

Yandex

Audio transcription project

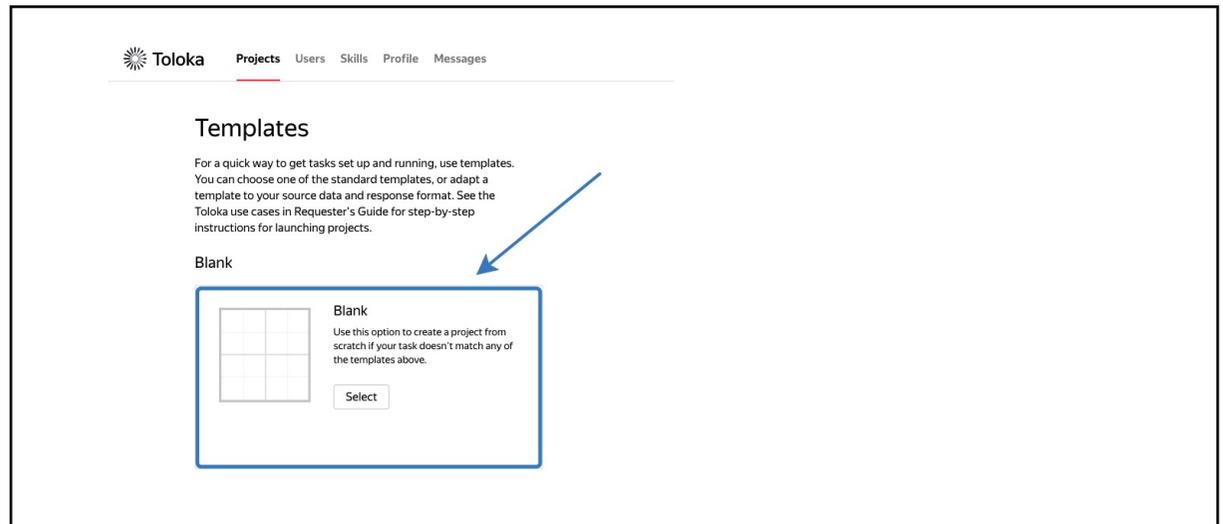
Alexey Drutsa, Dmitry Ustalov, Valentina Fedorova, Olga Megorskaya, Daria Baidakova

Project setup

1. Click the button **+ Create project**.



2. Choose a **Blank** template.



3. Fill in a title and a description of your project, then press **Save**.

Edit project

Back to the old interface Cancel Finish

1 General information

Name to show performers *

Description for performers

Private comment ?

Save

★★★★★

Transcribe the recording

Type what you hear in the recording.

0\$ per task ~0\$ per hour

4. Make sure “Template builder” is chosen. Delete the existing config and paste [the new code from here](#).

2 Task interface

Editor

HTML / JS / CSS ? **Template builder ?**

Config

```
1
2 vars: {
3   "author": "assessorlead",
4   "task": "new",
5   "validString": {
6     "type": "condition.all",
7     "conditions": [
8       {
9         "type": "condition.schema",
10        "schema": {
11          "type": "string",
12          "pattern": "^[a-zA-Z\\*\\s]{1,}$"
13        }
14      }
15    ],
16    "hint": "Use only lowercase letters and spaces"
17  }
18 ,
19 view: {
20   "type": "view.list",
21   "items": [
```

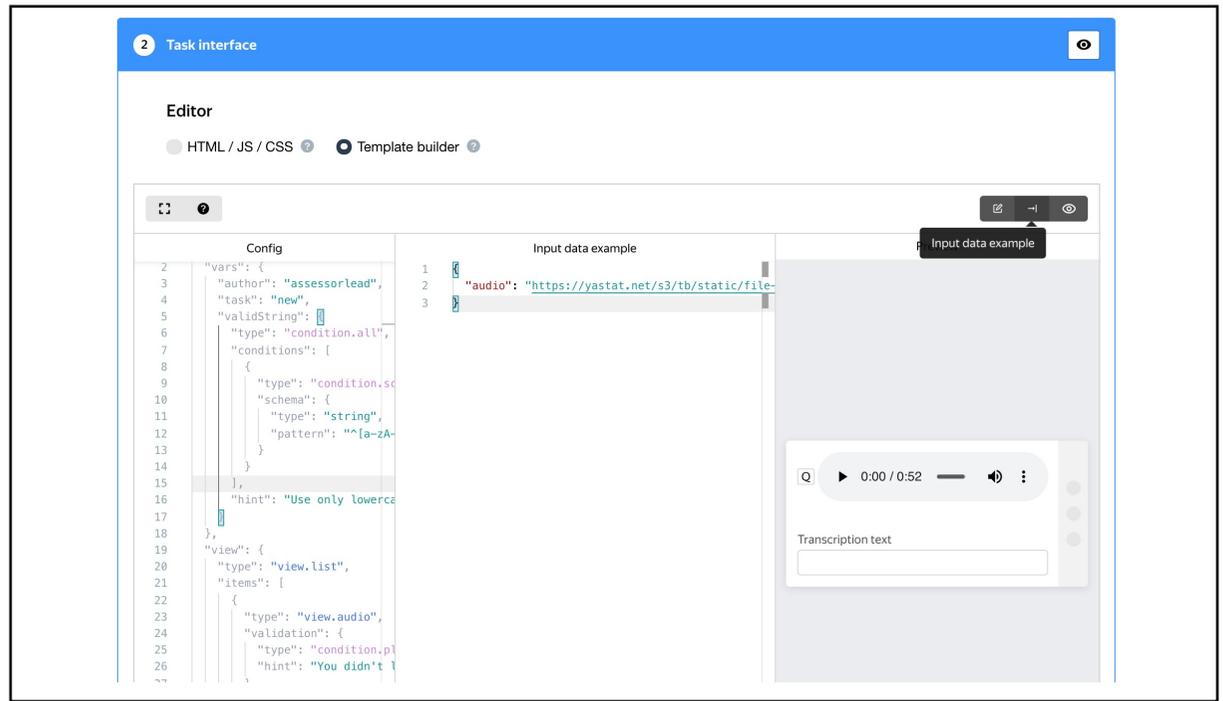
Preview

0:00 / 0:52

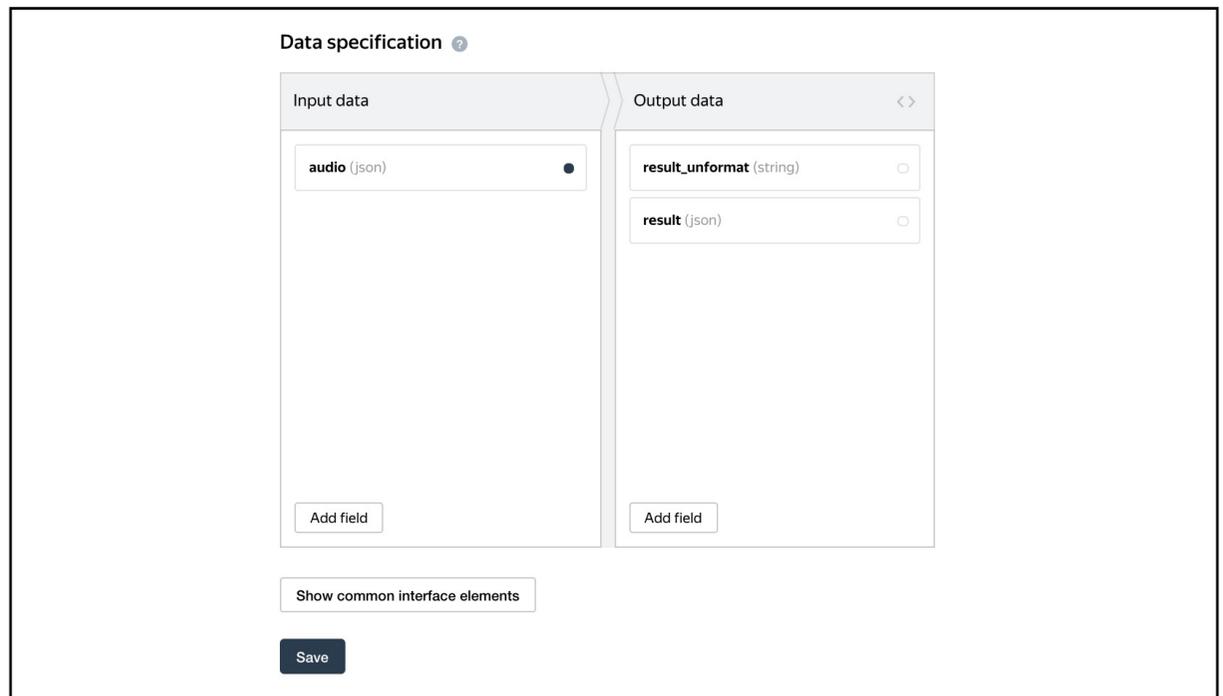
Transcription text

5. Press on **Input Data example** in the top right corner.

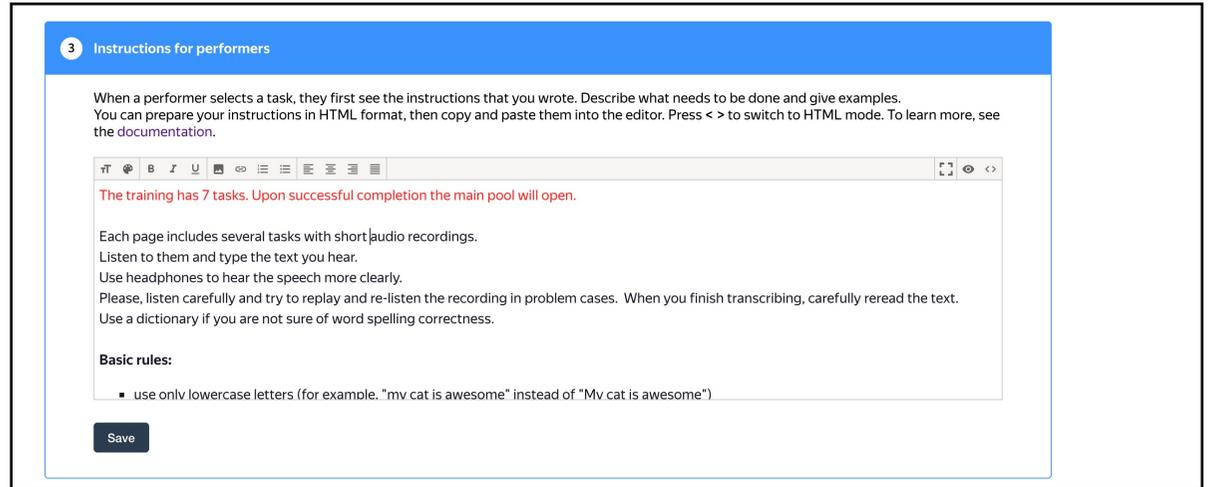
6. Paste [the following code](#) for Input Data Example so it looks like on the picture.



7. Make sure the specifications look like this and save the template:



8. Write an instruction or just paste [our version](#). Then **Save** this block.



3 Instructions for performers

When a performer selects a task, they first see the instructions that you wrote. Describe what needs to be done and give examples. You can prepare your instructions in HTML format, then copy and paste them into the editor. Press < > to switch to HTML mode. To learn more, see the documentation.

The training has 7 tasks. Upon successful completion the main pool will open.

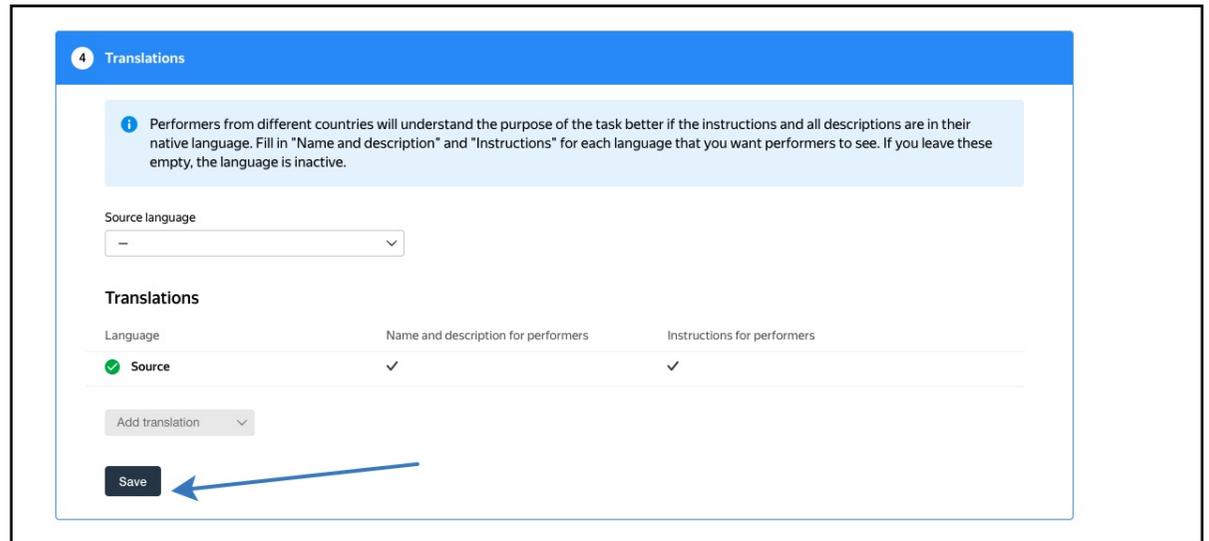
Each page includes several tasks with short audio recordings. Listen to them and type the text you hear. Use headphones to hear the speech more clearly. Please, listen carefully and try to replay and re-listen the recording in problem cases. When you finish transcribing, carefully reread the text. Use a dictionary if you are not sure of word spelling correctness.

Basic rules:

- use only lowercase letters (for example, "mv cat is awesome" instead of "Mv cat is awesome")

Save

9. Leave the "Translations" section as it is and **Save** this block.



4 Translations

Performers from different countries will understand the purpose of the task better if the instructions and all descriptions are in their native language. Fill in "Name and description" and "Instructions" for each language that you want performers to see. If you leave these empty, the language is inactive.

Source language

—

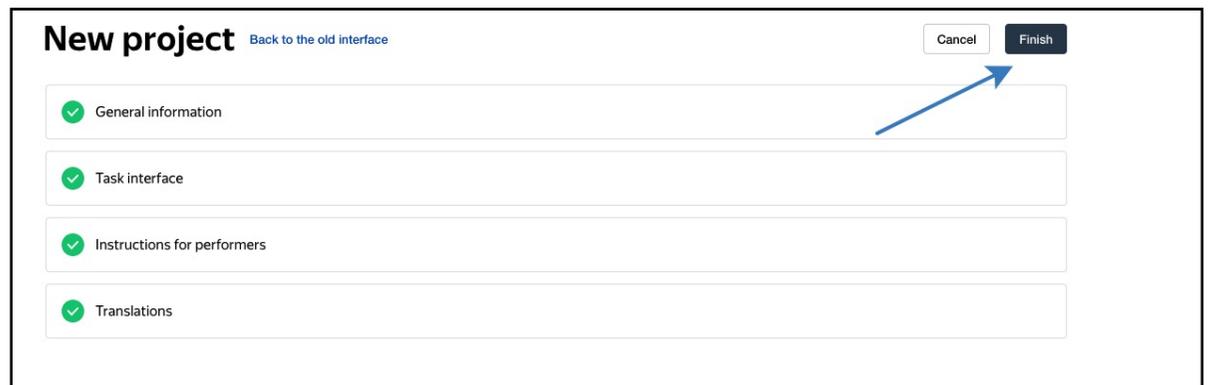
Translations

Language	Name and description for performers	Instructions for performers
✓ Source	✓	✓

Add translation

Save

10. **Save** the project.

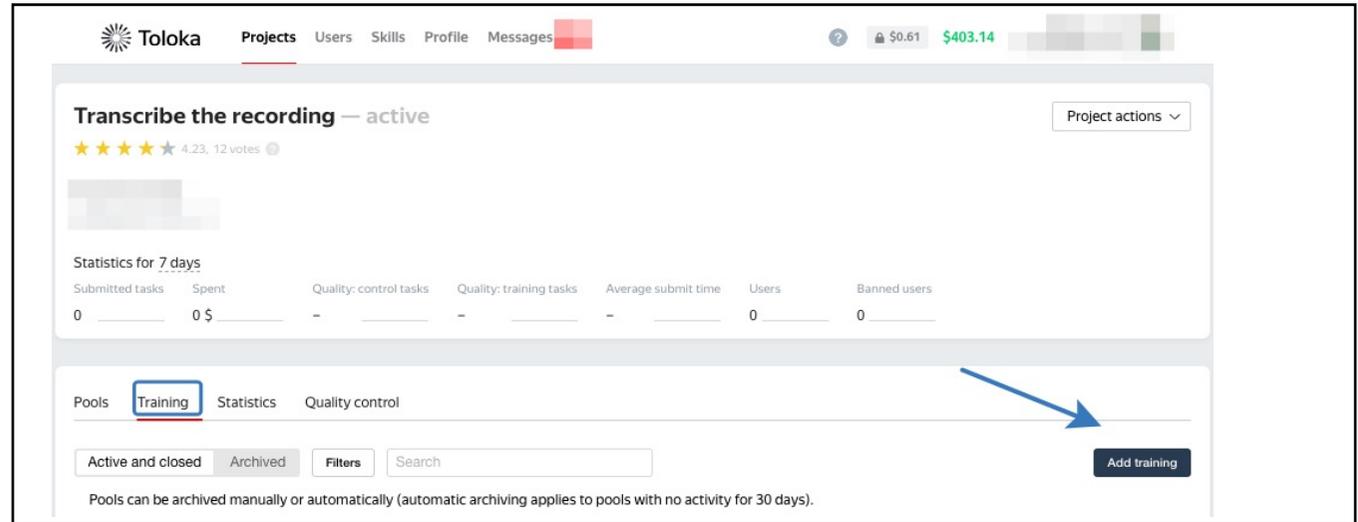


New project [Back to the old interface](#) **Cancel** **Finish**

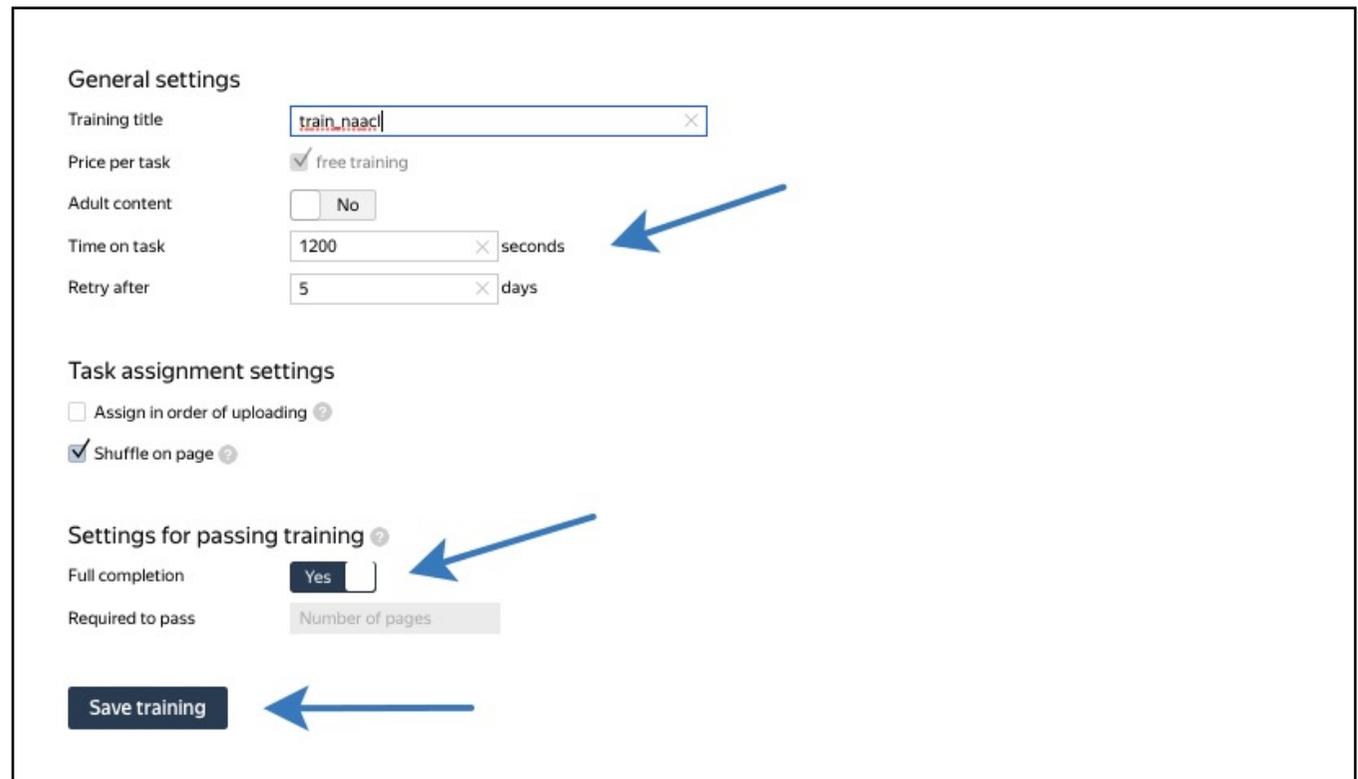
- ✓ General information
- ✓ Task interface
- ✓ Instructions for performers
- ✓ Translations

Training pool creation

1. Go to the “**Training**” tab and click the “**Add training**” button.



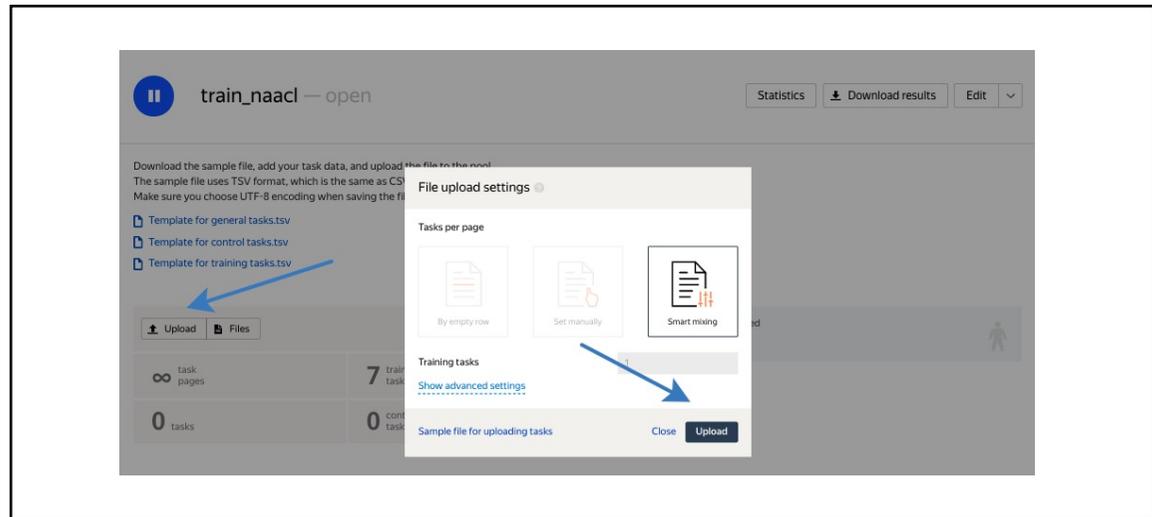
2. Keep the project instructions and fill in pool settings. Then click the “**Save training**” button.



Read more about training pools in the Toloka [Requester's Guide](#).

3. Upload training tasks to the pool. You can download the file with the tasks here:

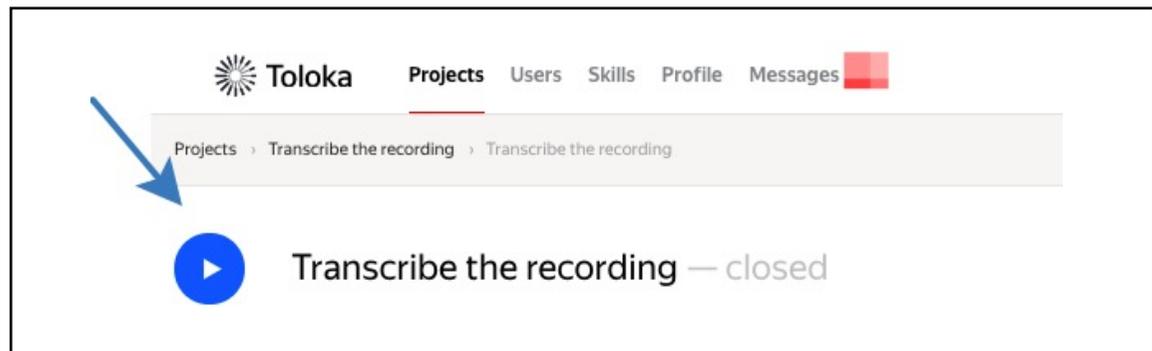
https://tlk.s3.yandex.net/naacl2021/data_train.tsv



4. If the upload went fine, you will see the task counter updated. PRESS PLAY

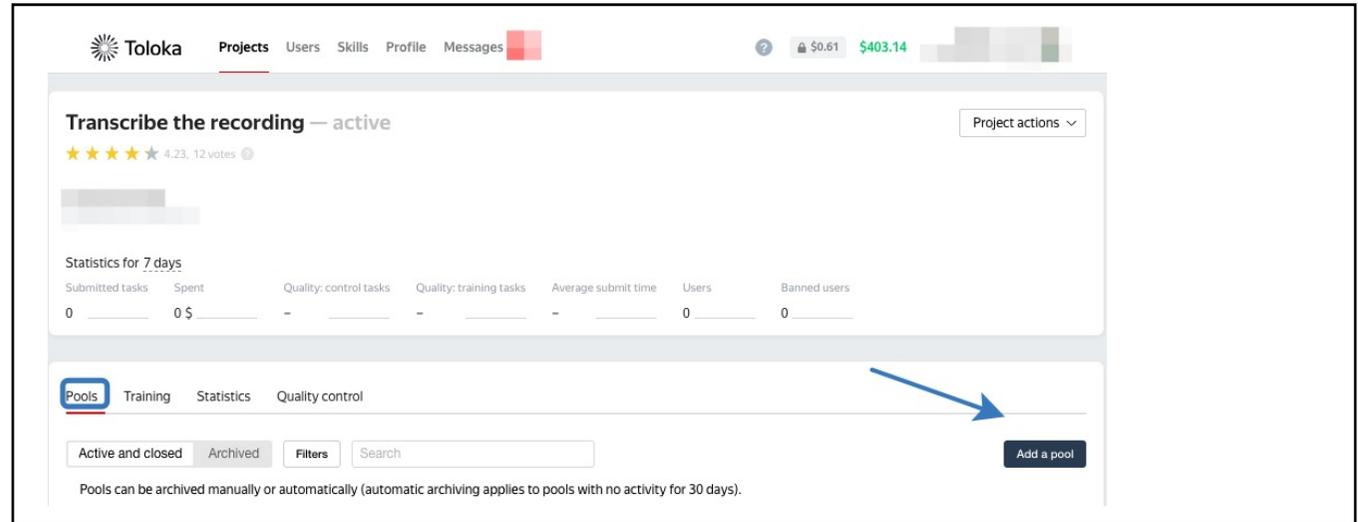


5. Run training pool by pressing large blue button.

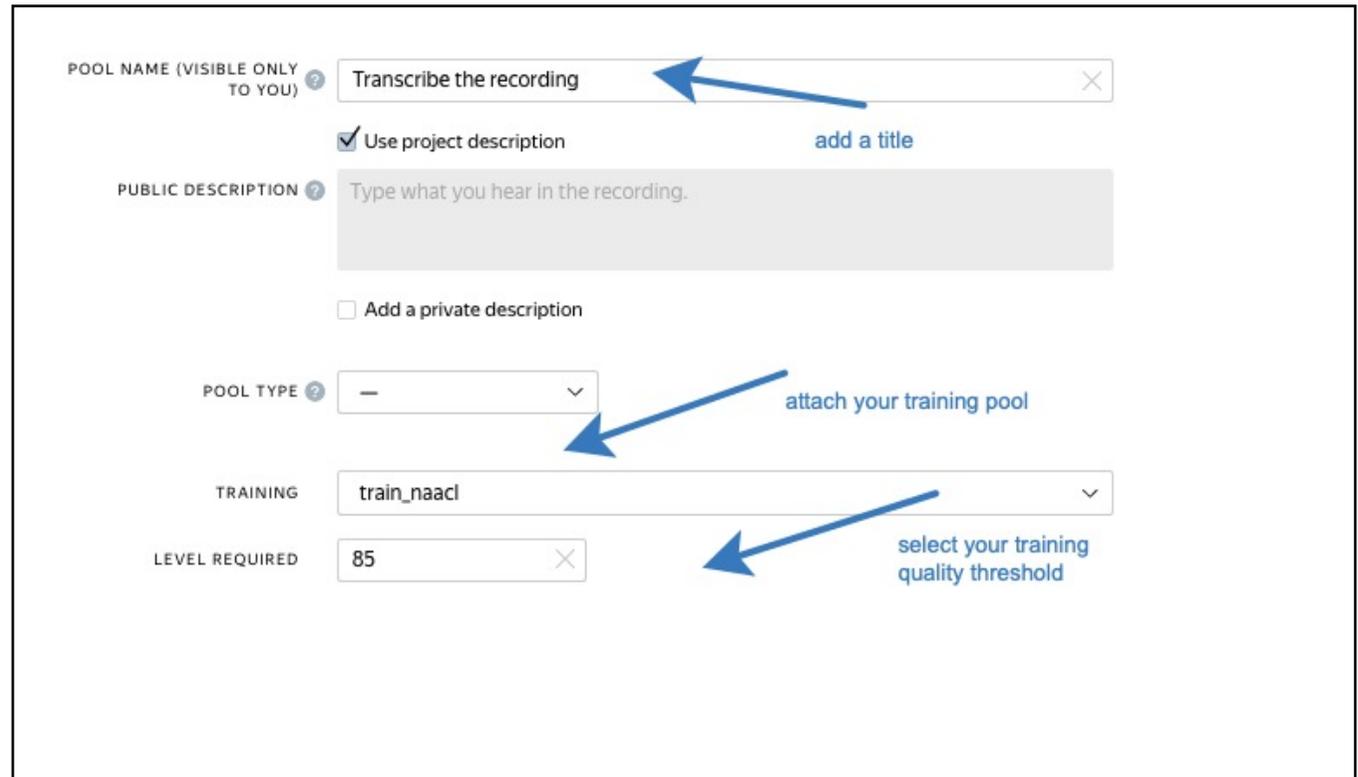


Main pool creation

1. Go to the “**Pools**” tab and click the “**Add a pool**” button.



2. Fill in the pool settings (see more on the next page).



3. Set up a price.

Price per task suite

Each task suite can have one or multiple tasks on the same page. Enter the total price for all tasks in the suite.

PRICE IN US DOLLARS FEE

[+ Dynamic pricing](#)

4. Add basic performer filters.

Performers

Filter performers who can access the task. Toloka has users from different countries, so don't forget to filter by language and region. [Learn more](#)

[Copy settings from...](#)

ADULT CONTENT Yes the task will only be visible for those who agreed to work with adult content

English is required

PERFORMER PROFILE

Languages = English

AND

CALCULATED DATA

Client = Toloka web version

OR = Toloka for mobile

the task will be displayed both for PC and mobile users

5. Set up **Control tasks** quality control rule.

Ban performers who give incorrect responses to control tasks.

Example:

This rule will be triggered when the performer completes 3 control tasks in the pool. If the performer gives at least 3 responses to the control tasks and the percentage of correct responses is less than 60%, they lose access to the project for 5 days.

If the percentage of correct responses is over 60%, the performer can pass to the next task page. The rule will be triggered after the next control task.

Optionally, add other [quality control rules](#).

6. Set up **Fast responses** quality control rule.

CONTROL TASKS ?

Recent control task responses to use

If

and

then

FAST RESPONSES ?

Recent task suites to use

Minimum time per task suite

If

then

7. Select **overlap 3** for each task.

We will aggregate the results after we finish our pools.

Overlap
Specify how many performers you want to complete each task in the pool.

OVERLAP

DYNAMIC OVERLAP Off

8. Set up a rating threshold — the percentage of top-rated users who will see your task.

Speed/quality ratio
Set additional filters to restrict performer access based on their rating in Toloka. This boosts quality but may slow down project completion because there will be fewer performers available to complete tasks. [Learn more...](#)

Top % Online Time

Specify the percentage of top-rated active users who can access tasks in the pool.

7922 Speed 4753 Quality

60% top-rated performers were selected.
The task is available to 4753 active users.

9. Fill in time allowed per task page completion.

Parameters

TIME PER TASK SUITE IN SECONDS

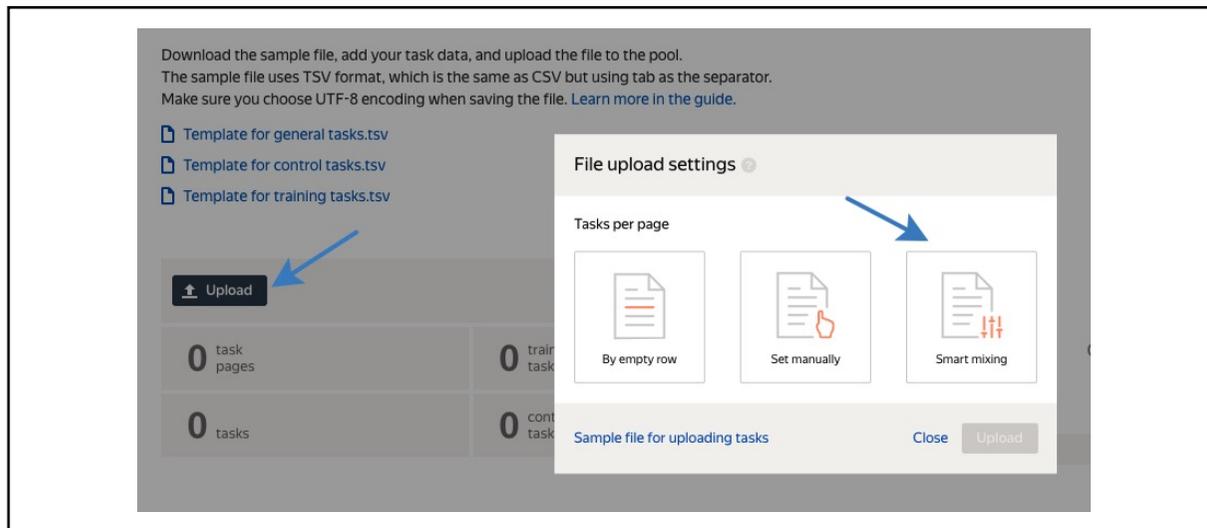
POOL CLOSING DATE

KEEP TASK ORDER No

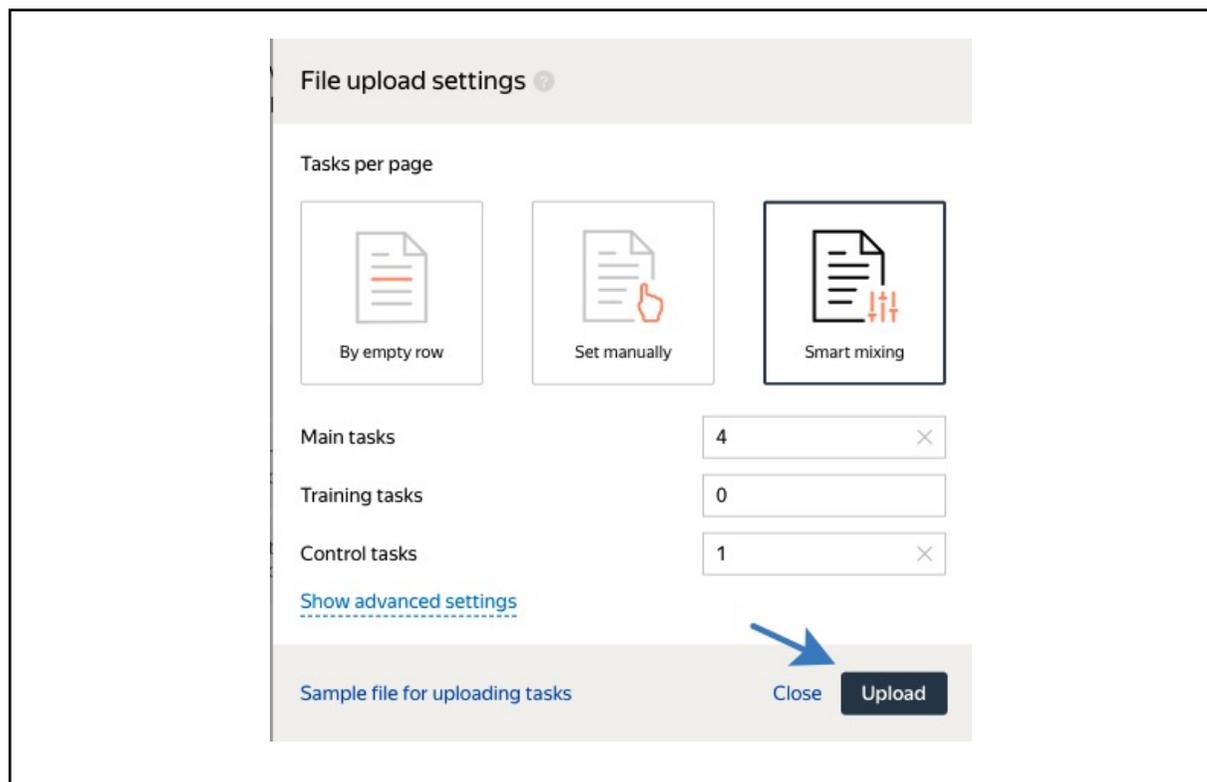
WAITING TIME FOR THE POOL TO CLOSE IN SECONDS

POOL PRIORITY WITHIN THE PROJECT

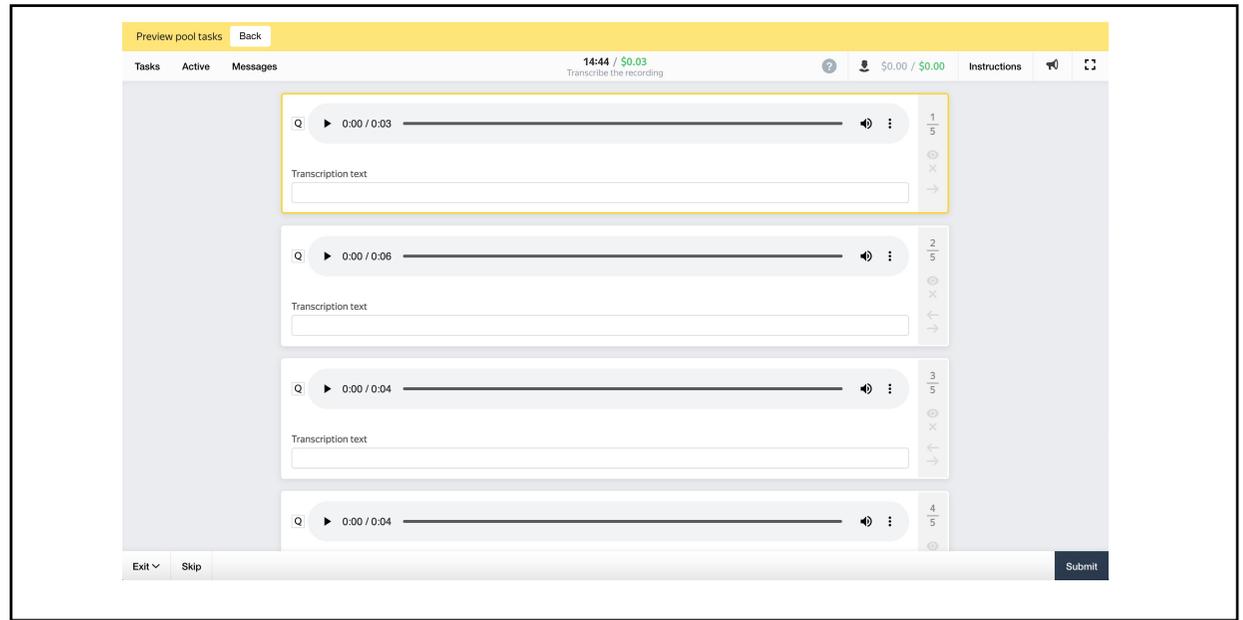
10. Click on Upload button and choose **Smart mixing** option for your task upload.



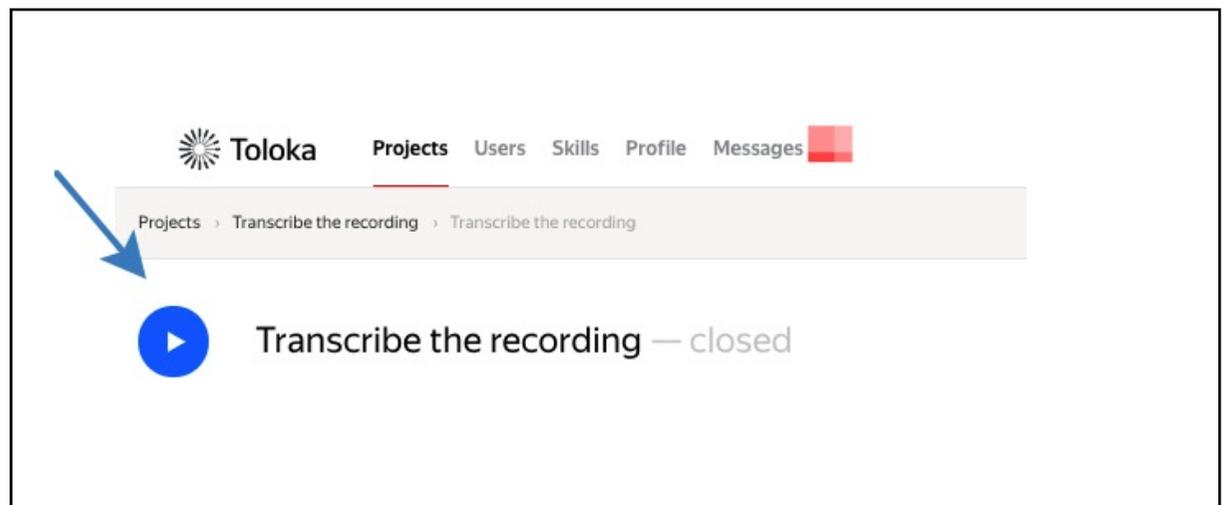
11. Leave the settings in place and upload the dataset. You can find it here:
https://tlk.s3.yandex.net/naacl2021/data_main.tsv



12. Preview the pool and make sure the tasks submit fine.

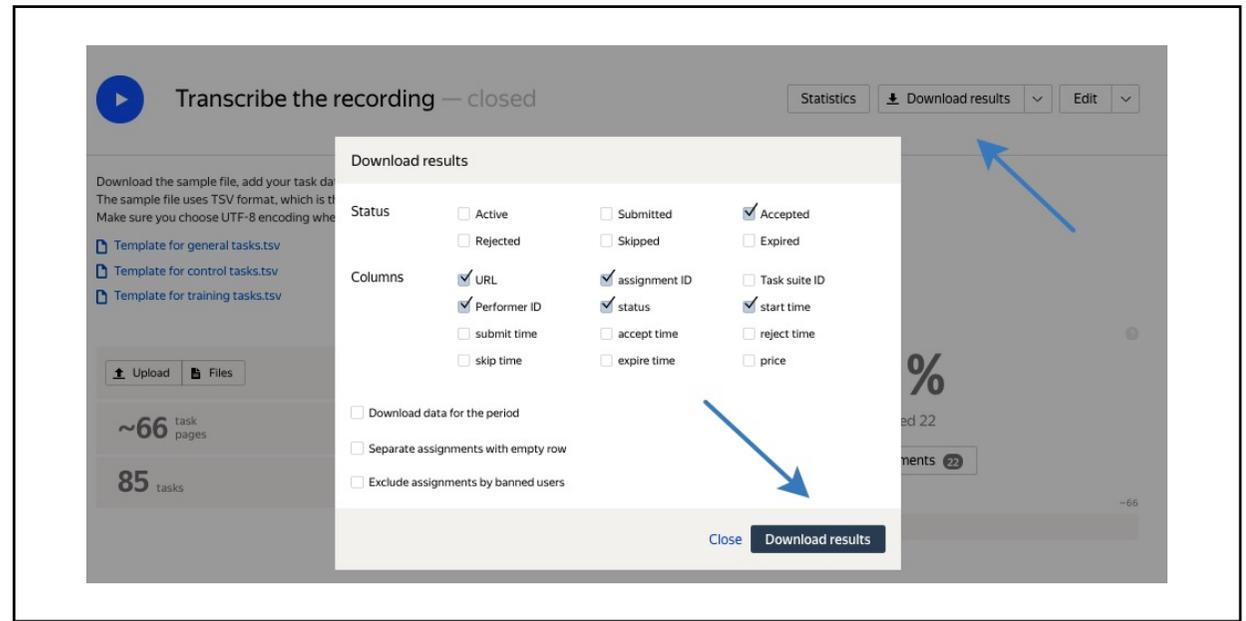


13. Run **the main** and **the training** pool.



14. Once the labeling is completed, you can download the results by clicking on the **Download Results** button.

Make sure to untick the box **“Separate assignments with empty row”**.



15. Aggregate the results by using Text Aggregation. You can [find the code in Toloka Crowd-Kit](#). Please, make sure that you download the results and name the file **assignments.tsv**

Useful resources

Tutorial slides

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

Training Pool Dataset

https://tlk.s3.yandex.net/naacl2021/data_train.tsv

Main Pool Dataset

https://tlk.s3.yandex.net/naacl2021/data_main.tsv

Text aggregation using Crowd-Kit

<https://tlk.s3.yandex.net/naacl2021/HRRASA.ipynb>

Toloka

<https://toloka.ai/>