
Toloka

Toloka

WWW 2021 hands-on tutorial

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

Alexey Drutsa, Dmitry Ustalov, Nikita Popov,
Daria Baidakova

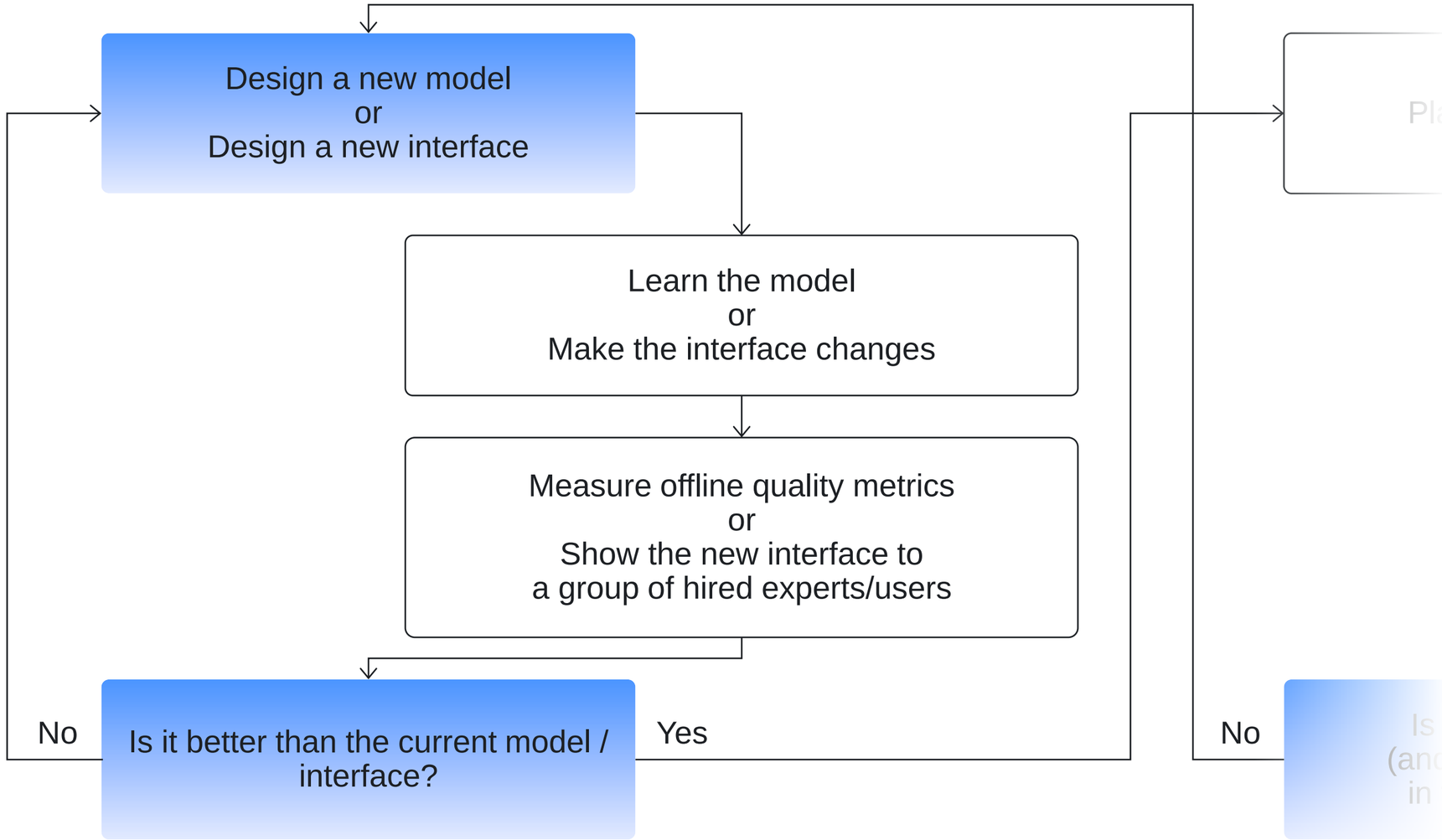
Toloka

Introduction

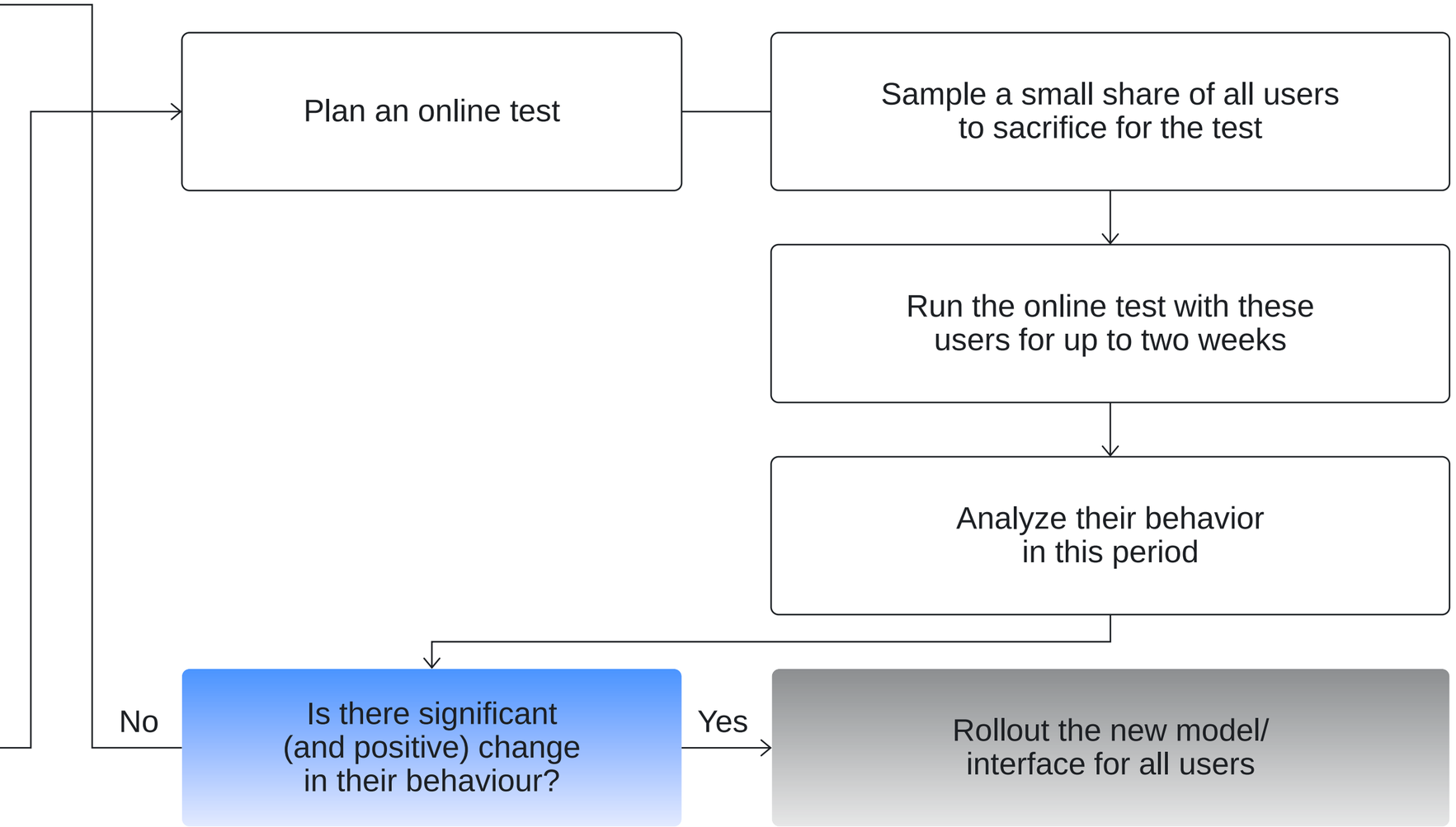
Alexey Drutsa,
Head of Efficiency and Growth Division at Toloka

Ranking evaluation

Scheme of web service evaluation



Scheme of web service evaluation



Evaluation

Offline eval

Online eval

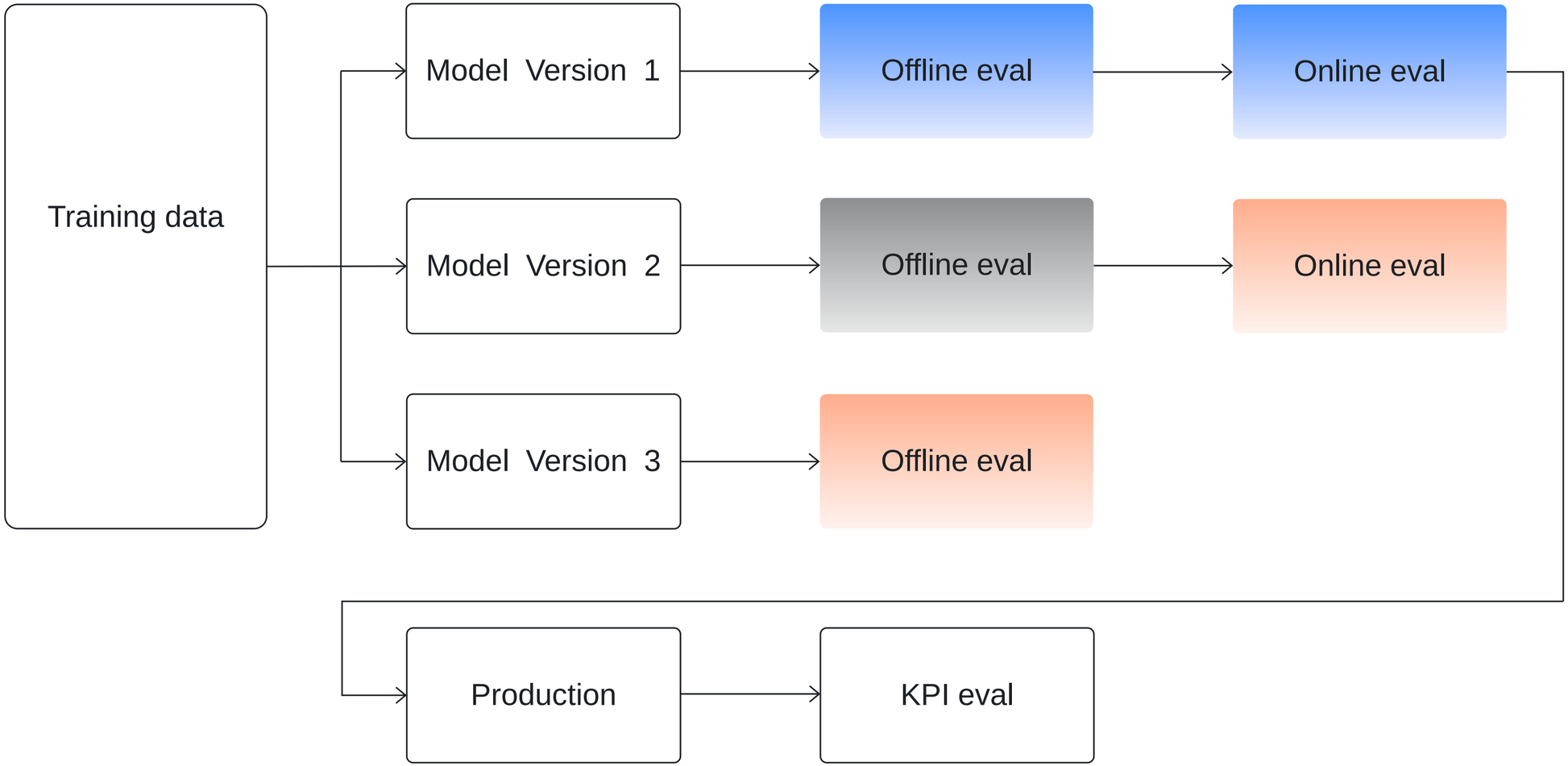
The topic of this tutorial

E.g., see our tutorials

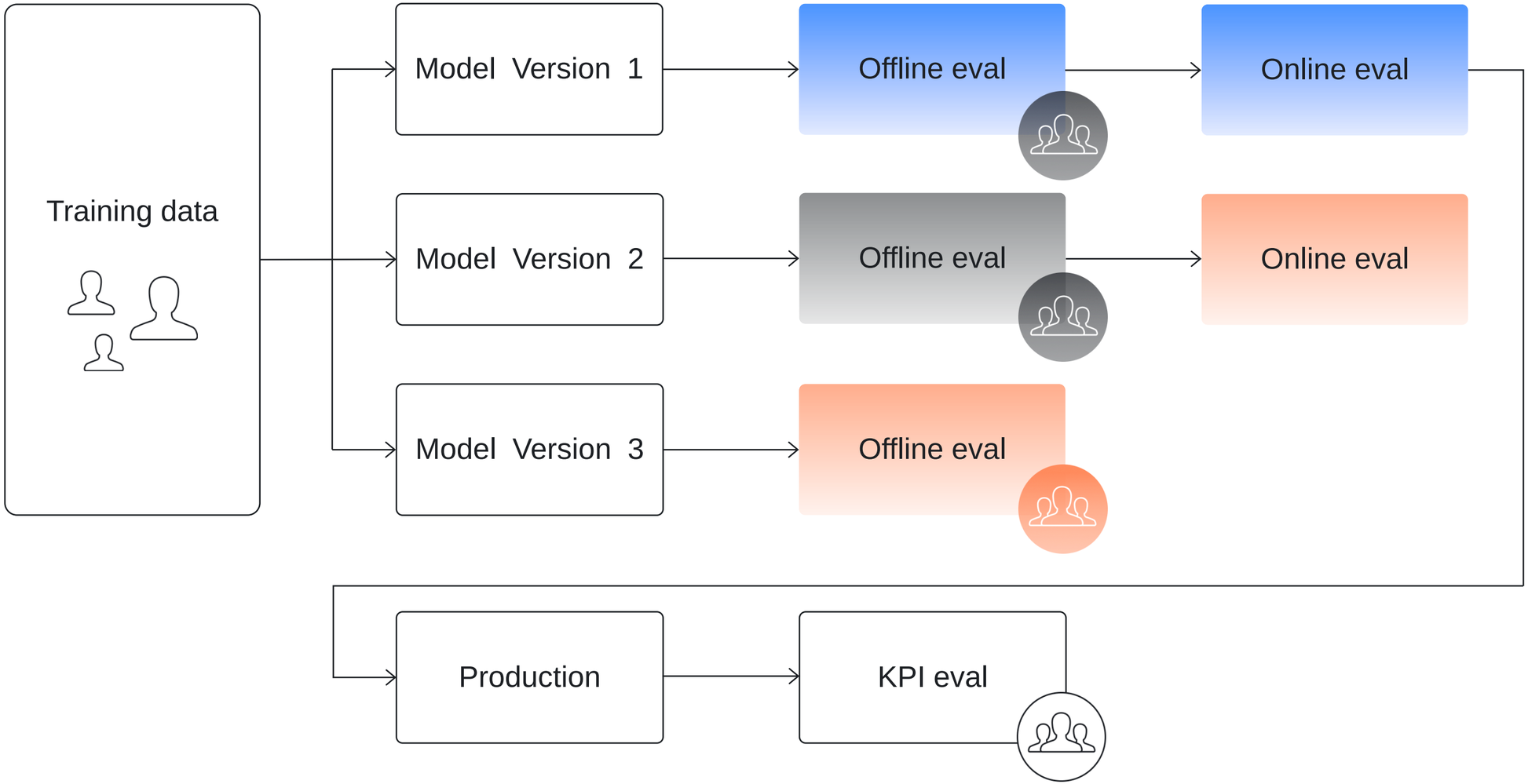
- At TheWebConf 2018
- At KDD 2018
- At SIGIR 2019

https://research.yandex.com/tutorials/online-evaluation

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

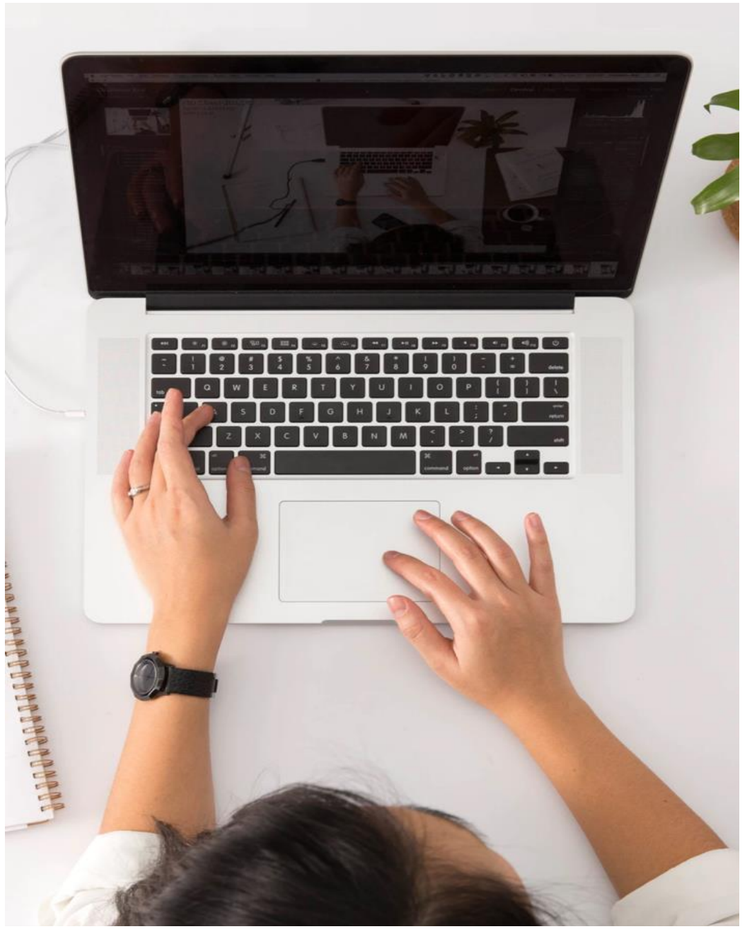
XX century — style management

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

https://media.glassdoor.com/l/6e/c3/ee/17/perfectserve-global-partner-cubicles.jpg

It can be different

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



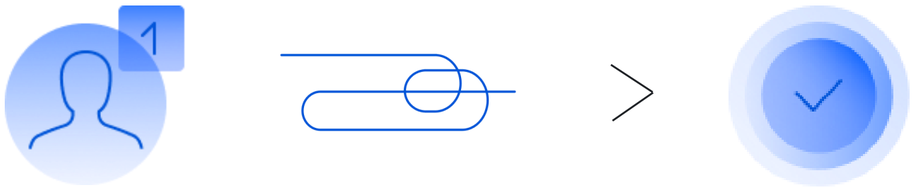
Crowdsourcing: specific way to design a business process

A big task

Cloud of performers

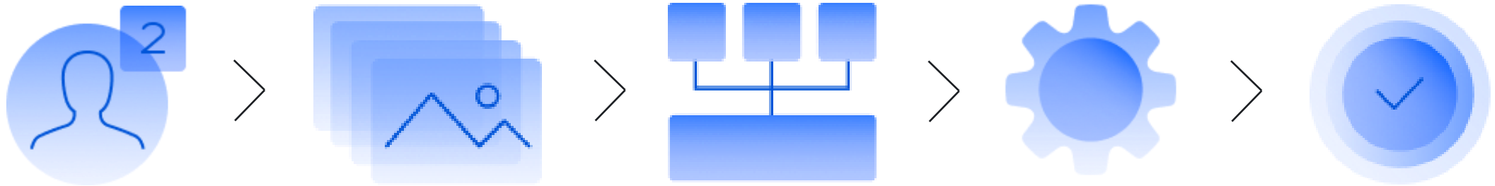
Result

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



Expert approach: rely on an expertise of a particular performer

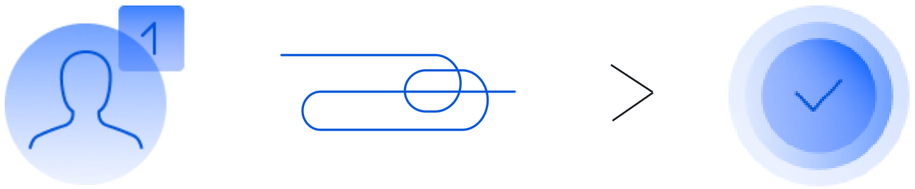
- Expensive
- Unmeasurable
- Hard to scale



Crowdsourcing approach

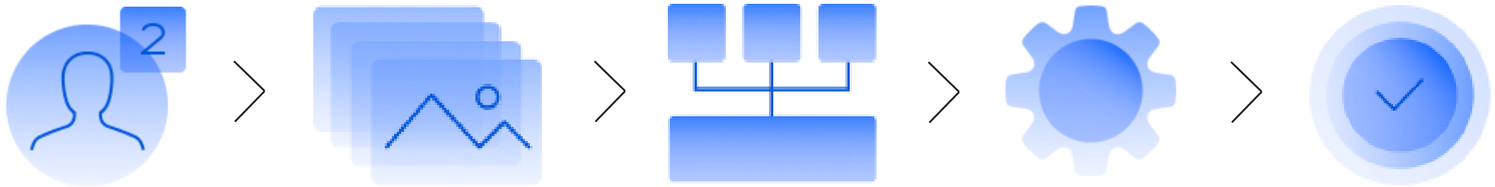
- Measurable
- Scalable
- Manageable

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

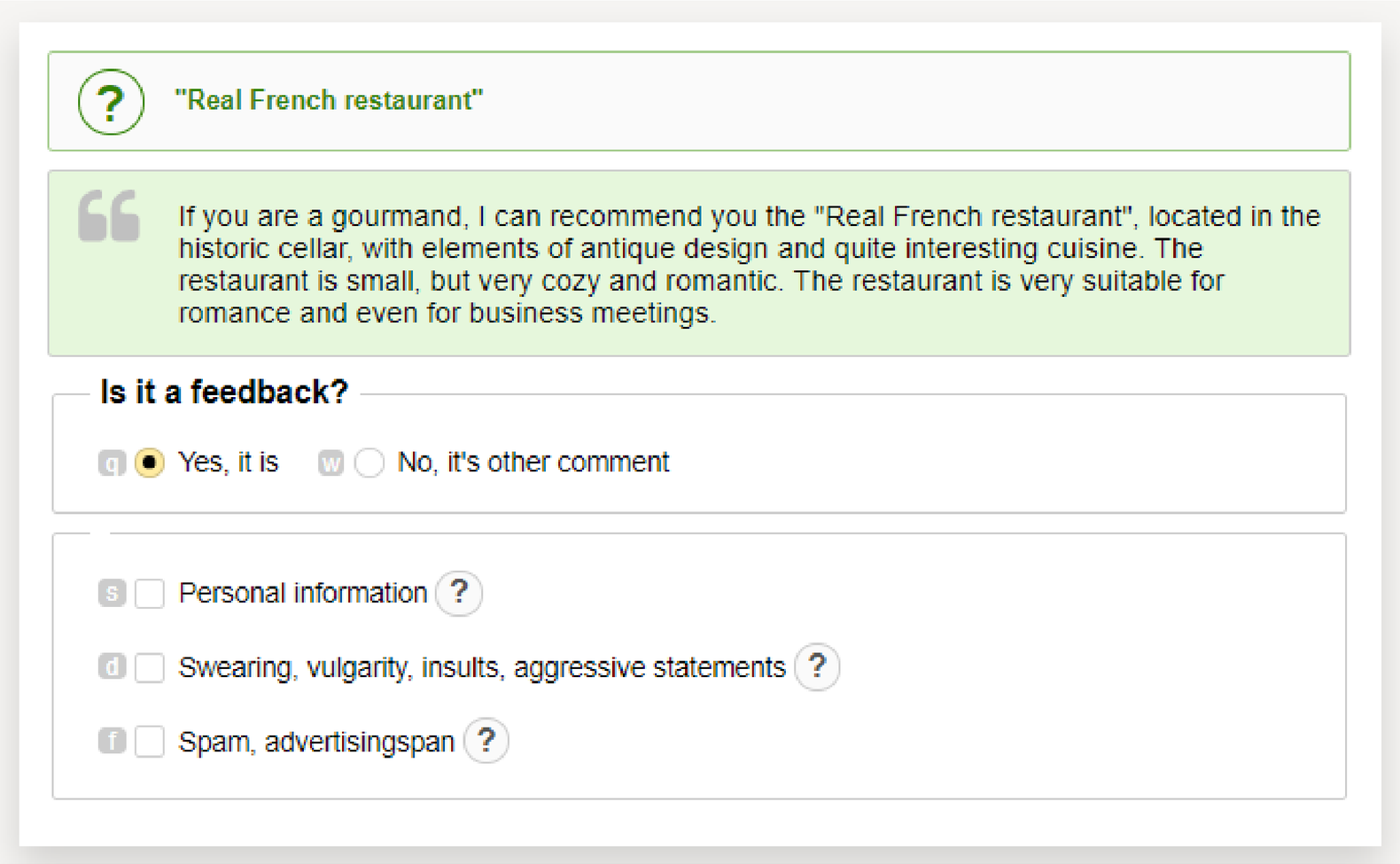
Example: binary classification

Is this cat white?

Yes

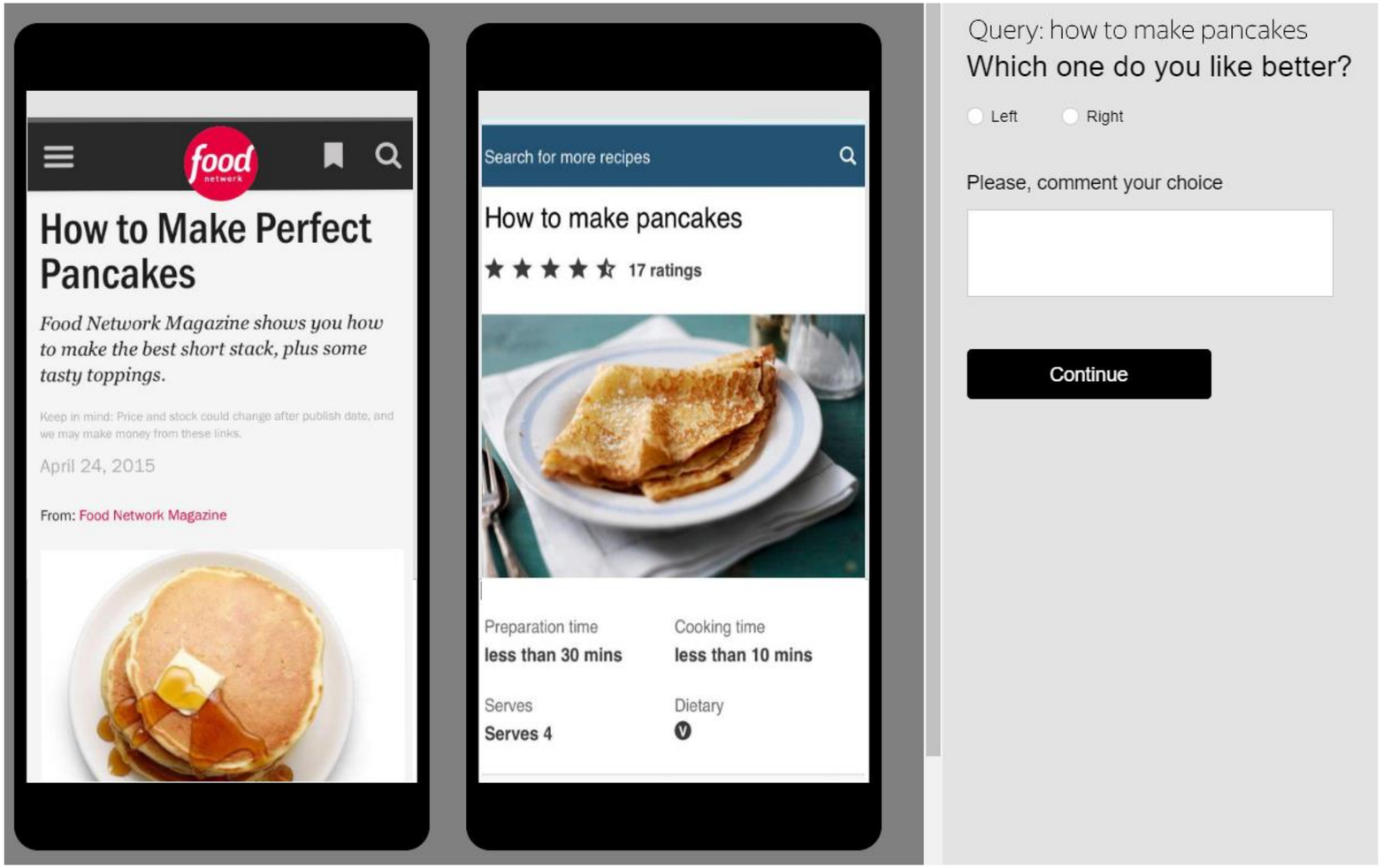
No

Example: multi classification

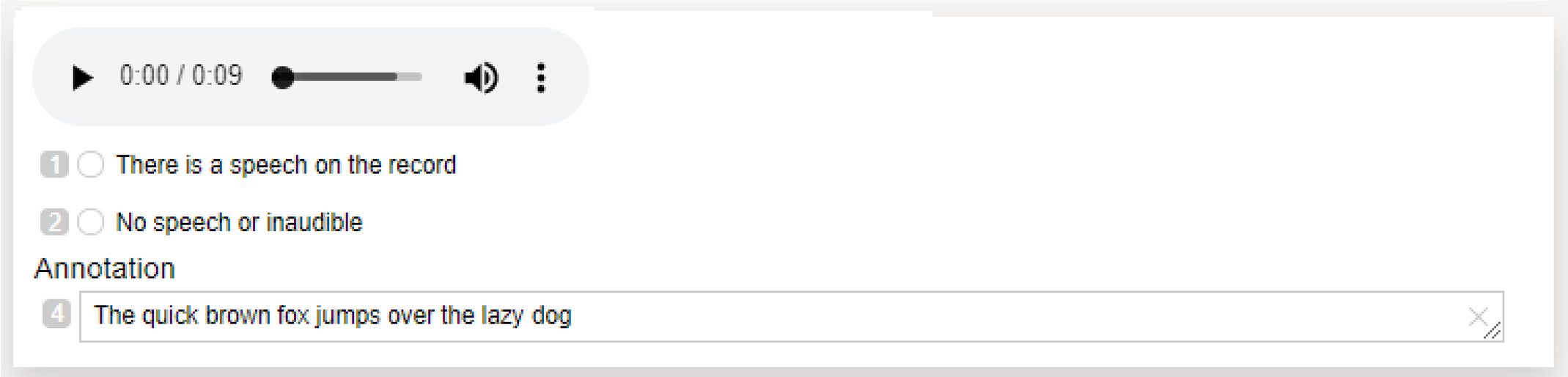


Example: multi classification with ordered labels

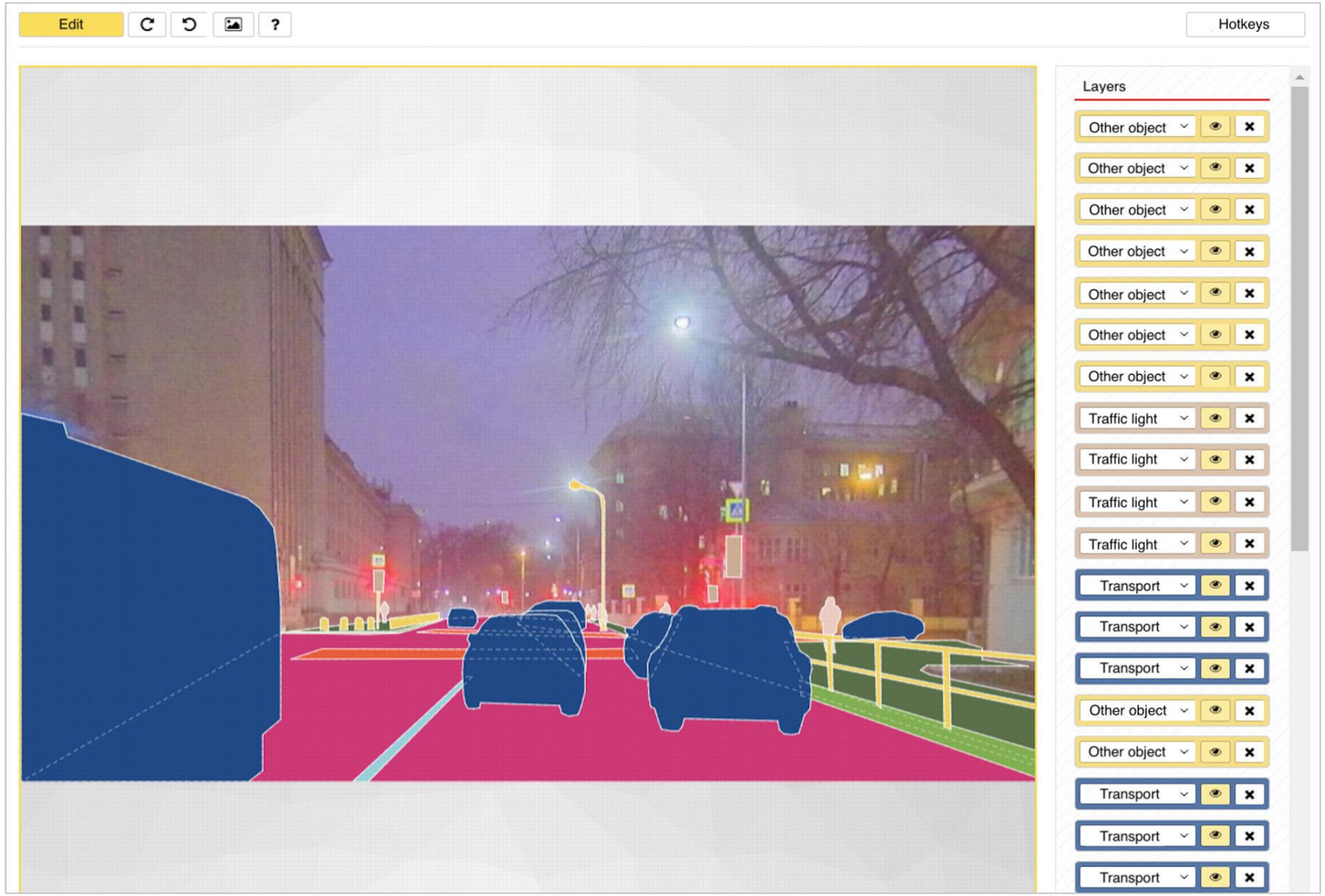
Examples: pairwise comparison



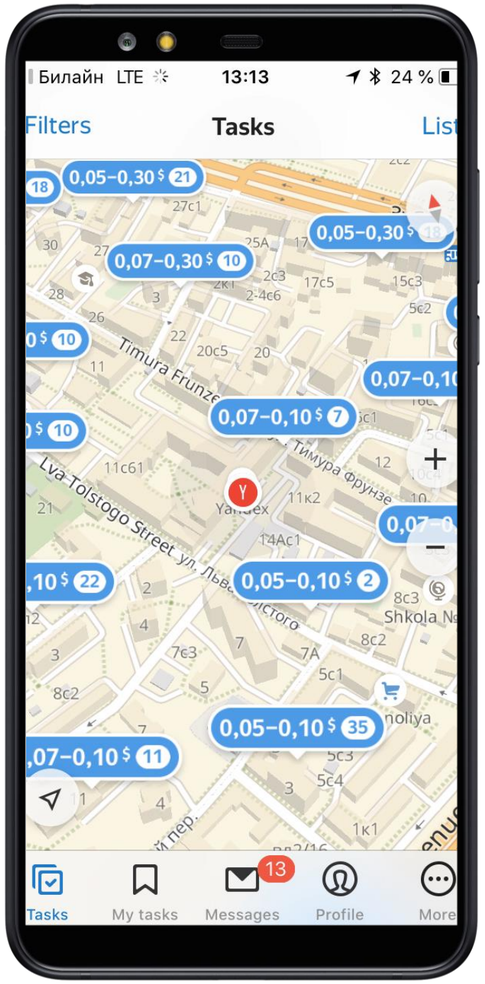
Examples: transcription with textual answers



Examples: object segmentation



Examples: spatial crowdsourcing



A crowdsourcing platform: two-sided market

Performers

Platform

Requesters

Crowdsourcing platforms: examples

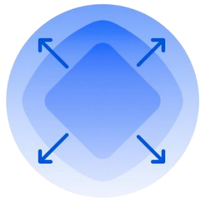
- Amazon Mechanical Turk
- Toloka
- Microworkers
- Gigwalk
- ClickWorker
- CloudFactory
- CrowdSource
- DefinedCrowd
- …

Pros of crowdsourcing platforms

24/7

Variety of skilled performers

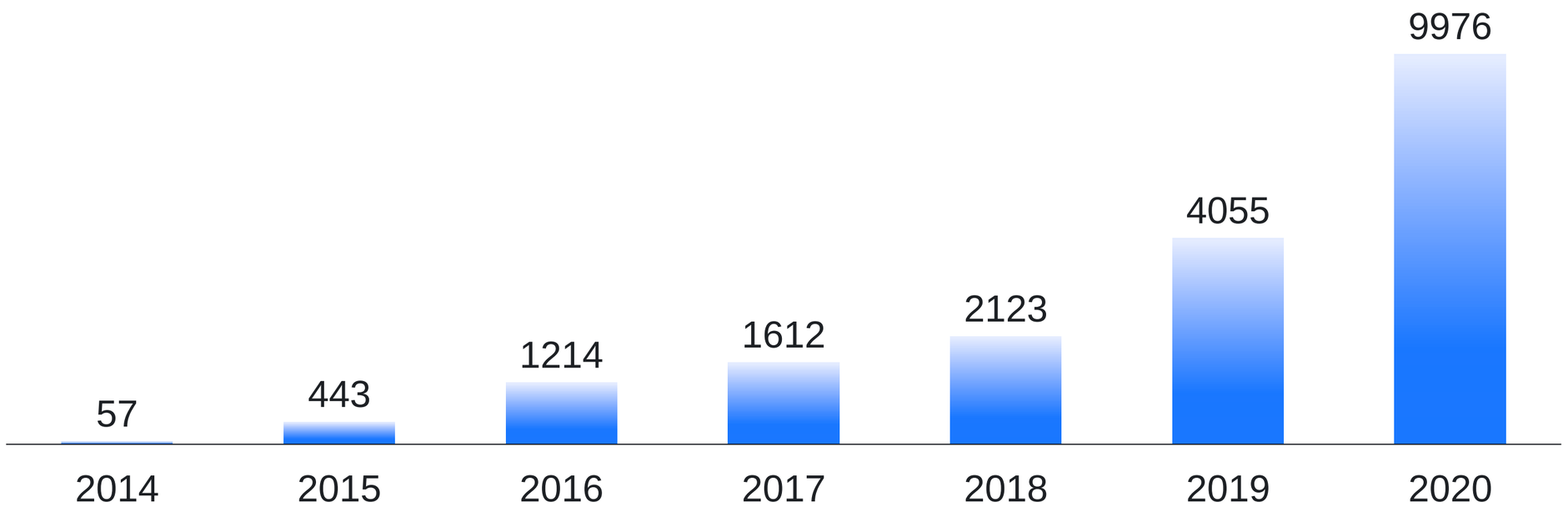
Vast region coverage



Ongoing processes

Crowdsourcing growth: Yandex experience

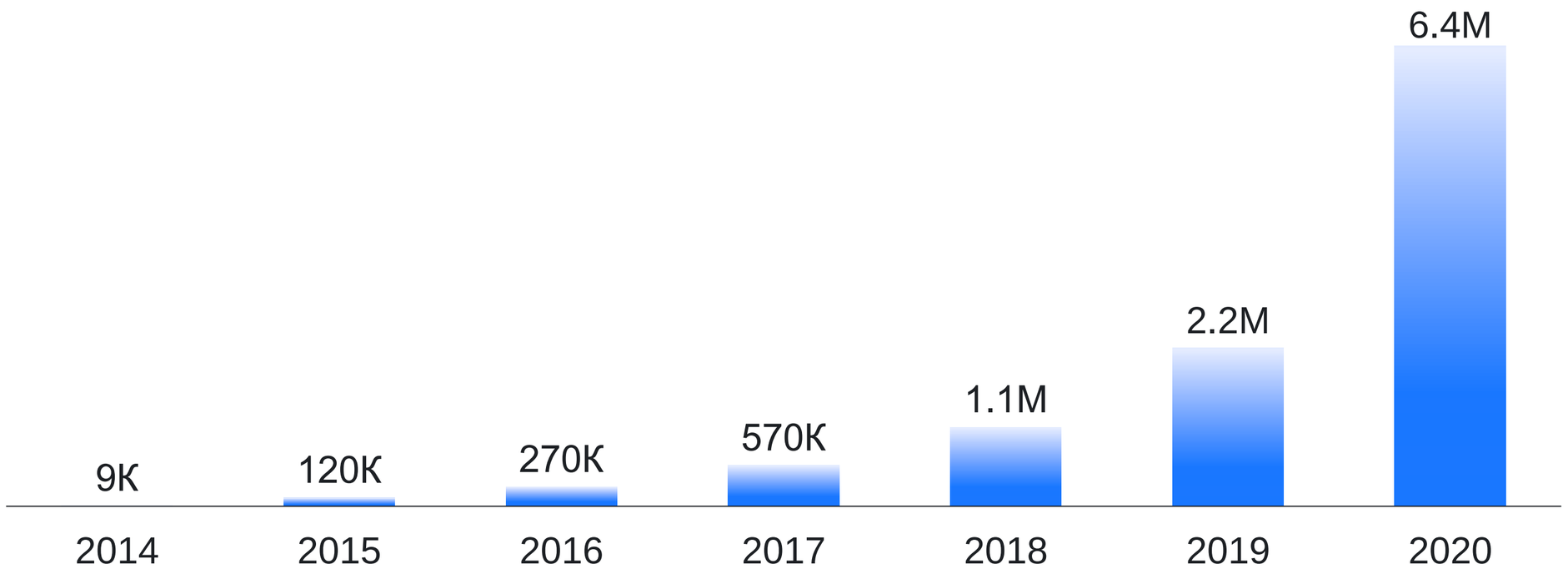
Different projects in Toloka



* 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

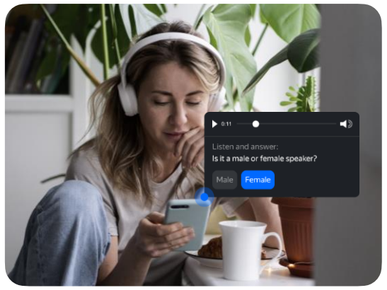
17M+
tasks

Toloka: real-life cases

Object classification

1000 photos

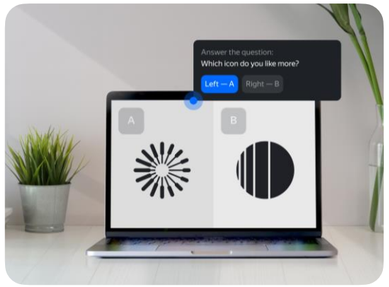
15 min $1.2



Audio transcription

100 recordings
25 minute long

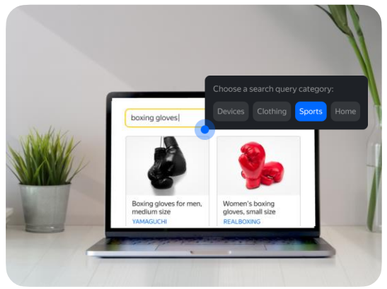
20 min $6



Side-by-side comparison

1000 photos

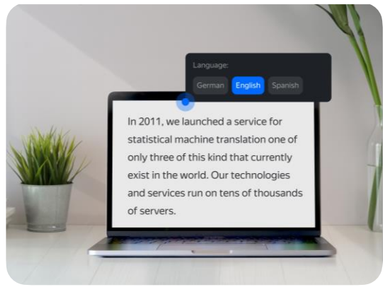
10 min $2.4



Object segmentation

1 000 objects in
100 photos

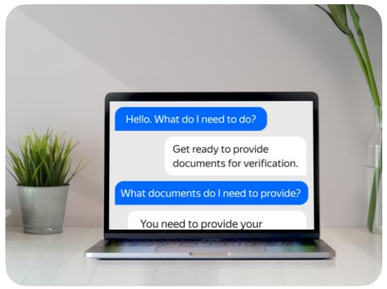
5 hrs $3.6



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

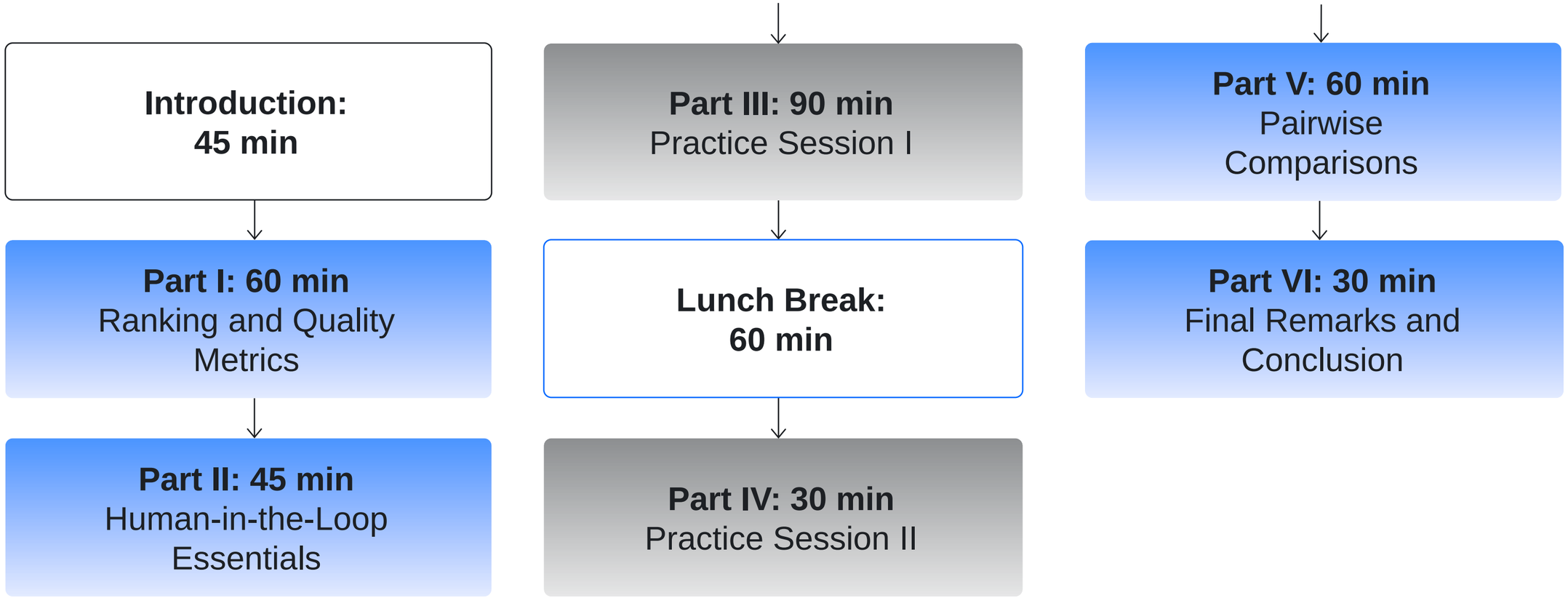
Theory

- 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

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

Grants for research
based on crowdsourcing