



Text classification manual

Create a project

1. Click **Create project**.

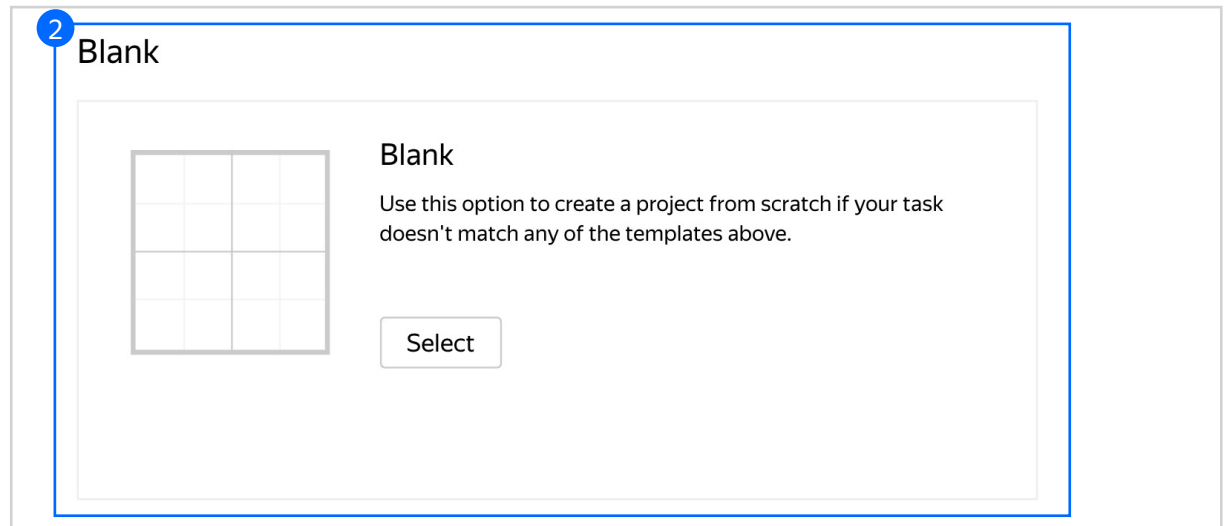
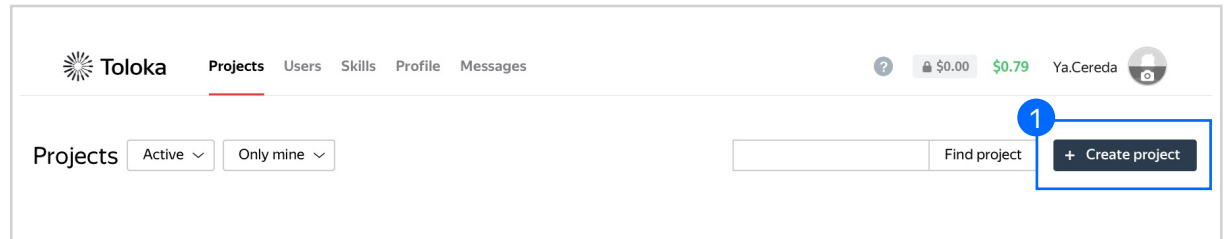
2. Choose the **Blank** template.

You can also use the **Text Classification** preset and update it according to your task specifics.

Read more about configuring the [task interface](#) in the Requester's Guide.

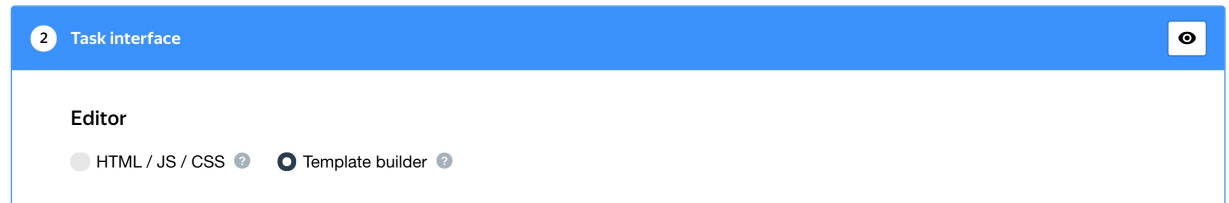
3. Enter a clear project name and description. Click **Save**.

Note: The project name and description will be visible to the performers.

This screenshot shows the 'General information' form for creating a project. The form has a blue header with a circled '1' and the text 'General information'. It contains two text input fields: 'Name to show performers *' with the placeholder text 'Is this headline clickbait?' and 'Description for performers' with the placeholder text 'Look at the a news headline and decide if it is clickbait or not.' Below these fields is a '+ Private comment' link. At the bottom, there is a checkbox labeled 'Performers will select tasks on the map' with a question mark icon. A blue box with a circled '3' highlights the 'Save' button at the bottom left of the form. To the right of the form, there is a preview box showing a five-star rating, the title 'Is this headline clickbait?', a description 'Look at the a news headline and decide if it is clickbait or not.', and pricing information: '0\$ per task' and '~0\$ per hour'.

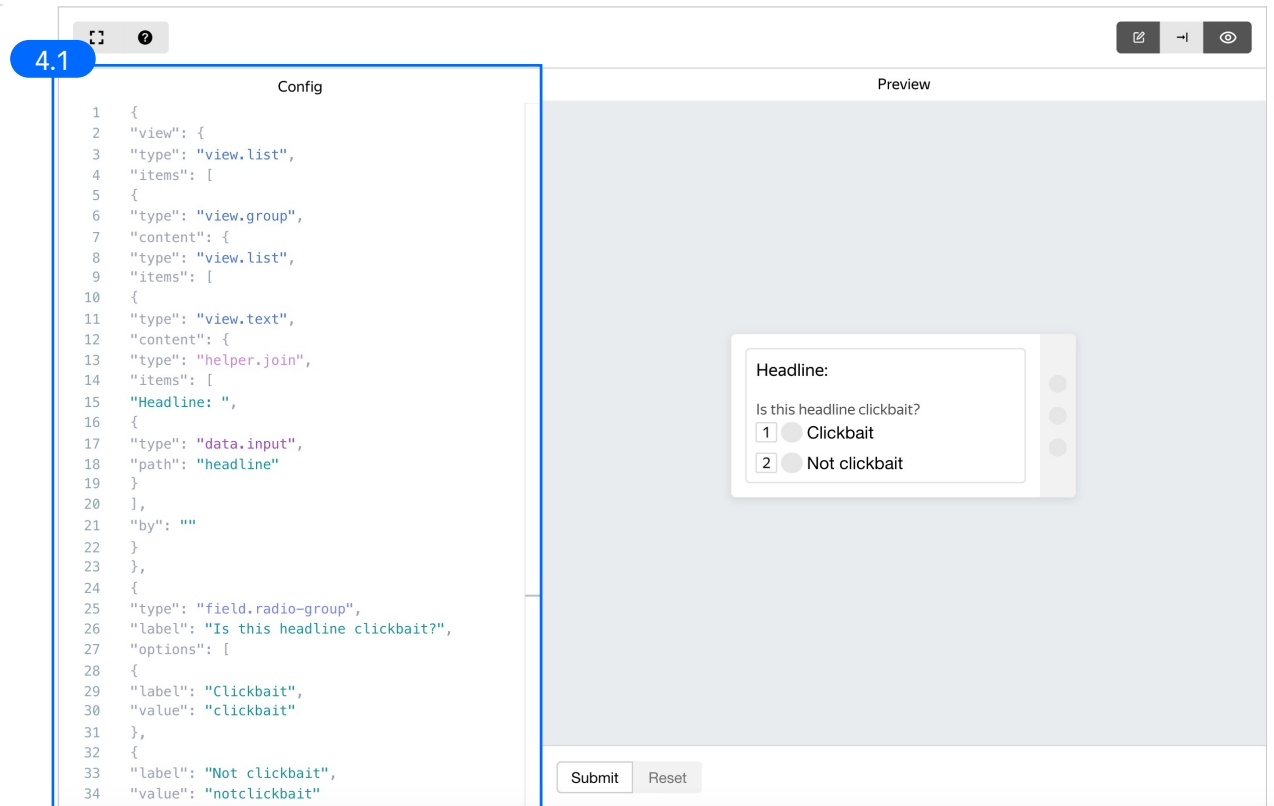
4. Update the task interface in the **Template Builder** block.

Read more about the [Template Builder](#) in the Requester's Guide.



- 4.1. Delete the existing config and paste the code provided at the end of this manual (in the appendix).

Check the [Interfaces section](#) of our Knowledge Base for more tips on interface design.



4.2. Make sure the specifications look like this:

Note: Specifications are a description of input data that will be used in a project and the output data that will be collected from the performers.

Read more about [input and output data specifications](#) in the Requester's Guide.

The screenshot shows a 'Data specification' interface with two columns: 'Input data' and 'Output data'. The 'Input data' column contains a field labeled 'headline (string)' with a dropdown arrow. The 'Output data' column contains a field labeled 'category (string)' with a dropdown arrow. Both columns have an 'Add field' button at the bottom.

5. Write comprehensive instructions. Make sure to add examples of what should be regarded as clickbait and define the possible grey areas. There is a validation rule in the template. It checks whether an answer has been chosen.

Click **Save**.

Get more tips on designing [instructions](#) in our Knowledge Base.

The screenshot shows the 'Instructions for performers' interface. It has a blue header with the title 'Instructions for performers'. Below the header, there is a text area with instructions: '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](#).' Below the text area is a rich text editor with a toolbar. The editor contains the following text: 'About the task', 'In this task you need to classify headlines on 2 categories: **Clickbait** and **Not clickbait**.', 'What is Clickbait headline?', and 'Clickbait refers to the practice of writing sensationalized or misleading headlines. **Clickbait headline** designed to make readers want to click on a hyperlink especially when the link leads to content of dubious value. Typically clickbait titles cover not very useful content, so visitors tend not to stay for too long, that's why it's bad.' At the bottom left of the editor, there is a blue circle with the number '5' and a 'Save' button.

6. Leave the **Translations** block as default and click **Save**.

4 Translations

Performers from different countries will understand the purpose of the task better if it's in their own language. Translate the task name, description, and instructions into each language that you want performers to see. Otherwise, the language is inactive. If you want to translate the task interface, you have to create it using Template Builder.

Source language
—

Translations

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

Add translation

6 Save

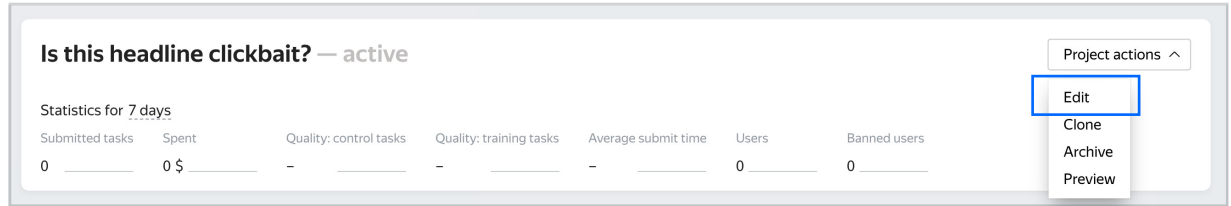
7. Click **Finish** to save the project.

Edit project

[Back to the old interface](#) Cancel **7** Finish

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

Note: To edit project parameters, click the button in the list of projects or go to **Project actions** → **Edit** on the project page.



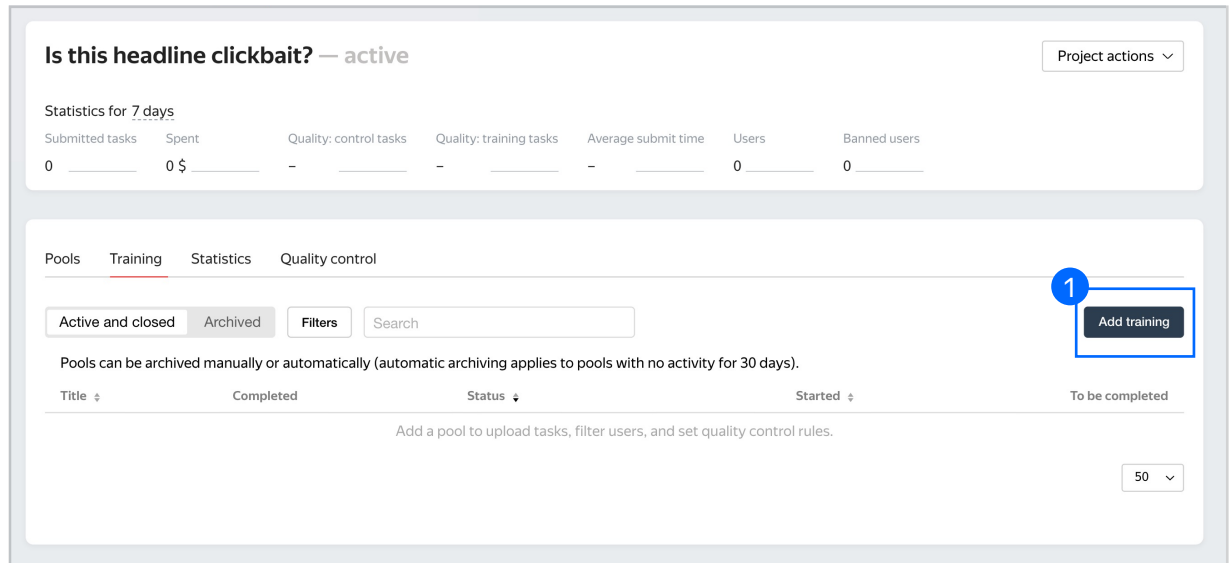
The screenshot shows a project page titled "Is this headline clickbait? — active". Below the title, there are statistics for 7 days: Submitted tasks (0), Spent (0 \$), Quality: control tasks (-), Quality: training tasks (-), Average submit time (-), Users (0), and Banned users (0). On the right side, there is a "Project actions" dropdown menu with options: Edit, Clone, Archive, and Preview. The "Edit" option is highlighted with a blue box.

Create a training pool

1. Click on the **Training** tab and then click **Add a pool**.

Training is an essential part of almost every crowdsourcing project. It allows you to select performers who have really mastered the task, and thus improve quality. Training is also a great tool for scaling your task because you can run it any time you need new performers.

Read more about [selecting performers](#) in our Knowledge Base.



The screenshot shows the "Training" tab selected in the project interface. At the top, there is a header "Is this headline clickbait? — active" and a "Project actions" dropdown menu. Below the header, there are statistics for 7 days: Submitted tasks (0), Spent (0 \$), Quality: control tasks (-), Quality: training tasks (-), Average submit time (-), Users (0), and Banned users (0). The main content area has tabs for "Pools", "Training" (selected), "Statistics", and "Quality control". Under the "Training" tab, there are buttons for "Active and closed", "Archived", "Filters", and a "Search" input field. A blue circle with the number "1" is next to the "Add training" button. Below the buttons, there is a text box that says "Pools can be archived manually or automatically (automatic archiving applies to pools with no activity for 30 days)." and a table with columns: Title, Completed, Status, Started, and To be completed. At the bottom, there is a text box that says "Add a pool to upload tasks, filter users, and set quality control rules." and a "50" dropdown menu.

2. Use the existing project instructions.

Instructions

☒ Use project instructions ?

What is Clickbait headline?

Clickbait refers to the practice of writing sensationalized or misleading headlines. **Clickbait headline** designed to make readers want to click on a hyperlink especially when the

3. Specify the training pool settings:

3.1. Click **Save training**.

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

General settings

Training title

Price per task ☒ free training

Adult content ☐ No

Time per task suite seconds

Retry after days

Task assignment settings

☒ Assign in order of uploading ?

☒ Shuffle on page ?

Settings for passing training ?

Full completion ☒ Yes

Required to pass

3.1

4. Upload training tasks to the pool.
You can download a sample file with tasks [here](#):

[Origin](#)

License: MIT

File upload settings ?

Tasks per page

By empty row Set manually Smart mixing

Training tasks 10

[Show advanced settings](#)

[Sample file for uploading tasks](#) Close Upload

If the upload is successful, you will see the updated task counter.

Upload Files Delete Edit Preview

∞ task pages	10 training tasks
0 tasks	0 control tasks

5. Run the training pool

We recommend opening the training pool along with the main pool. Otherwise Tolokers will spend their time on training but get no access to real tasks, which is frustrating. Also, do not forget to close the training pools when there are no main tasks available anymore.

5

clickbait training — closed

Statistics Download results Edit

Download the sample file, add your task data, and upload the file to the pool. The sample file uses TSV format, which is the same as CSV but using tab as the separator. Make sure you choose UTF-8 encoding when saving the file. [Learn more in the guide.](#)

[Template for general tasks.tsv](#)
[Template for control tasks.tsv](#)
[Template for training tasks.tsv](#)

Upload Files Delete Edit Preview

task pages	10 training tasks
0 tasks	0 control tasks

0 Users who completed training

Create the main pool

1. Click Add a pool.

A pool is a set of paid tasks grouped into task pages. These tasks are sent out for completion at the same time.

Note: All tasks within a pool have the same settings (price, quality control, etc.)

Is this headline clickbait? — active

Project actions

Statistics for 7 days

Submitted tasks	Spent	Quality: control tasks	Quality: training tasks	Average submit time	Users	Banned users
0	0 \$	-	-	-	0	0

Pools Training Statistics Quality control

Active and closed Archived Filters Search

1 Add a pool

Pools can be archived manually or automatically (automatic archiving applies to pools with no activity for 30 days).

Title	Priority	Progress	Status	Started	To be completed
-------	----------	----------	--------	---------	-----------------

To launch a project, you first need to add a pool, set user filters and quality control rules, and upload tasks.

50

2. Give the pool any name you find suitable. You are the only one who will see it.

The description can be either public or private. Choose the option you prefer.

POOL NAME (VISIBLE ONLY TO YOU) ? Is this headline clickbait?

☒ Use project description

PUBLIC DESCRIPTION ? Look at the news headline and decide whether it is clickbait or not.

☐ Add a private description

3. Specify [pool parameters](#):

- 3.1. Select pool type. Choose **Other**.

Read more about [pool types](#) in the Requester's Guide.

POOL TYPE ? Other

Exam

Training

Retry

☒ Other

PRICE IN US DOLLARS ?

FEE ? 0.005

- 3.2. Attach the training you created earlier and select the accuracy level that is required to reach the main pool. This means that Tolokers who got less than 90% accuracy will not see this pool.

TRAINING clickbait training

LEVEL REQUIRED, % 90

3.3. Set the price per task suite (for example, \$0.01).

Text classification tasks are normally paid as basic tasks because these tasks do not take much time.

Read more about [pricing principles](#) in our Knowledge Base.

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.

3.3

PRICE IN US DOLLARS ? 0.01 X

FEE ? 0.005

+ Dynamic pricing

3.4. [Filter](#) performers who can access the task. Choose “No” in the **Adult content** block. Click **Add filter** to choose the **Languages** and **Client** options in the list.

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](#)

3.4

ADULT CONTENT ? ☐ No

Add filter

Create a skill

3.5. Choose **Languages = English** as your first filter. This way, performers who speak English will be invited to complete this task. Then choose **Toloka web version** and **Toloka for mobile** clients.

These filters will make it possible for performers to complete your task on their computers or mobile devices.

The screenshot shows the 'FORMER PROFILE' section of the Toloka skill creation interface. At the top, there is an 'Add filter' button and a 'Create a skill' button. Below this, a filter rule is being configured. A blue circle with the number '3.5' highlights the first filter: 'Languages' set to 'English'. Below this, an 'AND' connector is shown. A second filter rule is highlighted with a blue box: 'CALCULATED DATA' set to 'Client', with two options: 'Toloka web version' and 'Toloka for mobile'. To the right, a 'Calculated data' list is visible, including 'Browser', 'Bugfix version of the operating system', 'Client' (highlighted with a blue box), 'Client application bugfix version', 'Client application major version', 'Client application versions', 'Device type', 'Operating system', 'OS major version', 'OS minor version', 'OS versions', 'Performer rating', 'Region by IP', 'Region by phone number', and 'Type of client application'.

3.6. Set up [Quality control](#). Ban performers who give incorrect responses to control tasks. Click **+ Add Quality Control Rule**

Since tasks such as these have an answer that can be used as ground truth, we can use standard quality control rules like golden sets.

Read more about [quality control principles](#) in our Knowledge Base or check out [control tasks settings](#) in the Requester's Guide.

The screenshot shows the 'Quality control' settings interface. At the top, the title 'Quality control' is followed by the text 'Add rules to get more accurate responses. All rules work independently.' Below this, there are three settings: 'NON-AUTOMATIC ACCEPTANCE' with a toggle set to 'No', 'REVIEW PERIOD IN DAYS' set to '14', and 'CAPTCHA FREQUENCY' set to 'None'. A blue circle with the number '3.6' highlights a large dashed box at the bottom containing a '+ Add Quality Control Rule' button.

3.7. Click **Control tasks**. Set the number of responses and the percentage of correct responses.

CONTROL TASKS ?

Recent control task responses to use 10 X

3.7

If number of responses ≥ 1 X +

and correct responses (%) ≤ 90 X X

then suspend in pool

add number permanently

bad quality X

3.8. Set up the up the [Fast responses rule](#).

This rule allows you to ban performers who submit tasks at a suspiciously high speed.

FAST RESPONSES ?

Recent task suites to use 5 X

3.8

Minimum time per task suite 30 X

If number of fast responses ≥ 2 X +

then suspend in pool

add number permanently

bad quality X

3.9. Overlap. This is the number of users who will complete the same task.

Set an overlap of 3 to get a more confident final label.

To understand [how this rule works](#), go to the Requester's Guide.

Overlap

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

3.9

OVERLAP ? 3

DYNAMIC OVERLAP ? ☐ Off

3.10. Optionally, specify the percentage of top-rated performers in the [Speed / quality balance](#).

Note: This can slow down pool completion.

Speed/quality balance

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.

6958

☐

90% 80% 70% 60% 50% 40% 30% 20% 10%

Speed

Quality

All users selected
The task is available to **6958** active users.

3.11. Specify the time given to complete a task suite (for example, 1200 seconds).

To understand how much time it should take to complete a task suite, try doing it yourself.

3.11

Parameters

TIME PER TASK SUITE IN SECONDS ? 1200

POOL CLOSING DATE ? 2022-07-23

KEEP TASK ORDER ? ☐ No

WAITING TIME FOR THE POOL TO CLOSE IN SECONDS ? 0

POOL PRIORITY WITHIN THE PROJECT ? 0

4. Click **Save** to save Pool parameters.

Parameters

TIME PER TASK SUITE IN SECONDS ? 1200

POOL CLOSING DATE ? 2022-07-23

KEEP TASK ORDER ? ☐ No

WAITING TIME FOR THE POOL TO CLOSE IN SECONDS ? 0

POOL PRIORITY WITHIN THE PROJECT ? 0

Cancel

4 Save

Prepare and upload a file with tasks

1. Prepare a TSV file with tasks as shown in our [example](#).
[Origin](#)
License: MIT
2. [Upload pool tasks](#) from this file.

- 2.1. Select [Smart mixing](#) in **File upload settings** and specify the number of tasks of each type per page.

We recommend putting as many tasks on one page as a performer can complete in 1 to 5 minutes. That way, performers are less likely to get tired, and they won't lose a significant amount of data if a technical issue occurs.

To learn more about [grouping tasks](#) into suites, read the Requester's Guide.

Click **Upload** again.

Note: If you changed the name of the input field, change it in the file as well.

Is this headline clickbait? — closed

Statistics Download results Edit

Download the sample file, add your task data, and upload the file to the pool. The sample file uses TSV format, which is the same as CSV but using tab as the separator. Make sure you choose UTF-8 encoding when saving the file. [Learn more in the guide.](#)

[Template for general tasks.tsv](#)
[Template for control tasks.tsv](#)
[Template for training tasks.tsv](#)

2 Upload

0 task pages	0 training tasks
0 tasks	0 control tasks

0 % Completed 0

File upload settings ?

Tasks per page

By empty row Set manually Smart mixing

Main tasks 4

Training tasks 0

Control tasks 1

[Show advanced settings](#)

Sample file for uploading tasks Close Upload

2.1

3. [Create control tasks.](#)

Click **Edit** → **Create control tasks**.

Note: Control tasks are tasks that already contain the correct response. They are used for checking the quality of responses from performers. The performer's response is compared to the response you provided. If they match, it means the performer answered correctly.

▶ Is this headline clickbait? — closed

StatisticsDownload resultsEdit

Download the sample file, add your task data, and upload the file to the pool.
The sample file uses TSV format, which is the same as CSV but using tab as the separator.
Make sure you choose UTF-8 encoding when saving the file. [Learn more in the guide.](#)

[Template for general tasks.tsv](#)
[Template for control tasks.tsv](#)
[Template for training tasks.tsv](#)

3

UploadFilesDeleteEdit

0 task pages

0 training tasks

120 tasks

0 control tasks

0 %
Completed 0

00

Edit tasks

Use main tasks as a starting point to create control tasks or training tasks.

Control tasks are for checking the quality of responses from performers. They contain correct responses to compare with actual responses.

Training tasks are for teaching performers how to complete tasks. They contain correct responses and hints.

[Learn more](#)

Main 120Control tasks 0Training tasks 0

Create control tasks

Create training tasks

Download

ID ↕	Overlap ↕	Responses from performers ↕	Last updated ↕
...38757659	3	0	06/07/2021 13:48:32
...38757642	3	0	06/07/2021 13:48:32

3.1. Enter correct responses for your control tasks. Check the result output field, which compares the user's response to the control task. Look at the headline, select the response, click **Save** and go to **next**. Repeat until you have 20 control tasks.

You need to create 20 tasks, because in small pools, control tasks should account for 10–20% of all tasks.

To learn more about [creating control tasks](#), go to the Requester's Guide.

Projects > Is this headline clickbait? > Is this headline clickbait? > Uploaded tasks > Edit tasks

Create control task

1. Enter correct responses

2. Select the fields to use

Field	Value
category	

Headline: What's The Worst Thing To Happen To You As A Single Person At A Wedding

Is this headline clickbait?

1: ☒ Clickbait

2: ☐ Not clickbait

3.1

Save and go to next

Create 1 more

Distribution of correct responses for control tasks

Create control tasks to see charts of response distribution.

Tip. Make sure to include different variations of correct responses in equal amounts.

Go to **Uploaded tasks** → **Control tasks** → **Distribution of correct responses for control tasks** tab.

Projects > Is this headline clickbait? > Is this headline clickbait? > Uploaded tasks

Edit tasks

Use main tasks as a starting point to create control tasks or training tasks.

Control tasks are for checking the quality of responses from performers. They contain correct responses to compare with actual responses.

Training tasks are for teaching performers how to complete tasks. They contain correct responses and hints.

[Learn more](#)

Main 100 **Control tasks 20** Training tasks 0

Create from main tasks

Download

ID	Overlap	Responses from performers	Correct responses, %	Last update
...576b192b	∞	0		07/23/2021 11:35
...576b1910	∞	0		07/23/2021 11:35
...576b1930	∞	0		07/23/2021 11:34

Distribution of correct responses for control tasks

category

50% clickbait

50% notclickbait

3.2. To check the number of control tasks, go to the **Pool** page.

The screenshot shows the Toloka Pool interface for a task titled "Is this headline clickbait?". At the top, there's a play button icon and the task title. To the right are buttons for "Statistics", "Download results", and "Edit". Below this, there's a section with instructions: "Download the sample file, add your task data, and upload the file to the pool. The sample file uses TSV format, which is the same as CSV but using tab as the separator. Make sure you choose UTF-8 encoding when saving the file. [Learn more in the guide.](#)" Below the instructions are three links: "Template for general tasks.tsv", "Template for control tasks.tsv", and "Template for training tasks.tsv". In the center, there's a summary box with buttons "Upload", "Files", "Delete", "Edit", and "Preview". It shows "~75 task pages", "0 training tasks", "100 tasks", and "20 control tasks". A blue circle with the number "3.2" is overlaid on the "20 control tasks" text. To the right, there's a large "0%" progress indicator and a "Completed 0" label. At the bottom, there's a progress bar from 0 to -75.

4. Preview the pool.


Note: Remember that the tasks will be completed by actual Tolokers. Double check that everything is correct with your project configuration.

The screenshot shows the "Preview pool tasks" page. At the top, there's a yellow header with "Preview pool tasks" and a "Back" button. Below the header, there's a navigation bar with "Tasks", "Active", and "Messages". The main content area shows five task cards. Each card has a headline, a question "Is this headline clickbait?", and two radio button options: "Clickbait" and "Not clickbait". The cards are numbered 1 through 5. The first card is highlighted with a yellow border. The second card has a "2/5" indicator. The third card has a "3/5" indicator. The fourth card has a "4/5" indicator. The fifth card has a "5/5" indicator. At the top right of the main content area, there's a timer "19:42 / \$0.01" and a balance "\$0.00 / \$0.00". There are also buttons for "?", "Instructions", and a speaker icon.

5. Start the pool.

Make sure you start the training pool before you start the main pool.

5



Is this headline clickbait? — closed

StatisticsDownload resultsEdit

Download the sample file, add your task data, and upload the file to the pool.
The sample file uses TSV format, which is the same as CSV but using tab as the separator.
Make sure you choose UTF-8 encoding when saving the file. [Learn more in the guide.](#)

- Template for general tasks.tsv
- Template for control tasks.tsv
- Template for training tasks.tsv

UploadFilesDeleteEditPreview

~75 task pages	0 training tasks
100 tasks	20 control tasks

0 %

Completed 0

0~75

Receiving responses

1. Wait until the pool is completed.
Refresh the pool page to check progress.

2. Click the **Download results** button and run aggregation using the [Dawid-Skene model](#).

We use this aggregation model because our questions are of comparable difficulty, and we don't have many control tasks.

Read more about the [Dawid-Skene model](#) in the Requester's Guide or get at an overview of different [aggregation models](#) our Knowledge Base.

3. Go to the list of operations and wait until aggregation is complete.
Refresh the **Operations** page to check progress.

Note: Aggregation takes from 5 to 20 minutes. During this time, you can start configuring your next project. Refresh the Operations page to check progress.

The screenshot shows a project titled "Is this headline clickbait? — closed". It features a "Download results" button and a "View operations" button. A dropdown menu is open under "View operations", showing "Dawid-Skene aggregation model" and "Aggregation by skill". The project status is "100 % Completed 75, accepted 75". A green progress bar is at the bottom.

Download the sample file, add your task data, and upload the file to the pool.
The sample file uses TSV format, which is the same as CSV but using tab as the separator.
Make sure you choose UTF-8 encoding when saving the file. [Learn more in the guide.](#)

[Template for general tasks.tsv](#)
[Template for control tasks.tsv](#)
[Template for training tasks.tsv](#)

Upload	Files	Edit	Preview
75 task pages	0 training tasks		
100 tasks	20 control tasks		

100 %
Completed 75, accepted 75
[View assignments](#)

The screenshot shows the "Operations" page with a table of operations. The first operation is "Dawid-Skene aggregation model" with a progress bar at 0% and status "Running".

Operations

PROJECT: Is this headline clickbait? POOL: Is this headline clickbait?

Id	Type	Started	Completion time	Progress	Status	Files
5b15d...	Dawid-Skene aggregation model	07/23/2021 12:41:28	-	0%	Running	-

- When aggregation is complete, download the TSV file with the results.

Operations

PROJECT

Is this headline clickbait? ▾

POOL

Is this headline clickbait? ▾

Id ▴ ▾	Type ▴ ▾	Started ▴ ▾	Completion time ▴ ▾	Progress	Status ▴ ▾	Files
5b15d...	Dawid-Skene aggregation model	07/23/2021 12:41:28 PM	07/23/2021 12:43:45 PM	<div></div> 100%	Success	<div>4</div> <div>Download</div>

10 ▾

Appendix

Interface code

Step 4.1.

```
{
  "view": {
    "type": "view.list",
    "items": [
      {
        "type": "view.group",
        "content": {
          "type": "view.list",
          "items": [
            {
              "type": "view.text",
              "content": {
                "type": "helper.join",
                "items": [
                  "Headline: ",
                  {
                    "type": "data.input",
                    "path": "headline"
                  }
                ],
                "by": ""
              }
            },
            {
              "type": "field.radio-group",
              "label": "Is this headline clickbait?",
              "options": [
                {
                  "label": "Clickbait",
                  "value": "clickbait"
                },
                {
                  "label": "Not clickbait",
                  "value": "notclickbait"
                }
              ],
              "data": {
                "type": "data.output",
                "path": "category"
              },
              "validation": {
                "type": "condition.required",
                "hint": "you need to select one answer"
              }
            }
          ]
        }
      },
      {
        "type": "plugin.toloka",
        "layout": {
          "kind": "scroll",
          "taskWidth": 300
        },
        "1": {
          "type": "action.set",
          "data": {
            "type": "data.output",
            "path": "category"
          },
          "payload": "clickbait"
        },
        "2": {
          "type": "action.set",
          "data": {
            "type": "data.output",
            "path": "category"
          },
          "payload": "notclickbait"
        },
        "type": "plugin.hotkeys"
      }
    ]
  }
}
```