



Toloka

# Improving Recommender Systems with Human-in-the-Loop

Dmitry Ustalov, Natalia Fedorova, Nikita Pavlichenko,  
Maxim Kunakov, Fedor Zhdanov



# Part I

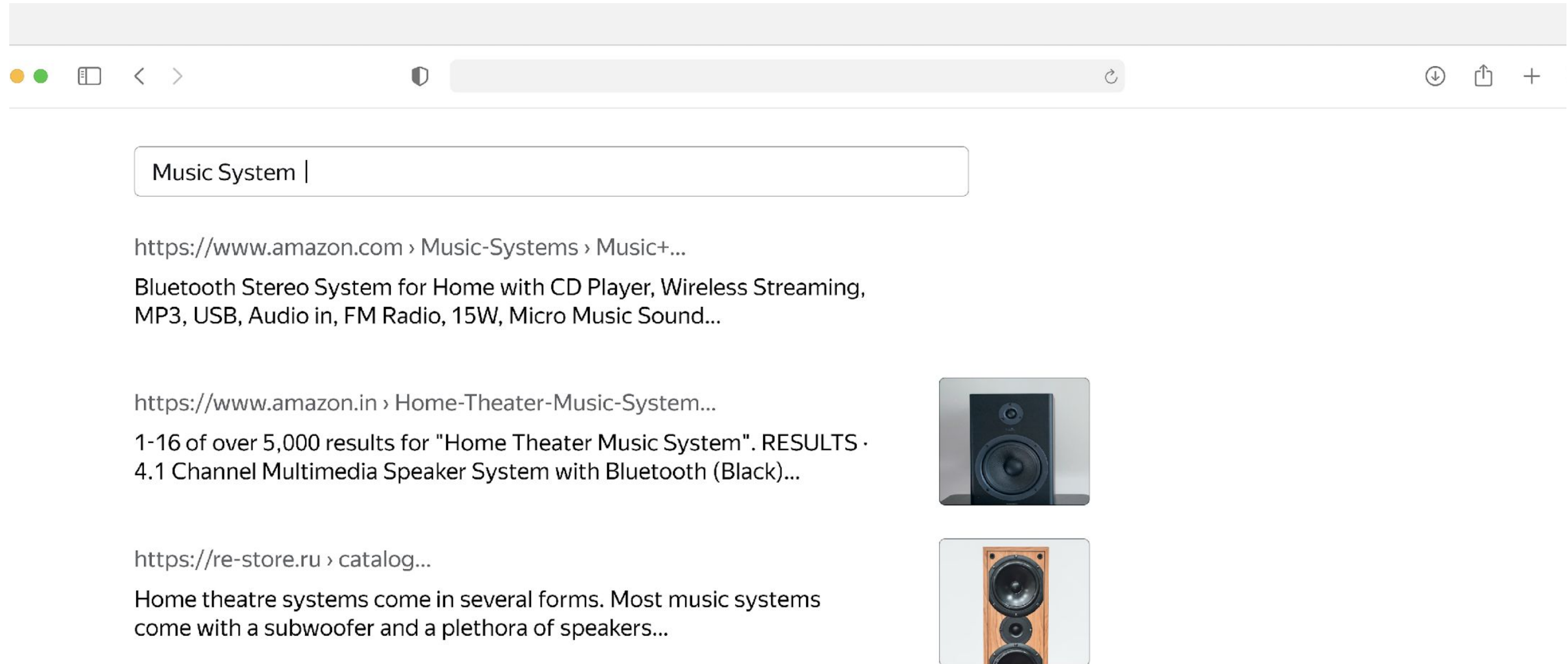
# Introduction

Fedor Zhdanov,  
Head of Machine Learning at Toloka

# Recommender Systems

- Recommender systems employ Machine Learning to produce recommendations
- Even state-of-the-art recommender models do not correlate with actual human preferences
- In this tutorial, we show how to gather real human judgments on recommendations using crowdsourcing
- But the methodology can be used beyond crowdsourcing

# Ranking and Recommender Evaluation




The screenshot shows a web browser window with a search bar containing the text "Music System |". Below the search bar, there are three search results. The first result is from "https://www.amazon.com" and is titled "Bluetooth Stereo System for Home with CD Player, Wireless Streaming, MP3, USB, Audio in, FM Radio, 15W, Micro Music Sound...". The second result is from "https://www.amazon.in" and is titled "1-16 of over 5,000 results for 'Home Theater Music System'. RESULTS · 4.1 Channel Multimedia Speaker System with Bluetooth (Black)...". The third result is from "https://re-store.ru" and is titled "Home theatre systems come in several forms. Most music systems come with a subwoofer and a plethora of speakers...". To the right of the second and third results are small images of speakers. The second image shows a black speaker, and the third image shows a wooden speaker.


Music System |

<https://www.amazon.com> › Music-Systems › Music+...  
Bluetooth Stereo System for Home with CD Player, Wireless Streaming, MP3, USB, Audio in, FM Radio, 15W, Micro Music Sound...

<https://www.amazon.in> › Home-Theater-Music-System...  
1-16 of over 5,000 results for "Home Theater Music System". RESULTS · 4.1 Channel Multimedia Speaker System with Bluetooth (Black)...



<https://re-store.ru> › catalog...  
Home theatre systems come in several forms. Most music systems come with a subwoofer and a plethora of speakers...

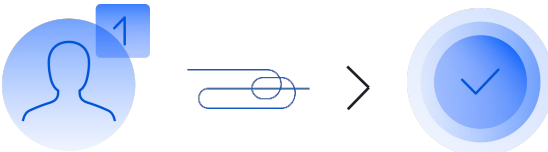


# Crowdsourcing

# Crowdsourcing as engineer-oriented approach

## In-house “expert”: managing people

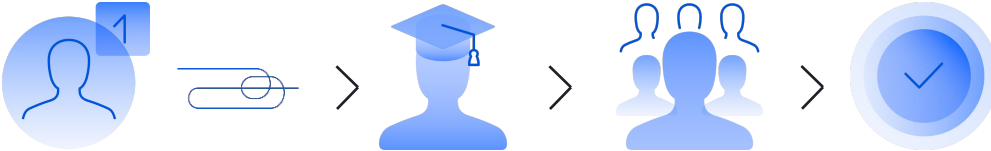
Direct managing in-house crowd



- ▶ Easy to setup
- ▶ Expensive
- ▶ Unmeasurable
- ▶ Impossible to scale

## BPO / vendor

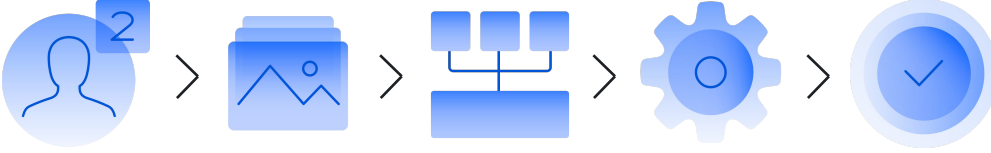
Access to crowd via third-party BPO who manage them



- ▶ Quick access to crowd
- ▶ Expensive
- ▶ Unmeasurable
- ▶ Hard to scale

## Crowdsourcing

Technologically managing crowd as yet another computing power



- ▶ Measurable
- ▶ Scalable
- ▶ Manageable
- ▶ Requires advanced tech

# Crowdsourcing examples: use cases to improve search relevance

Product + search query

A browser window showing a crowdsourcing task. The description is "Music system". Below it, the question is "Is this search result relevant to this query?". There are five radio button options: 1 Relevant, 2 Rather relevant, 3 Irrelevant, 4 Can't say, and P 404.

Category + search query

A browser window showing a crowdsourcing task. The instruction is "Classify how relevant the category is to the search query". The query is "Kitchen table" and the category is "Dining room furniture". There are seven radio button options: 1 Excellent, 2 Good, 3 Fair, 4 Bad, 5 Adult, 6 Junk, and 7 Unreadable text.

Image + search query

A browser window showing a crowdsourcing task. The question is "Is the below item relevant to someone who performed the given search?". The search query is "coffee bean grinder". An image of a coffee bean grinder is shown. To the right, there are five radio button options: E Exact, S Substitute, C Compliment, I Irrelevant, and U Unjudicable.

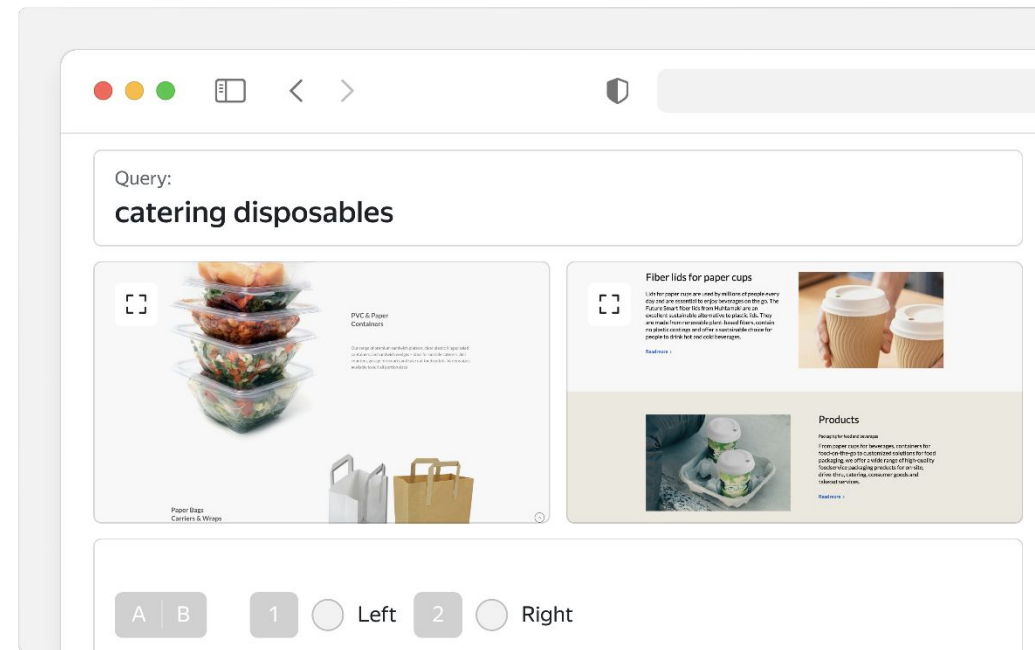
Filters + category

A browser window showing a crowdsourcing task. The instruction is "Classify filter relevance to the product category". The filter is "high heel" and the category is "Women's shoes". There are two text input fields: "S Google first text" and "D Google second text". Below them are four radio button options: 1 Excellent, 2 Good, 3 Fair, and 4 Bad.

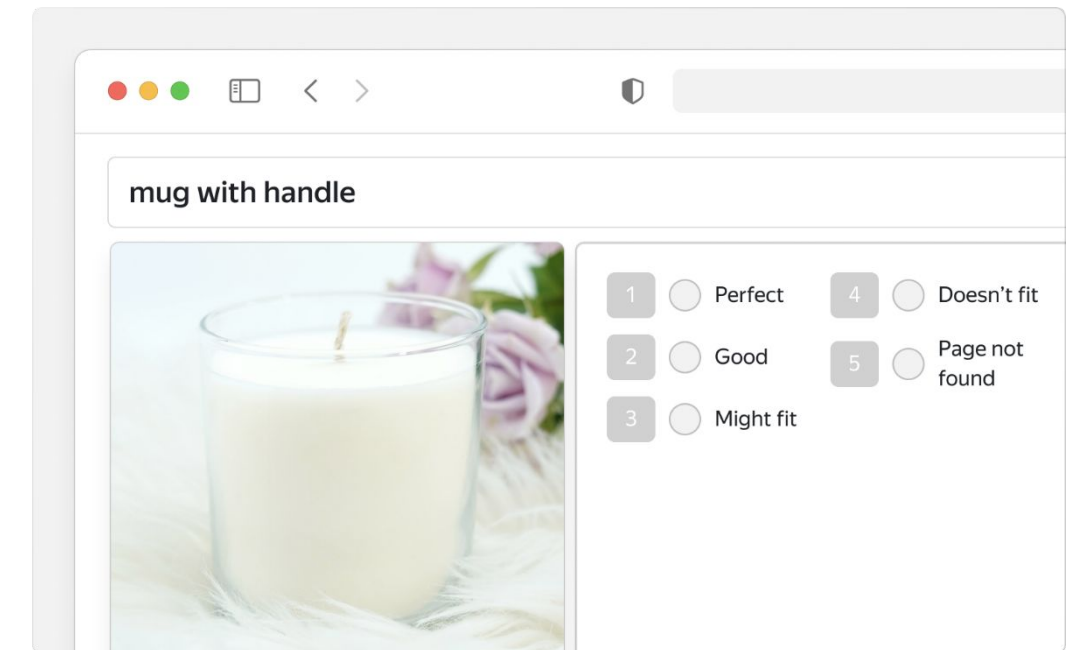


# Crowdsourcing examples: use cases to improve search relevance

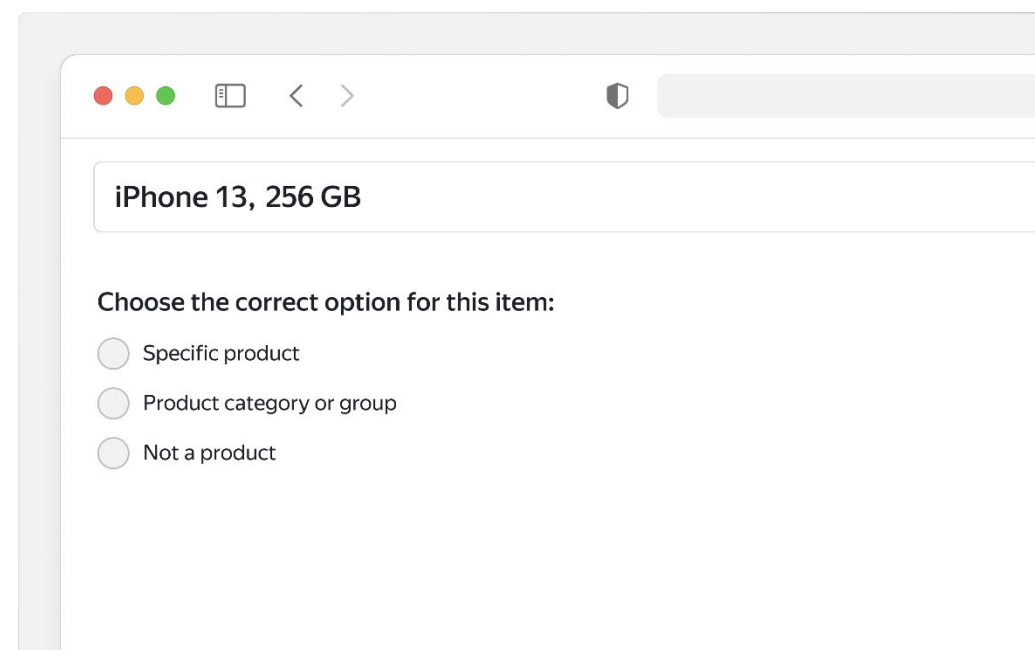
Side-by-side comparison of search results



Identify spam or irrelevant matches

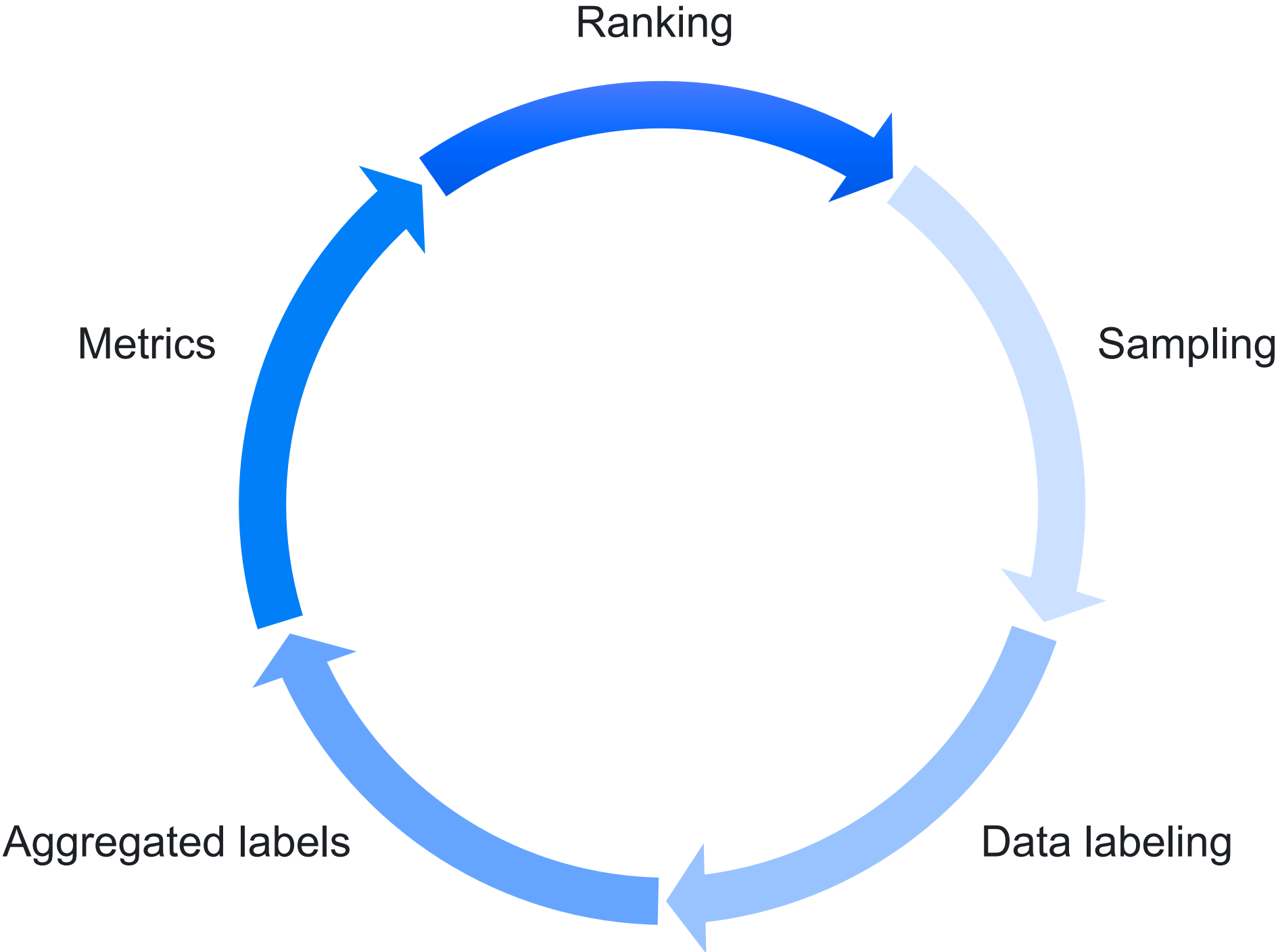


Classify type of search query (broad vs narrow)



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

# Human-in-the-Loop Pipeline for Offline Metrics



# Learning outcomes

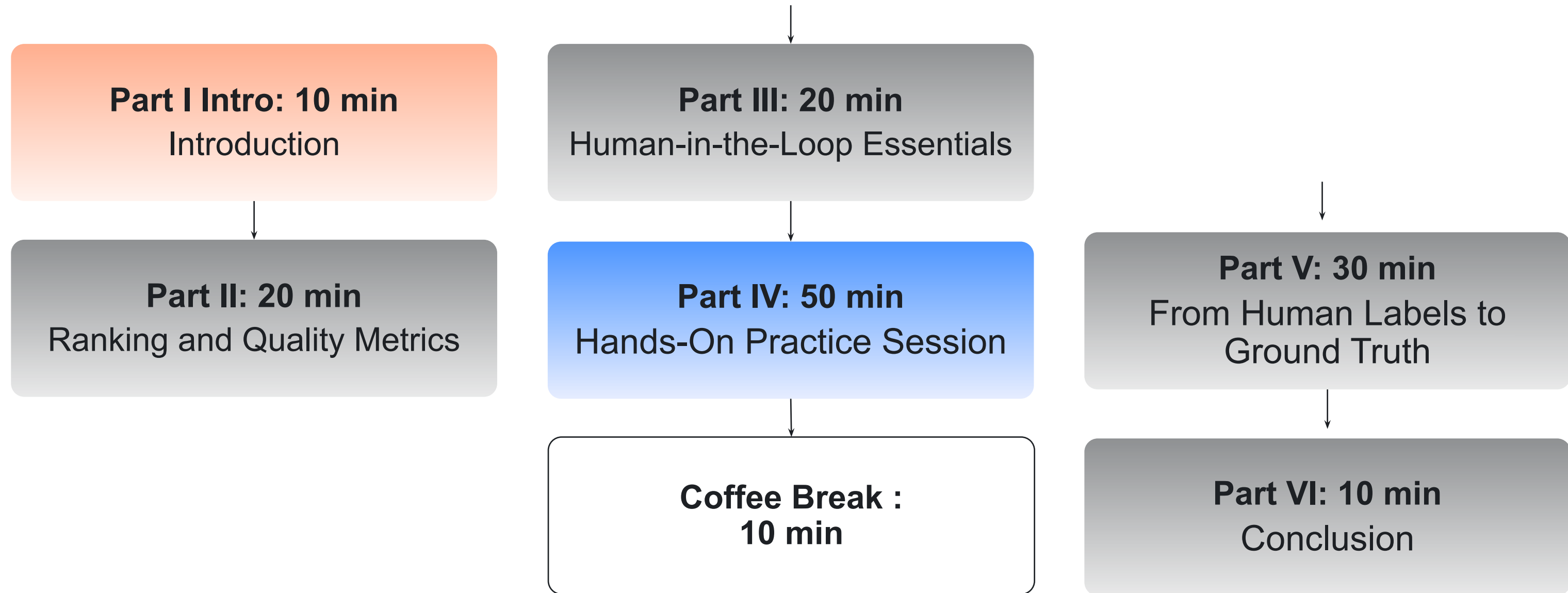
## Theory

- ▶ Offline approach for ranking evaluation
- ▶ Use crowdsourcing for industrial applications

## Practice

- ▶ Build scalable data labeling pipelines
- ▶ Run crowdsourcing projects with real annotators
- ▶ Program Human-in-the-Loop via public Python libraries (Toloka-Kit)

# Tutorial Schedule



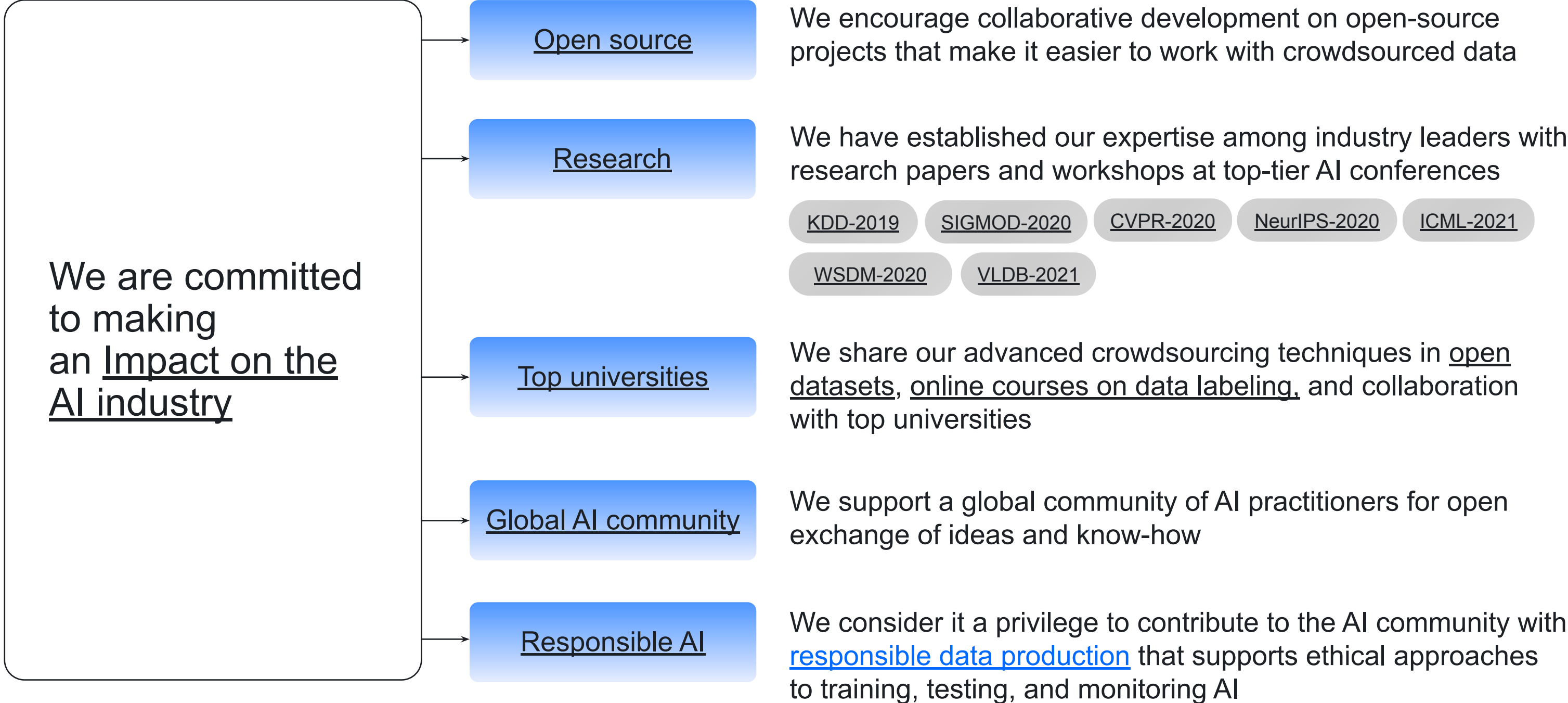
# Toloka Research Grants Program

- ▶ 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/>

# Our team helps the AI industry



Join our Slack:  
**recsys\_2022**



<https://bit.ly/3eYIX2P>

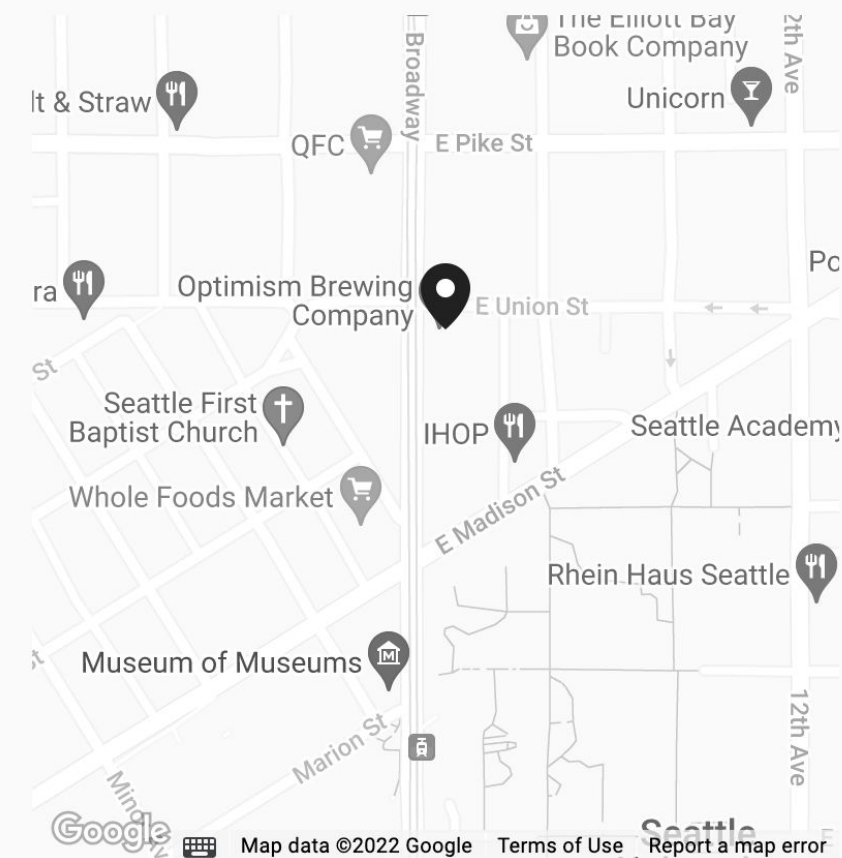


# Invite you to Toloka happy hour

# 21 September at 6 p.m.

## OUR LOCATION

1158 Broadway  
Seattle, WA, 98122  
United States



# Thank You!

**Fedor Zhdanov**

Head of Machine Learning at Toloka



[fedorzh@toloka.ai](mailto:fedorzh@toloka.ai)



<https://toloka.ai/events/recsys-2022/>



<https://www.linkedin.com/company/toloka/>



<https://twitter.com/tolokaai>



<https://tolokacommunity.slack.com/ssb/>



<https://github.com/Toloka>