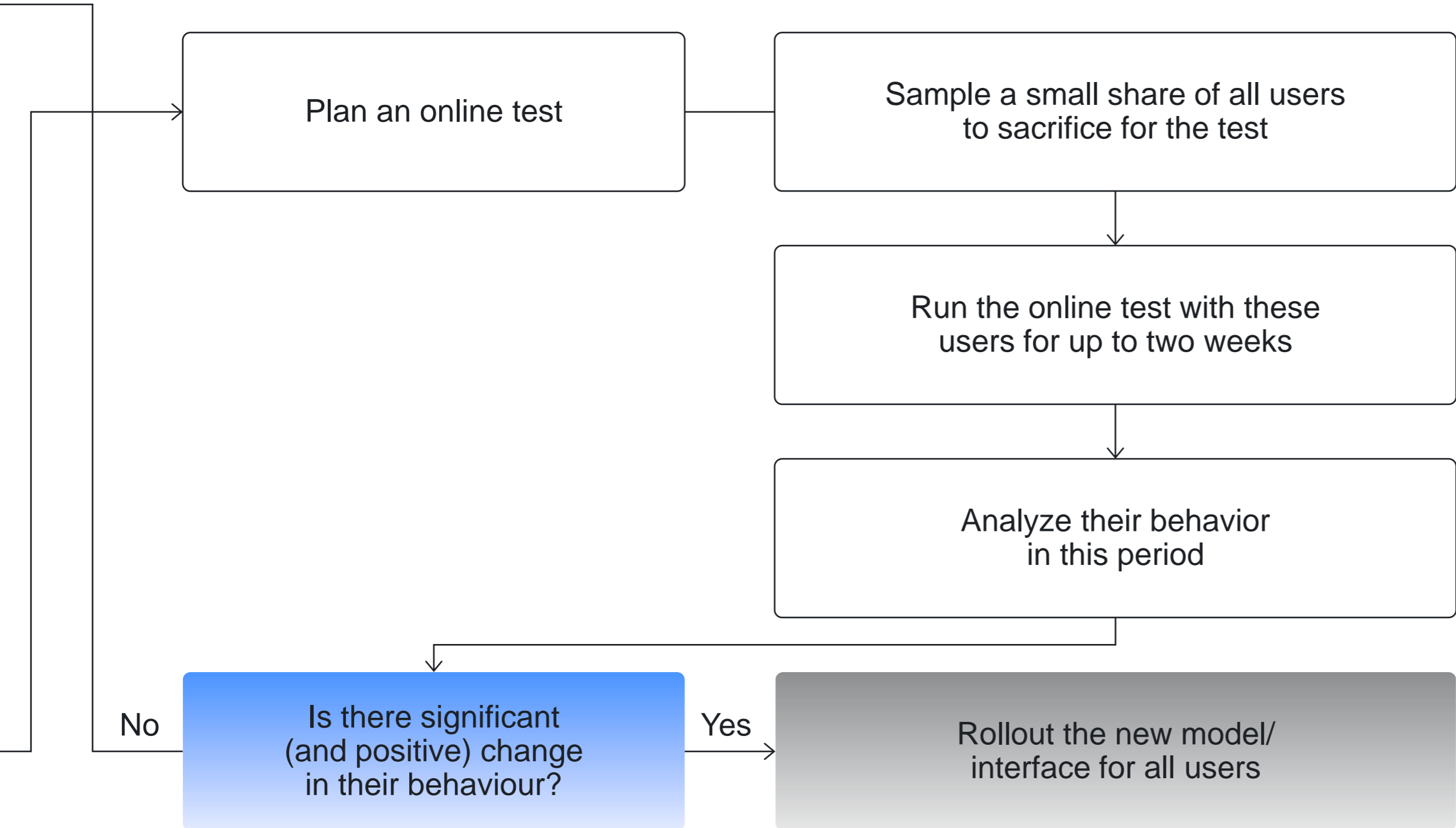


# Ranking evaluation

# Scheme of web service evaluation



# Scheme of web service evaluation



# Evaluation

Offline eval

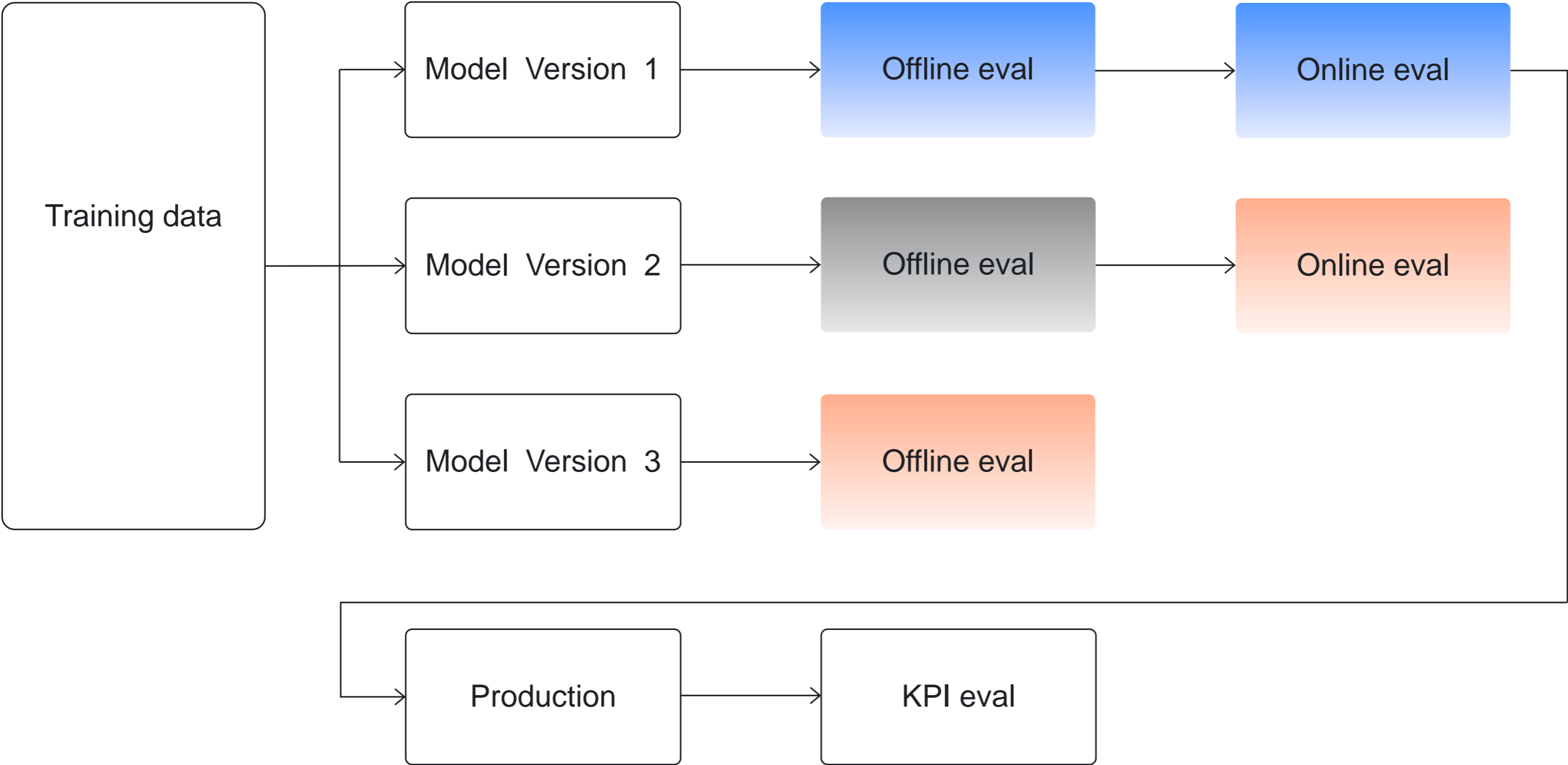
Online eval

The topic of  
this tutorial

**E.g., see our tutorials**

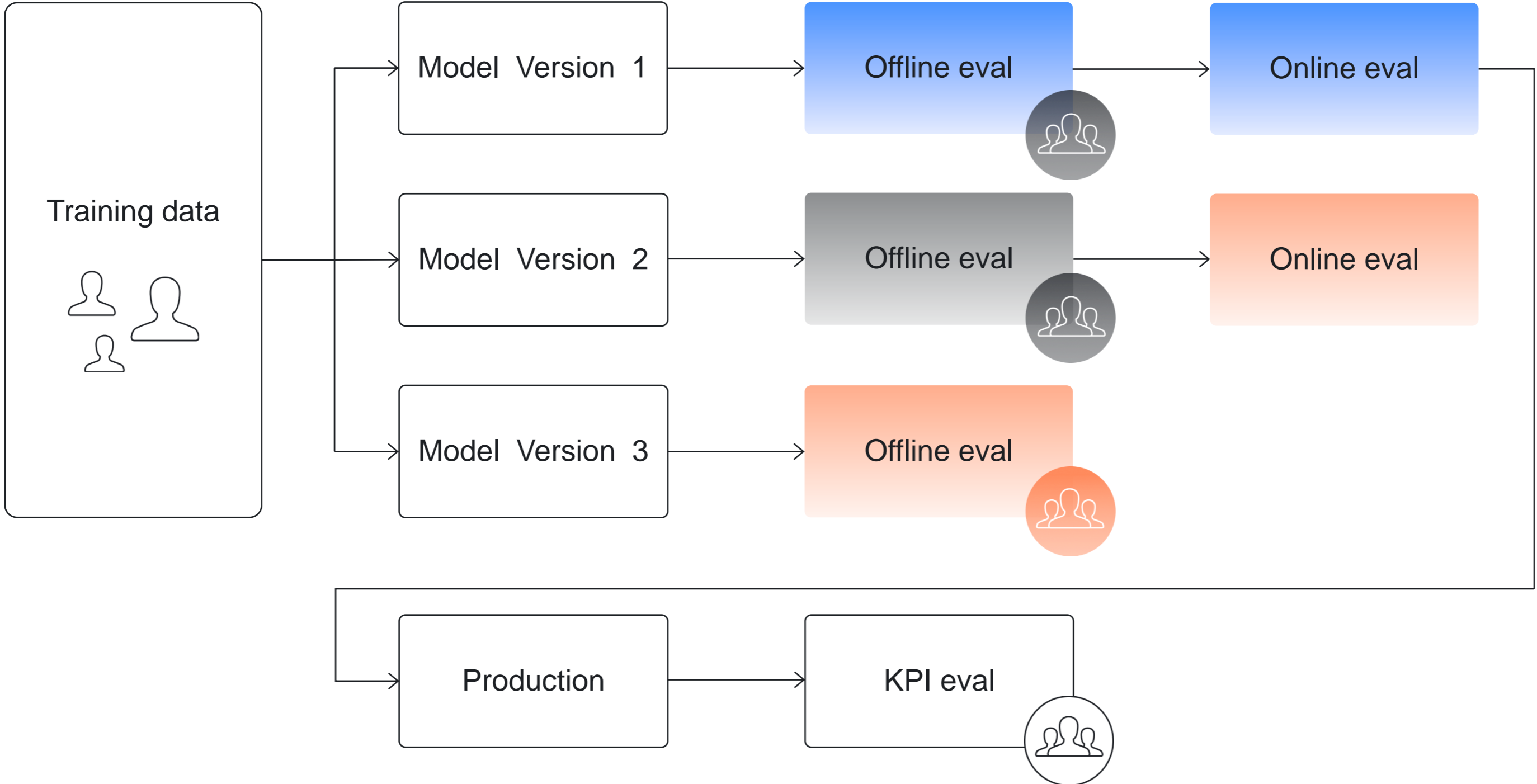
- ▶ At TheWebConf 2018
- ▶ At KDD 2018
- ▶ At SIGIR 2019

# ML production pipeline (not only ranking)





# ML production pipeline: Human-in-the-Loop



**Crowdsourcing as a  
powerful technology for  
data-based industries**

Search

Machine translation

Personal assistant

Self-Driving

Maps

Ads

E-commerce

Speech technologies

# Majority of data-based solutions require data often labelled by human



...at a large scale





# XX century — style management

- ▶ Routine tasks
- ▶ Regular work
- ▶ No ability to choose tasks



# It can be different

- ▶ Flexibility to choose from hundreds of tasks
- ▶ No requirements in regularity
- ▶ Switch to another task when bored



# Crowdsourcing: specific way to design a business process

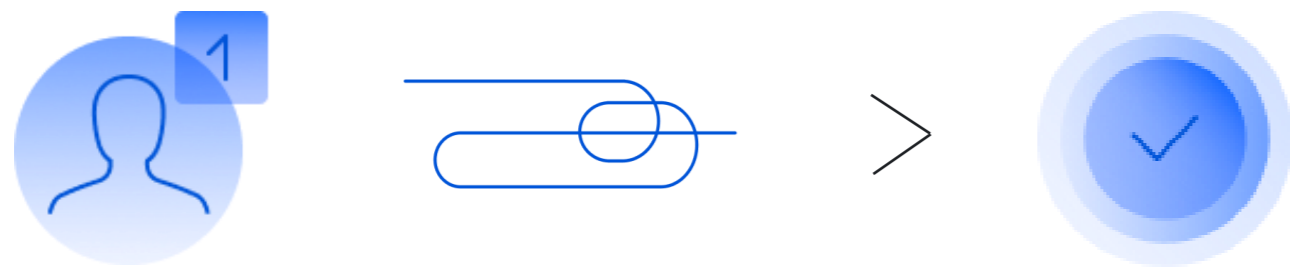


A big task

Cloud of performers

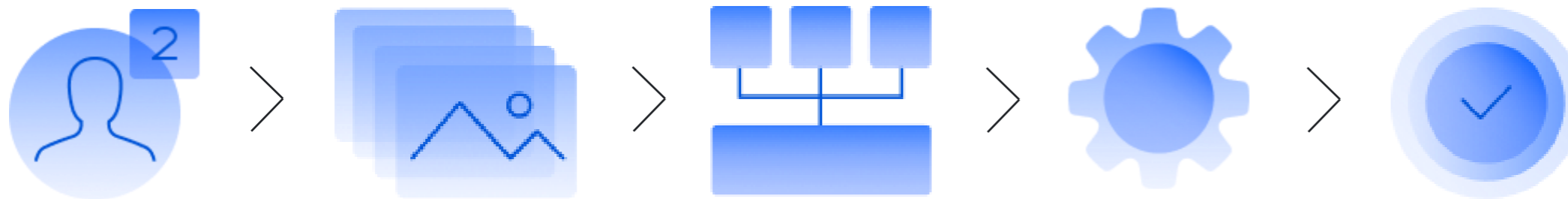
Result

# Crowdsourcing: require less from a performer, more — from a manager



**Expert approach: rely on an expertise of a particular performer**

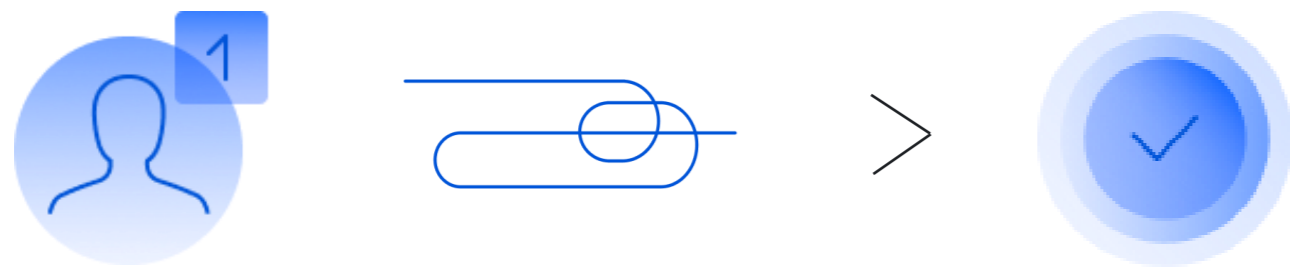
- ▶ Expensive
- ▶ **Unmeasurable**
- ▶ Hard to scale



**Crowdsourcing approach**

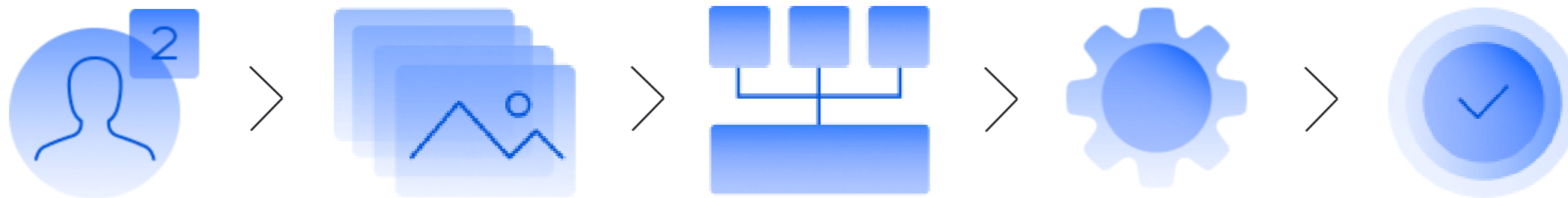
- ▶ Measurable
- ▶ Scalable
- ▶ Manageable

# Crowdsourcing: require less from a performer, more — from an **engineer-manager**



**Expert approach: rely on an expertise of a particular performer**

- ▶ Expensive
- ▶ Unmeasurable
- ▶ Hard to scale



**Crowdsourcing approach**

- ▶ Measurable
- ▶ Scalable
- ▶ Manageable



# Crowdsourcing can provide maximal flexibility to performers if

- ▶ On a platform side, efficient tools for quality management are available for requester
- ▶ Requester knows how to build smart crowdsourcing pipelines resistant to single performer's mistakes

# Core take away from the tutorial

Data labeling is an essential part of ML production

1. Data labeling is a room for innovations to beat competitors  
In real world, you compete within the whole pipeline of ML production  
Those who know how to manage the data labelling win
2. Do not outsource data labeling expertise  
if you do not outsource expertise in ML
3. Data labelling on a large scale is an engineering task  
Deal with the Crowd as with yet another computing cluster

# Crowdsourcing: examples

# Crowdsourcing applications: examples

Task type	Used in
Information assessment	Ranking of search results
Content categorization	Text and media moderation, data cleaning and filtering
Content annotation	Metadata tagging
Pairwise comparison	Offline evaluation, media duplication check
Object segmentation, including 3D	Image recognition for self-driving car
Audio and video transcription	Speech recognition for voice-controlled virtual assistant
Spatial crowdsourcing	Verify business information and office hours



# Example: binary classification

Is this cat white?

Yes

No





# Example: multi classification



"Real French restaurant"



If you are a gourmand, I can recommend you the "Real French restaurant", located in the historic cellar, with elements of antique design and quite interesting cuisine. The restaurant is small, but very cozy and romantic. The restaurant is very suitable for romance and even for business meetings.

**Is it a feedback?**

Yes, it is  No, it's other comment

Personal information

Swearing, vulgarity, insults, aggressive statements

Spam, advertisingspan

# Example: multi classification with ordered labels

Query: Machine learning  
URL: https://en.wikipedia.org/wiki/Machine\_learning

Open the original [Yandex](#) [Google](#)

1  Vital  
2  Useful  
3  Relevant+  
4  Relevant-  
5  Irrelevant  
6  Not displayed

en.wikipedia.org Machine learning - Wikipedia

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article [Talk](#) [Read](#) [Edit](#) [View history](#)

## Machine learning

From Wikipedia, the free encyclopedia

*For the journal, see [Machine Learning \(journal\)](#).  
"Statistical learning" redirects here. For statistical learning in linguistics, see [statistical learning in language acquisition](#).*

**Machine learning (ML)** is the scientific study of algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on



# Examples: pairwise comparison

The image displays two mobile recipe cards side-by-side, followed by a pairwise comparison interface.

**Left Card (Food Network Magazine):**

- Header: **How to Make Perfect Pancakes**
- Text: *Food Network Magazine shows you how to make the best short stack, plus some tasty toppings.*
- Disclaimer: Keep in mind: Price and stock could change after publish date, and we may make money from these links.
- Date: April 24, 2015
- Source: From: **Food Network Magazine**
- Image: A stack of pancakes with butter and syrup.

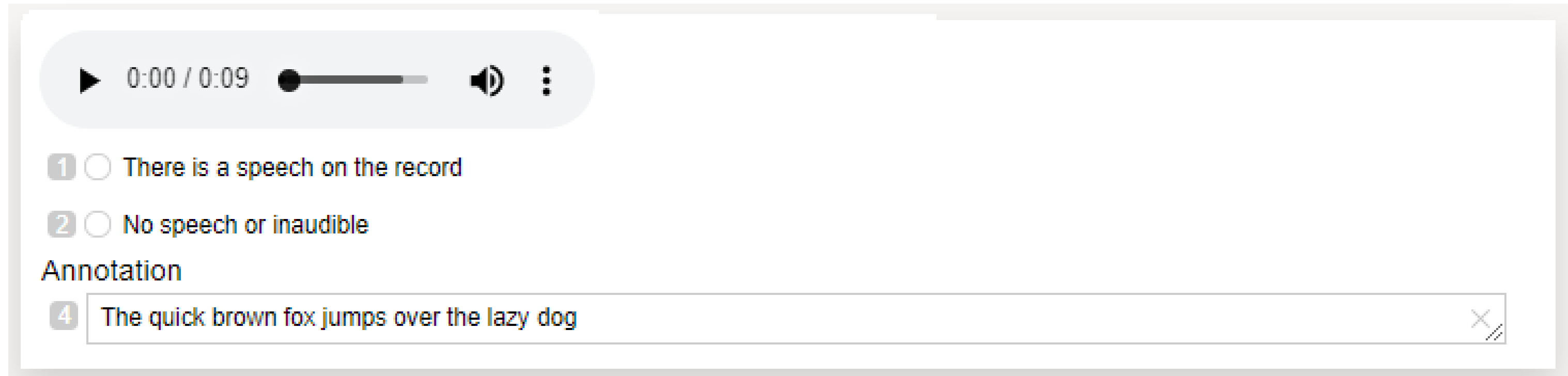
**Right Card (Search for more recipes):**

- Header: **How to make pancakes**
- Ratings: ★ ★ ★ ★ ☆ 17 ratings
- Image: A stack of pancakes on a plate.
- Preparation time: **less than 30 mins**
- Cooking time: **less than 10 mins**
- Serves: **Serves 4**
- Dietary: **V**

**Comparison Interface:**

- Query: how to make pancakes
- Question: Which one do you like better?
- Options:  Left  Right
- Text: Please, comment your choice
- Form: A text input field for comments.
- Button: **Continue**

# Examples: transcription with textual answers



0:00 / 0:09

1  There is a speech on the record

2  No speech or inaudible

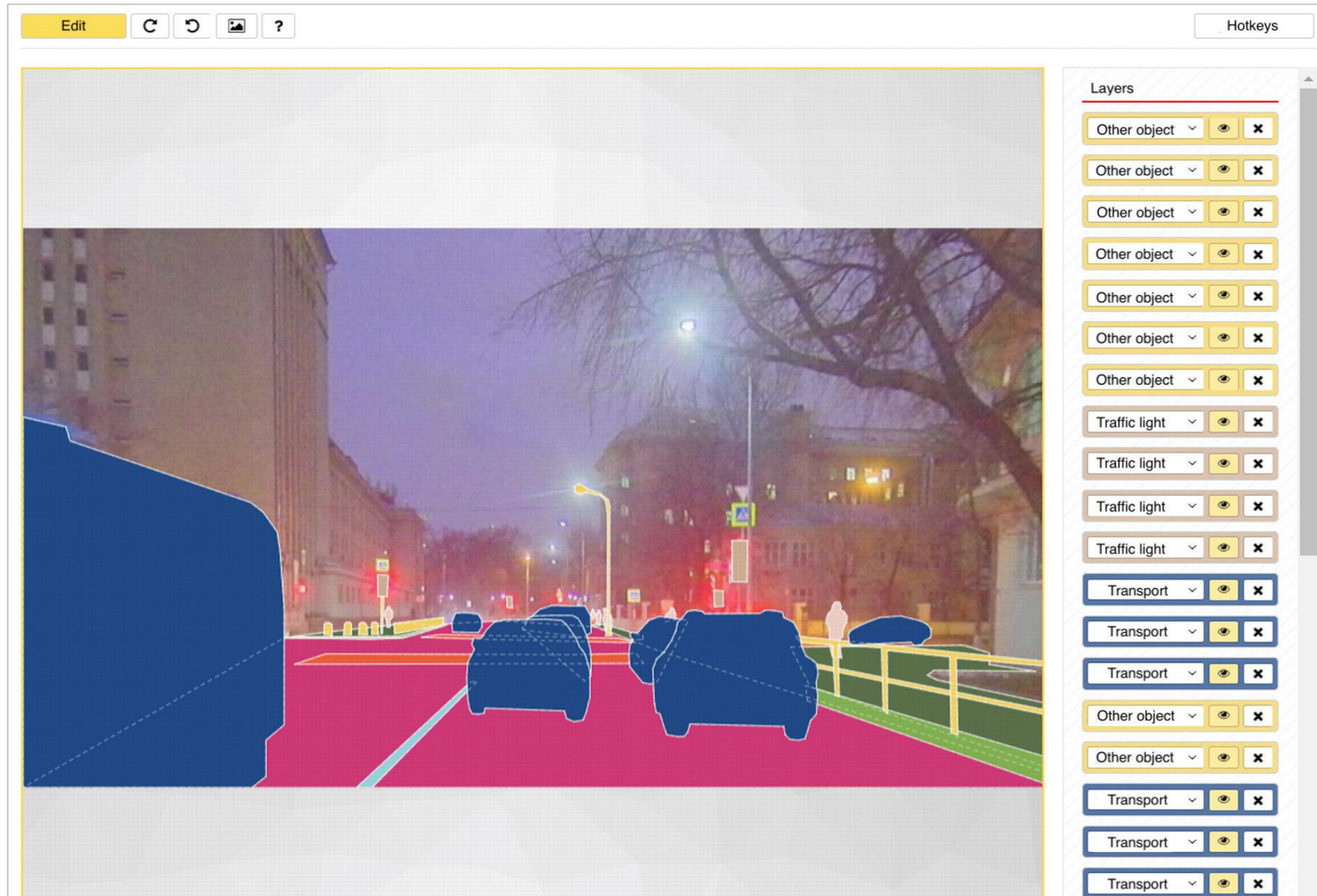
Annotation

4 The quick brown fox jumps over the lazy dog

The image shows a user interface for an audio player. At the top, there is a play button, a progress bar showing 0:00 / 0:09, a volume icon, and a menu icon. Below the player are two radio button options: '1 There is a speech on the record' and '2 No speech or inaudible'. Underneath these is a section labeled 'Annotation' containing a text input field with the text '4 The quick brown fox jumps over the lazy dog' and a close button (an 'X' icon).

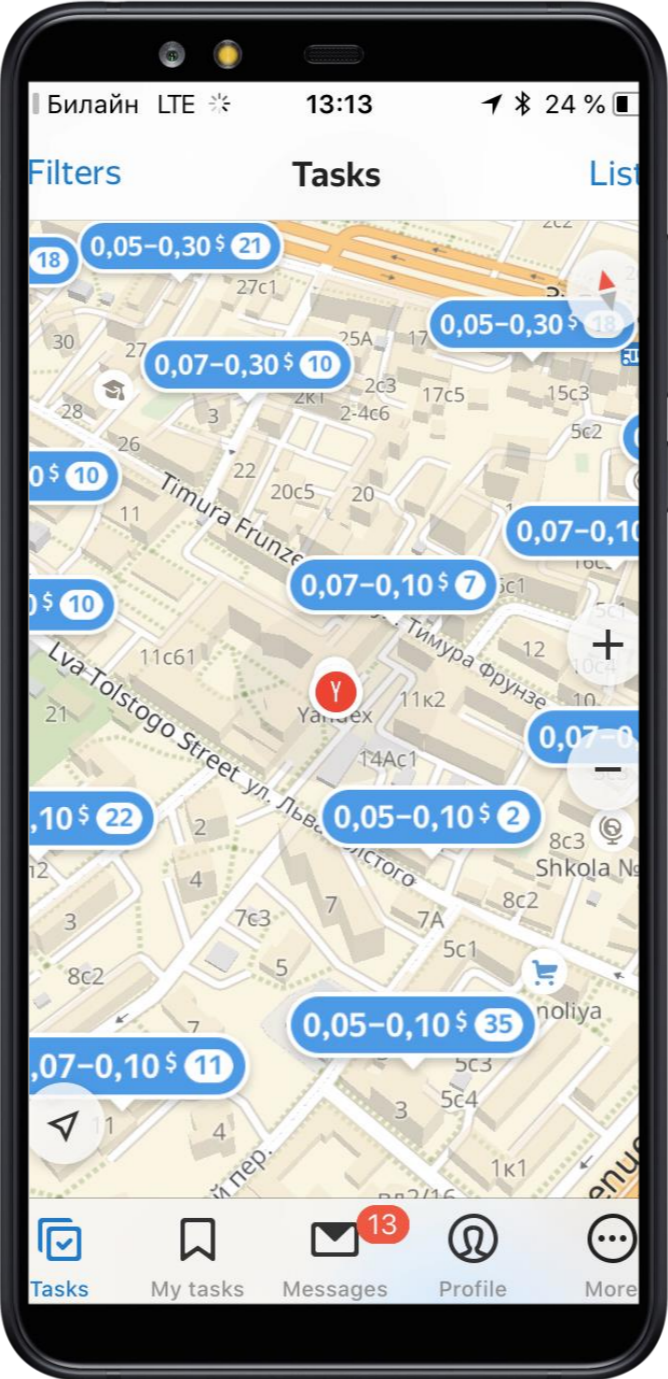


# Examples: object segmentation

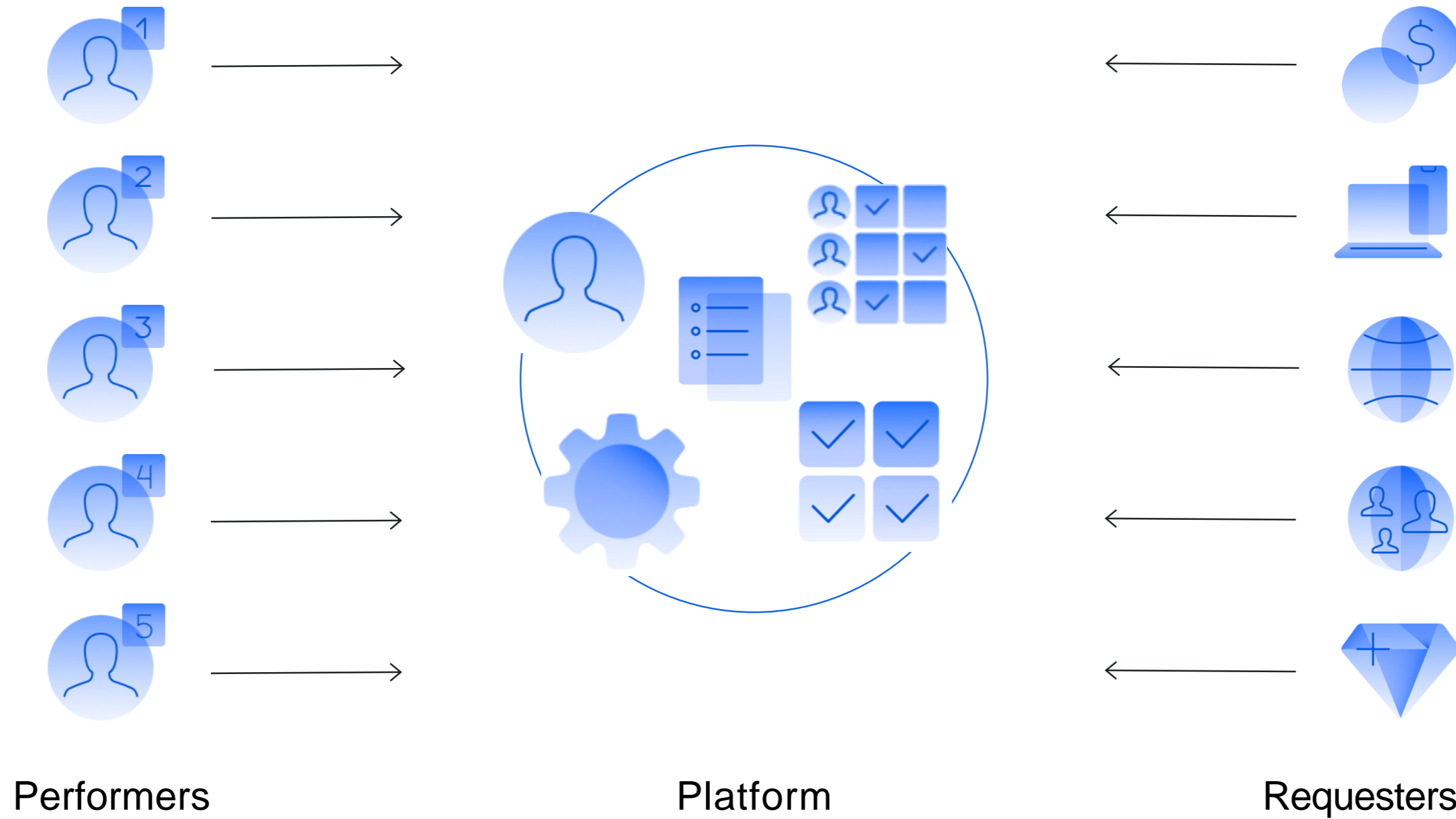




# Examples: spatial crowdsourcing



# A crowdsourcing platform: two-sided market



# Crowdsourcing platforms: examples

- ▶ Amazon Mechanical Turk
- ▶ Toloka
- ▶ Microworkers
- ▶ Gigwalk
- ▶ ClickWorker
- ▶ CloudFactory
- ▶ CrowdSource
- ▶ DefinedCrowd
- ▶ ...

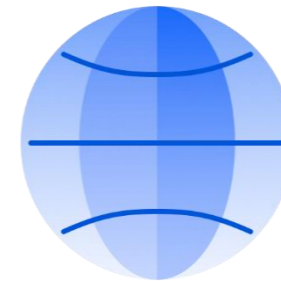
# Pros of crowdsourcing platforms



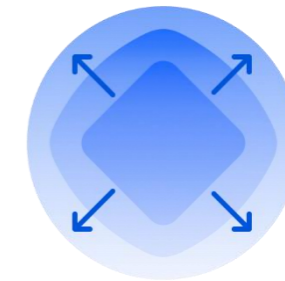
24/7



Variety  
of skilled  
performers



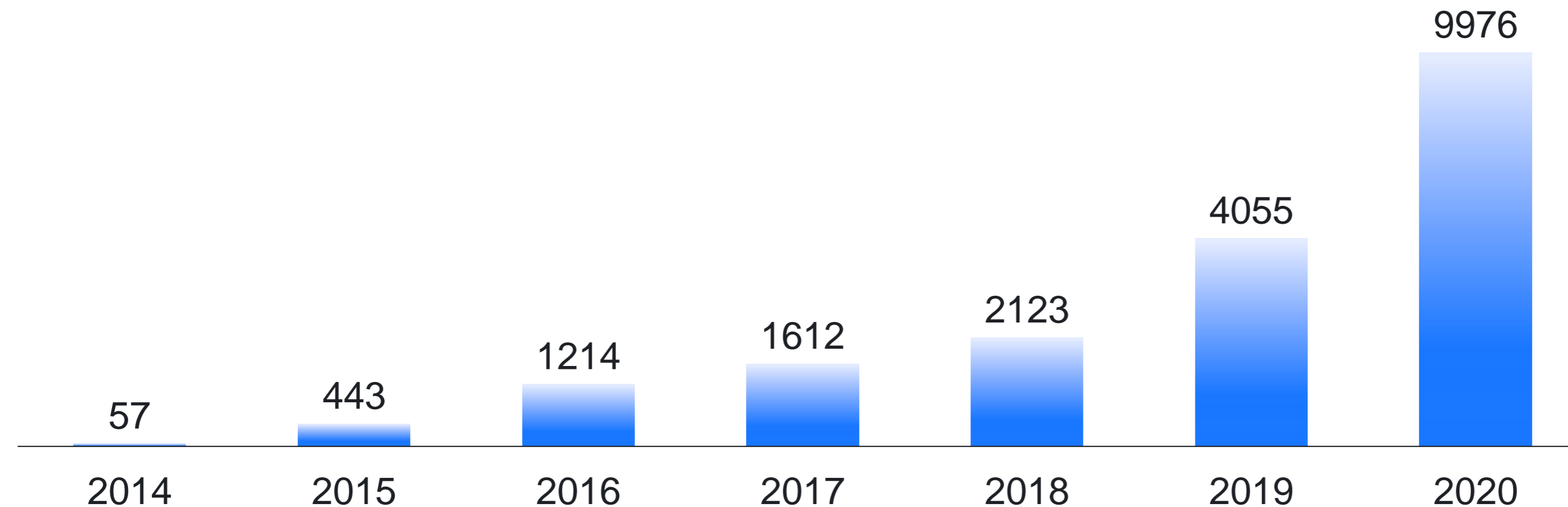
Vast  
region  
coverage



Ongoing  
processes

# Crowdsourcing growth: Yandex experience

Different projects in Toloka

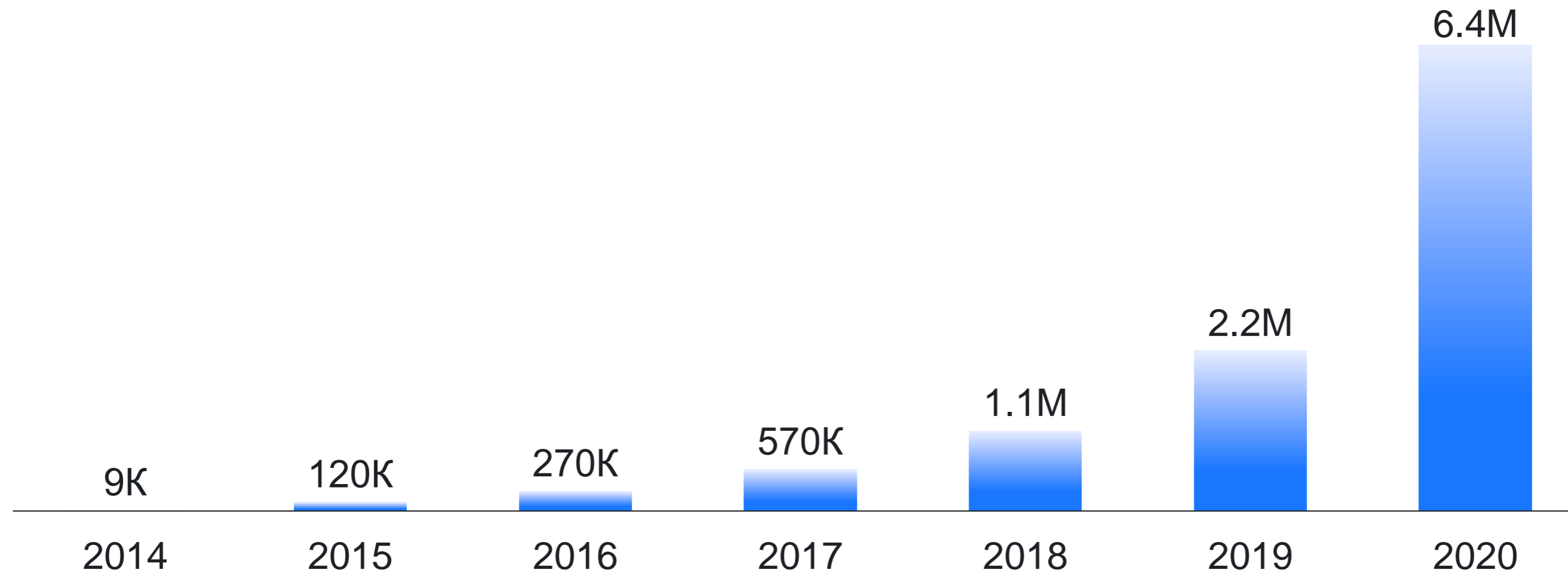


\* An extrapolation based on the first 3 months of 2020



# Crowdsourcing growth: Yandex experience

Active performers in Toloka



\* An extrapolation based on the first 3 months of 2020

# Everyday on Toloka



700+  
different  
projects



41K+  
performers



17M+  
tasks

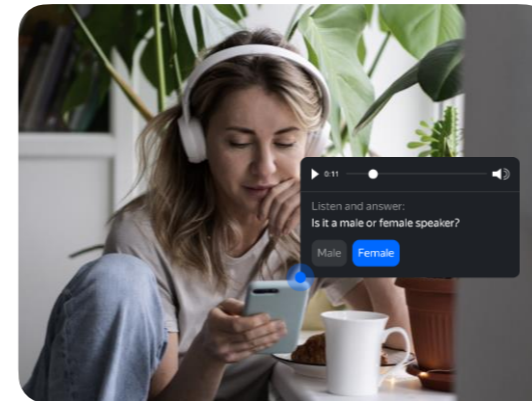
# Toloka: real-life cases



Object classification

# 1000 photos

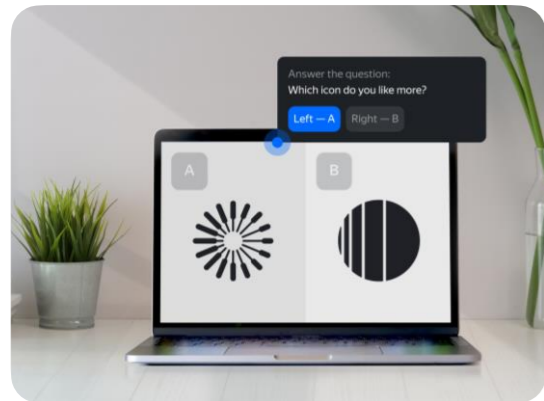
🕒 15 min 💰 \$1.2



Audio transcription

# 100 recordings  
25 minute long

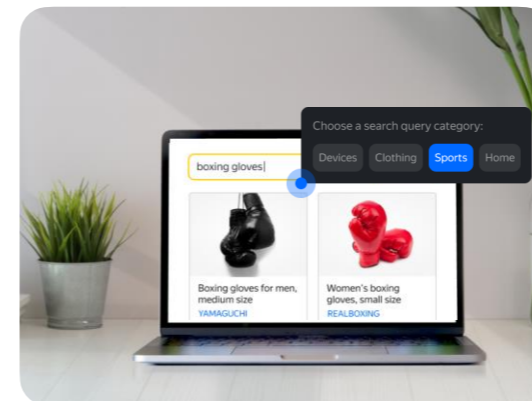
🕒 20 min 💰 \$6



Side-by-side comparison

# 1000 photos

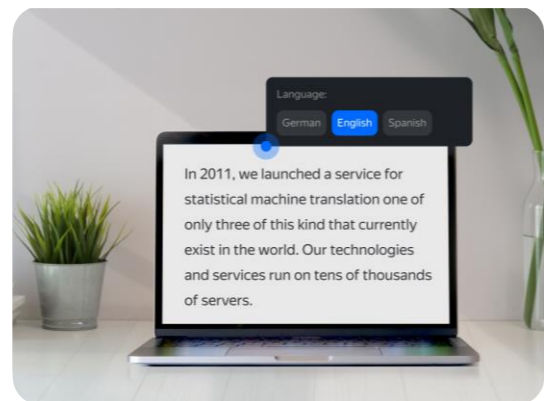
🕒 10 min 💰 \$2.4



Object segmentation

# 1 000 objects in  
100 photos

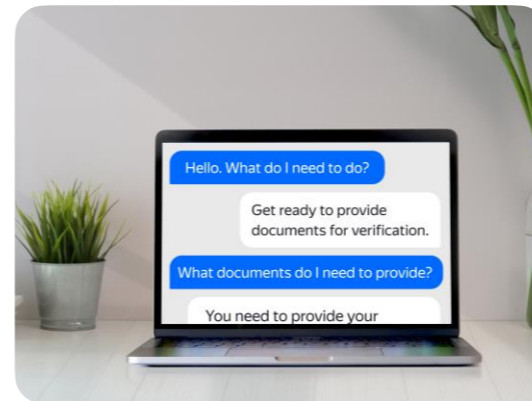
🕒 5 hrs 💰 \$3.6



Text Classification

# 1000 tasks

🕒 2 hrs 💰 \$18



Phrase generation for a chatbot

# 500 phrases on the same topic

🕒 15 min 💰 \$1

# Tutorial overview



**Why this tutorial?**  
**Practice!**



# You will learn how to

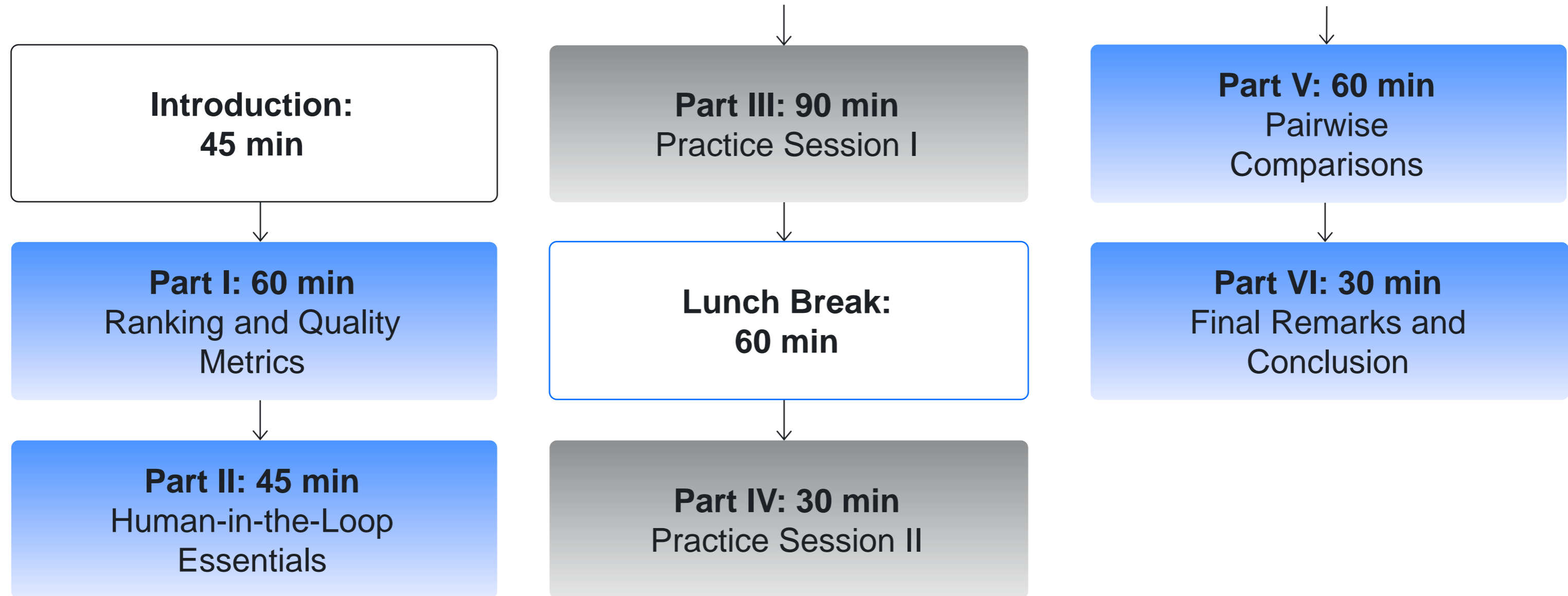
## Theory

- ▶ Evaluate ranking in an offline manner
- ▶ Use crowdsourcing in an efficient way for industrial scale

## Practice

- ▶ Build scalable data labeling pipelines
- ▶ Run crowdsourcing projects on real performers
- ▶ Programmatically build Human-in-the-Loop via public Python libraries

# Tutorial Schedule



# Toloka Research Grants Program

- ▶ We encourage the use of crowdsourcing for research purposes
- ▶ Recipients of the grant are awarded up to \$500 in credit to fuel their research



<https://toloka.ai/grants/>

# Thank you! Questions?

**Alexey Drutsa**

Head of Efficiency and Growth Division



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



<https://research.yandex.com/tutorials/crowd/www-2021>



Grants for research  
based on crowdsourcing