Toloka

Crowdsourcing for Information Retrieval
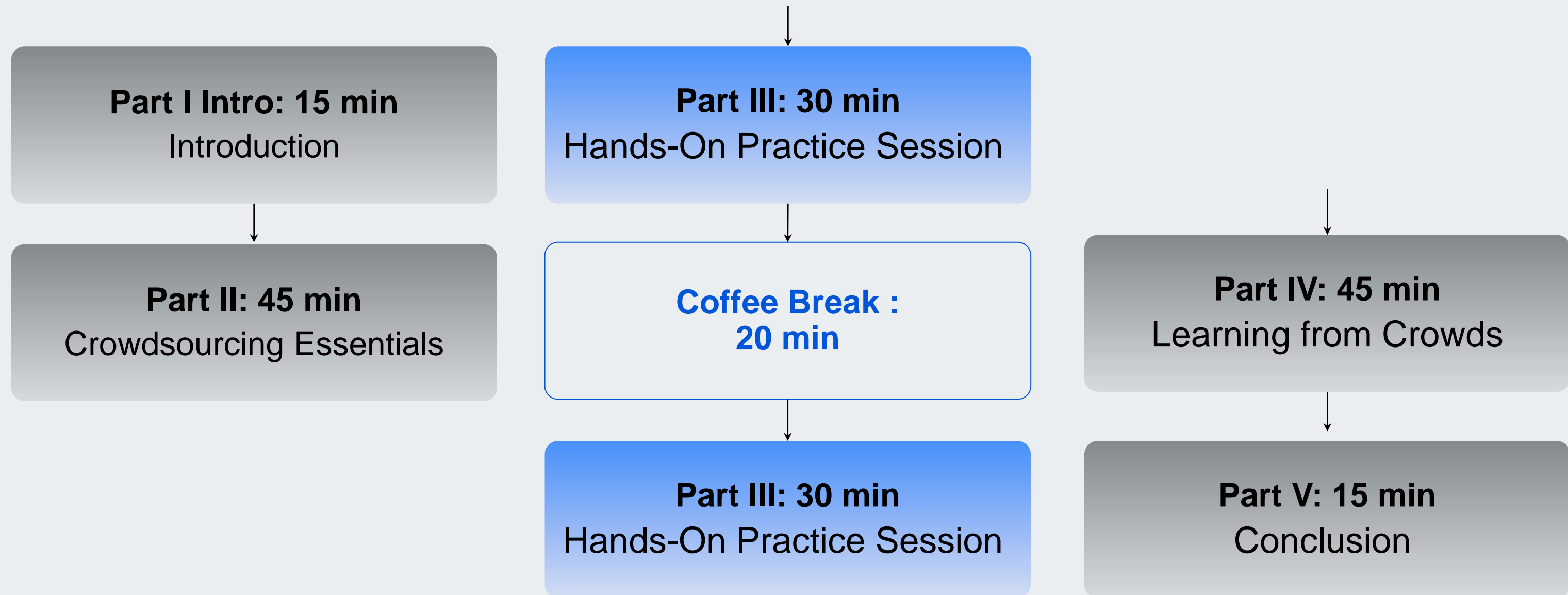
Dmitry Ustalov, Natalia Fedorova, Nikita Pavlichenko, Alisa Smirnova, Daniil Likhobaba

Toloka

Part III
Practice Session

Daniil Likhobaba,
Researcher at Toloka

Tutorial Schedule

Part I Intro: 15 min
Introduction

Part II: 45 min
Crowdsourcing Essentials

Part III: 30 min
Hands-On Practice Session

**Coffee Break :
20 min**

Part III: 30 min
Hands-On Practice Session

Part IV: 45 min
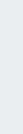Learning from Crowds

Part V: 15 min
Conclusion

※ Toloka

Practice session: scope

Imagine that you are developing a conversational agent and you need to create an intent classification pipeline for user queries

► You have a dataset of user queries

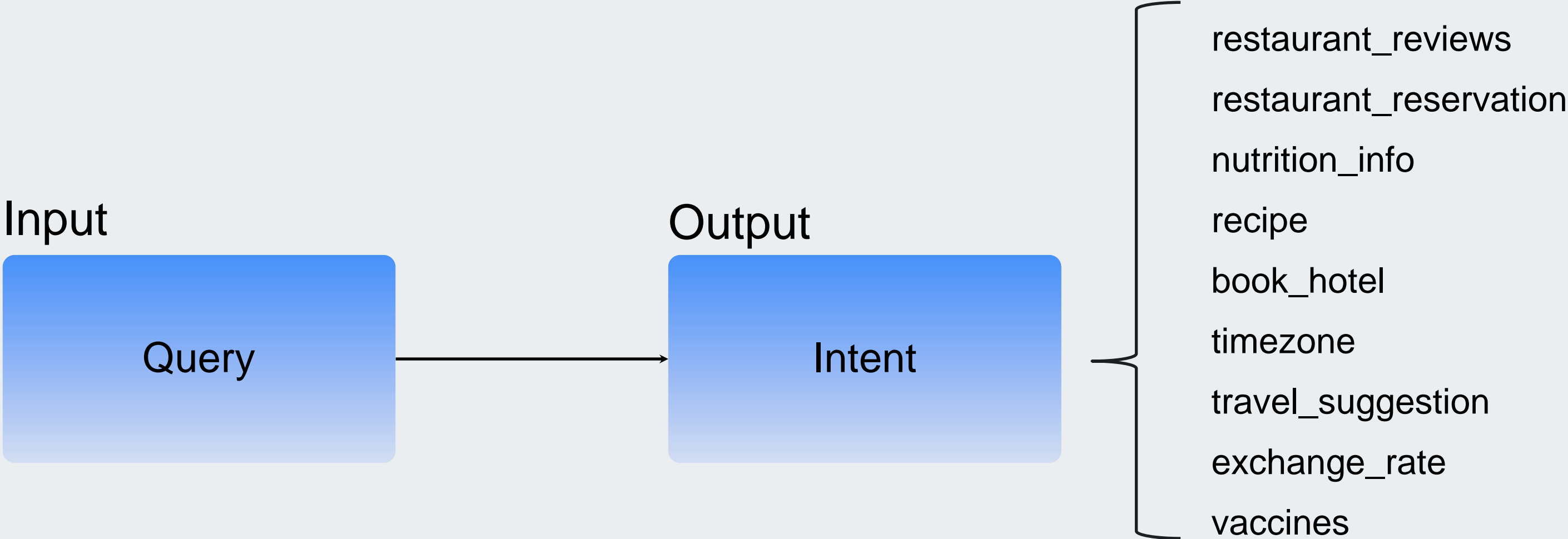This is your goal during the current part of our tutorial

► You need to get intents for these queries

► This collected data will be used to train a classification model

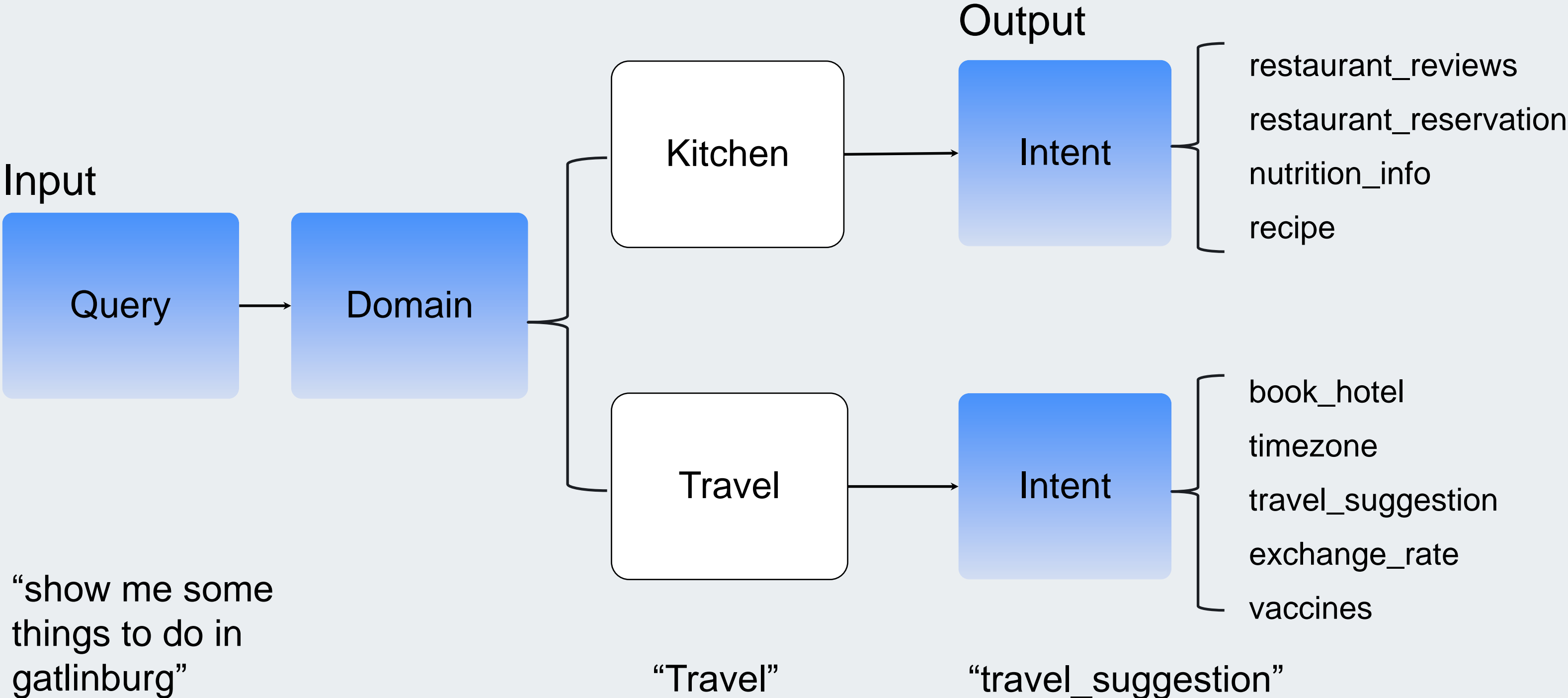This will be done during the next part of our tutorial

⁂ Toloka

5

How do we label it?

Input

Output

Query → Intent

restaurant_reviews

restaurant_reservation

nutrition_info

recipe

book_hotel

timezone

travel_suggestion

exchange_rate

vaccines

"show me some things to do in gatlinburg"

"travel_suggestion"

Toloka

Decomposition

Input

Query → Domain

"show me some things to do in gatlinburg"

Kitchen → Intent
- restaurant_reviews
- restaurant_reservation
- nutrition_info
- recipe

Travel → Intent
- book_hotel
- timezone
- travel_suggestion
- exchange_rate
- vaccines

Output

"Travel"

"travel_suggestion"

Toloka

Intent classification

Project setup

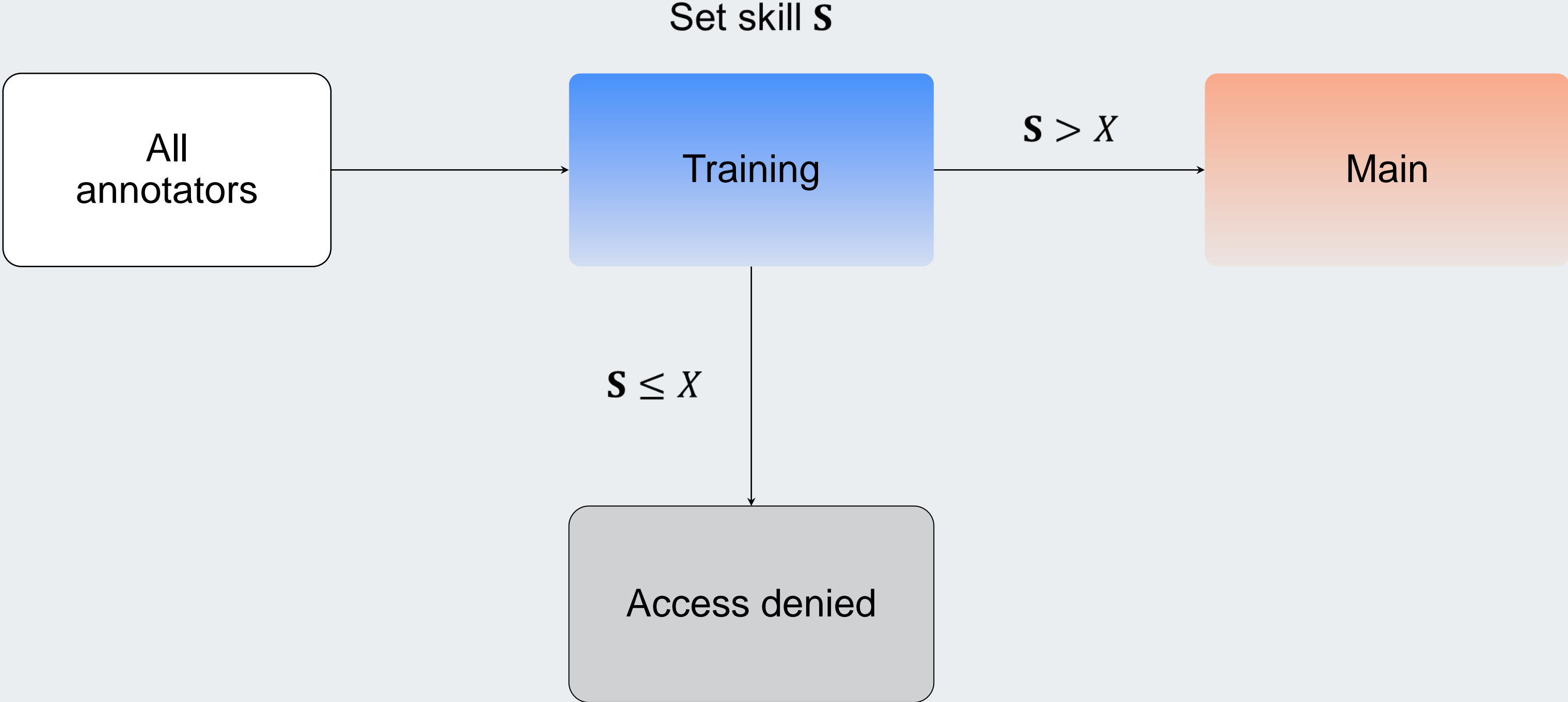**Task**

Choose the category and intent for a given text

Settings
► Type: Classification
► Training: yes
► Quality control: golden set
► Pay: $0.01 per 1 page with 6 tasks
► Time: 600 sec per 1 page

☀ Toloka

Crowd training

Set skill $\mathbf{S}$

All
annotators

Training

$\mathbf{S} > X$

Main

$\mathbf{S} \leq X$

Access denied

☀ Toloka

We are starting the practice session: 60 min

We share links to:

► A step-by-step guide to configure and run data labeling

► A dataset with queries and intents

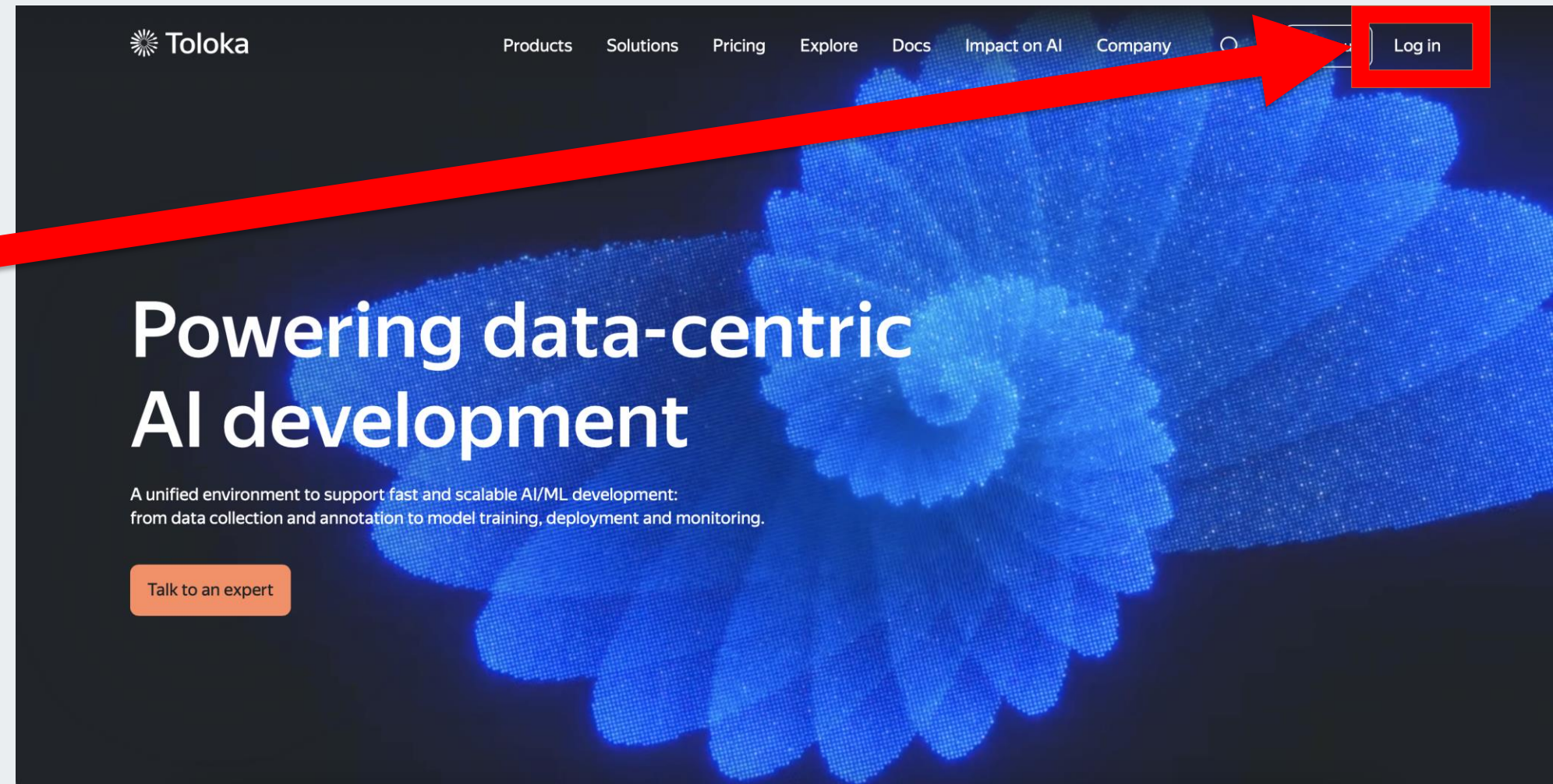Promo code can be obtained at
toloka.ai/academy/education-pricing

Step-by-step instructions:

bit.ly/412Am0H

Toloka

Sign up on Toloka as a requester
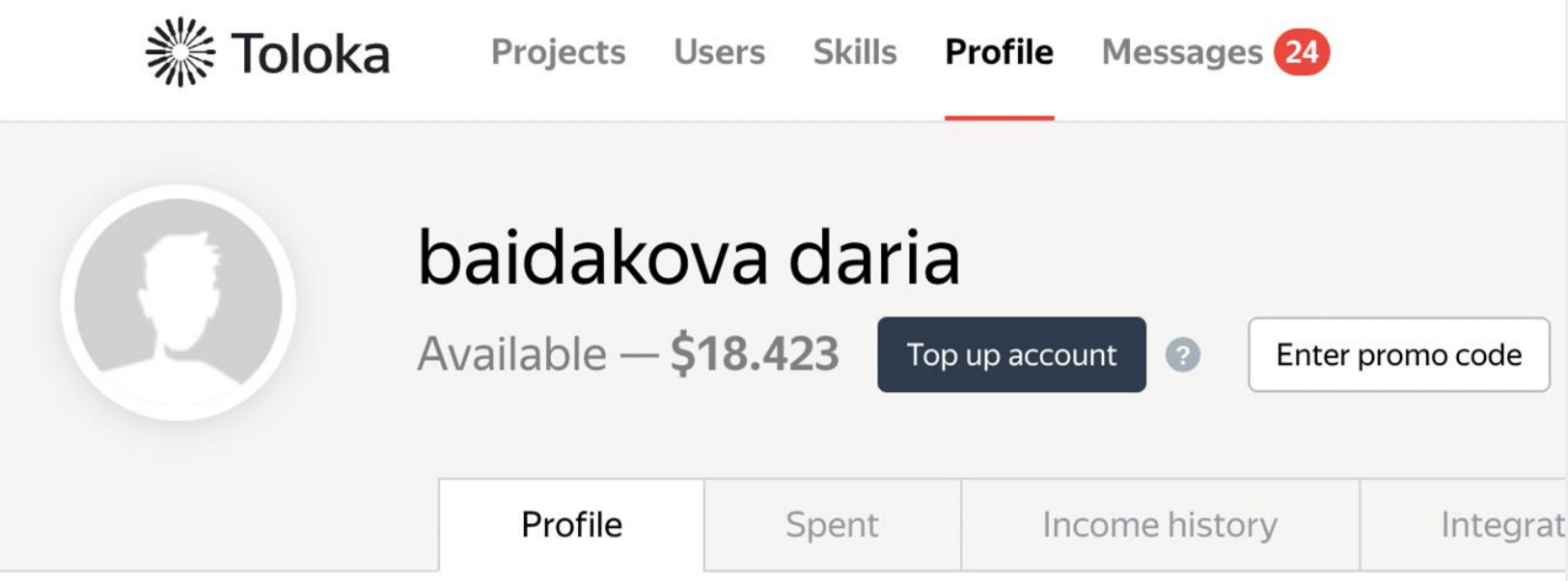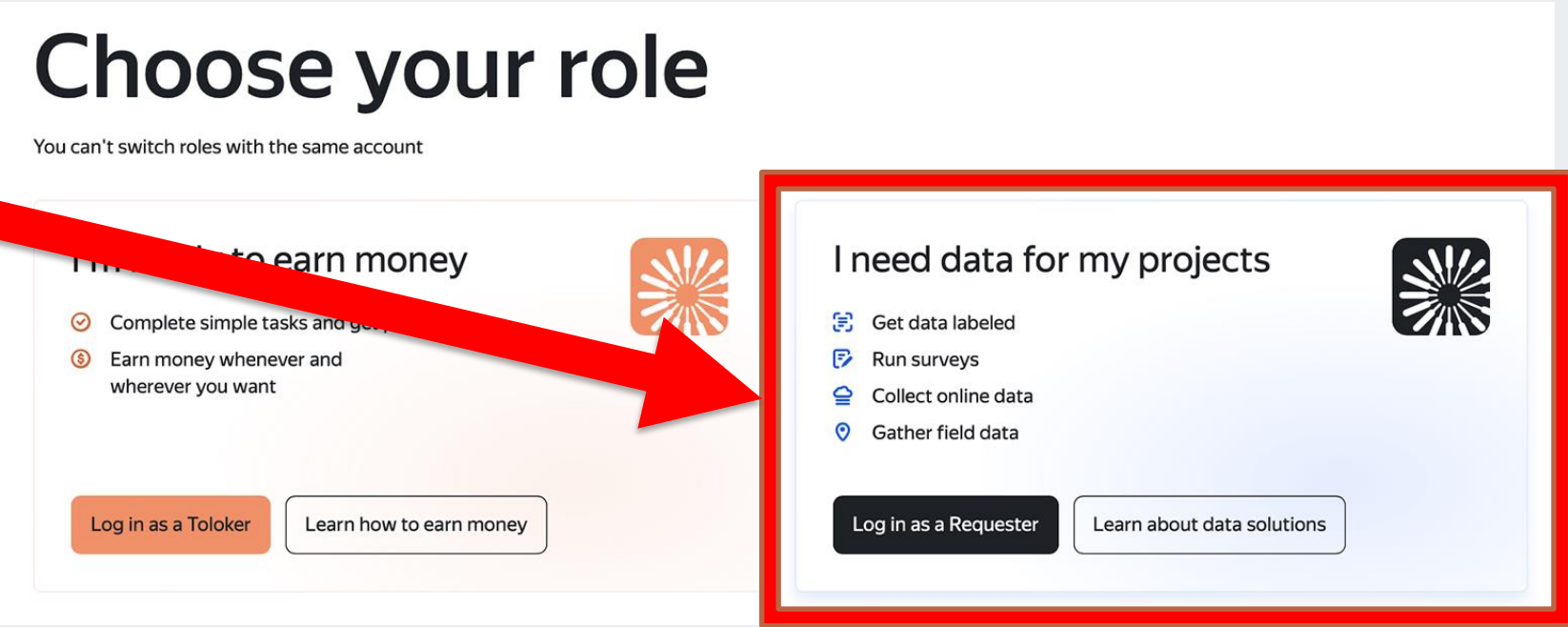
1. Go to toloka.ai

2. Click on "Log in"

Sign up on Toloka as a requester

3. Click on "I need data for my projects"

4. Use your email and password

5. Go to Profile and use promocode

Toloka

Thank You!

Daniil Likhobaba

Research Engineer at Toloka

research@toloka.ai

https://toloka.ai/

Linked in
https://www.linkedin.com/
company/toloka/

https://twitter.com/tolokaai

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

GitHub
https://github.com/Toloka