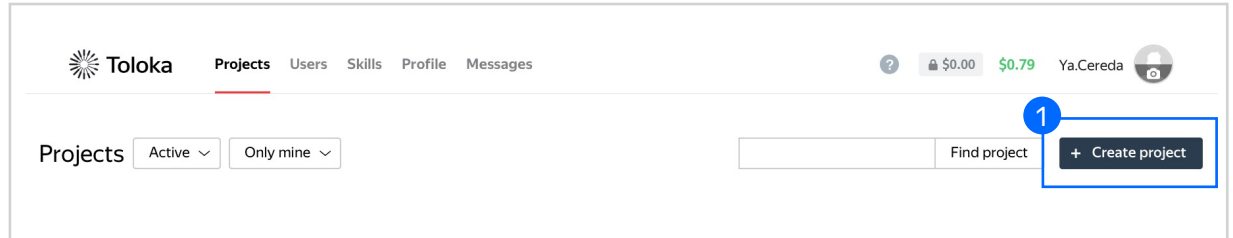Toloka

Sentiment analysis manual
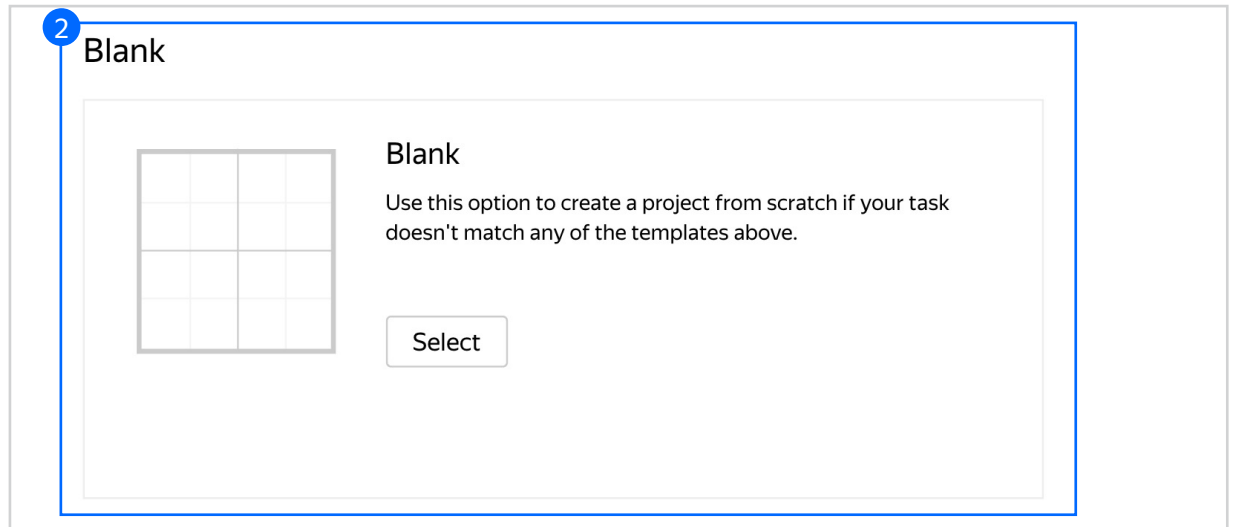
Create a project

1. Click **Create project.**

2. Choose the **Blank** template.
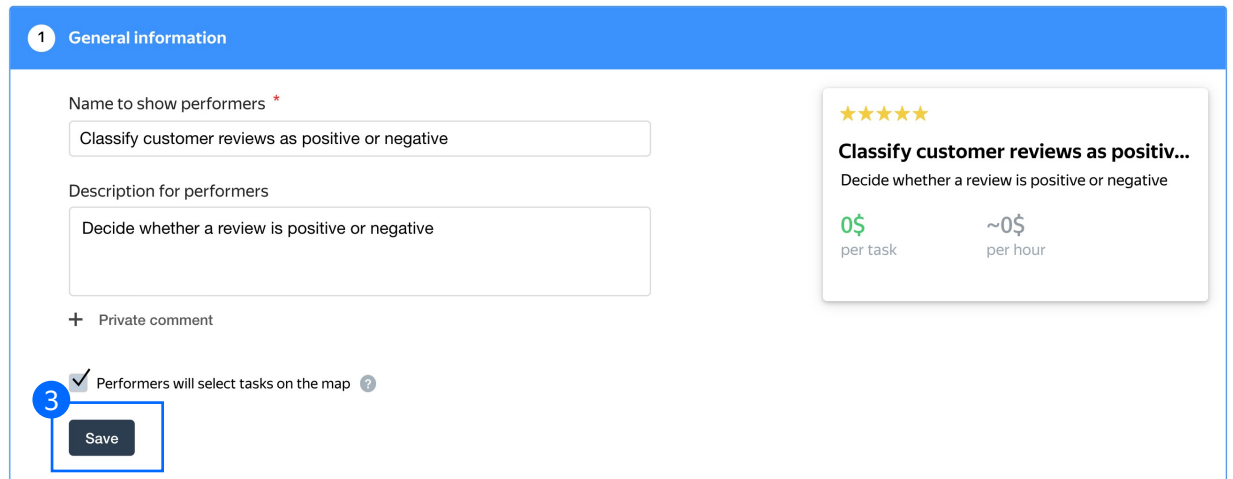
   You can also use the **Sentiment Analysis** preset and update it according to your task specifics.

   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.

※ Toloka   **Projects** Users Skills Profile Messages   ? 🔒 $0.00 $0.79 Ya.Cereda 📷

1

Projects [Active ⌄] [Only mine ⌄]   [   ] Find project   [+ Create project]

2 Blank

Blank

Use this option to create a project from scratch if your task doesn't match any of the templates above.

[Select]

1 General information

Name to show performers *

[Classify customer reviews as positive or negative]

Description for performers

[Decide whether a review is positive or negative]

+ Private comment

☑ Performers will select tasks on the map ?

3 [Save]

★★★★★

Classify customer reviews as positiv...

Decide whether a review is positive or negative

0$ ~0$
per task   per hour

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

Read more about the [Template Builder](#) in the Requester's Guide.
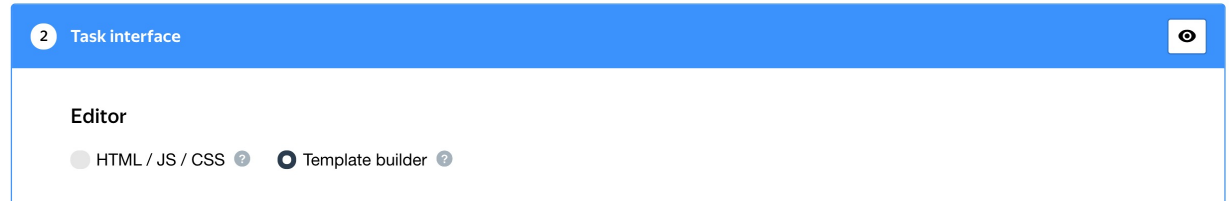
4.1. Delete the existing config and paste the code provided at the end of this manual (in the appendix).

In the **Preview** section, you can see how the template will work. The performer looks at a review and decides if it is positive or negative.

The validation rule in the template checks whether an answer has been chosen.

Check the [Interfaces section](#) of our Knowledge Base for more tips on interface design.
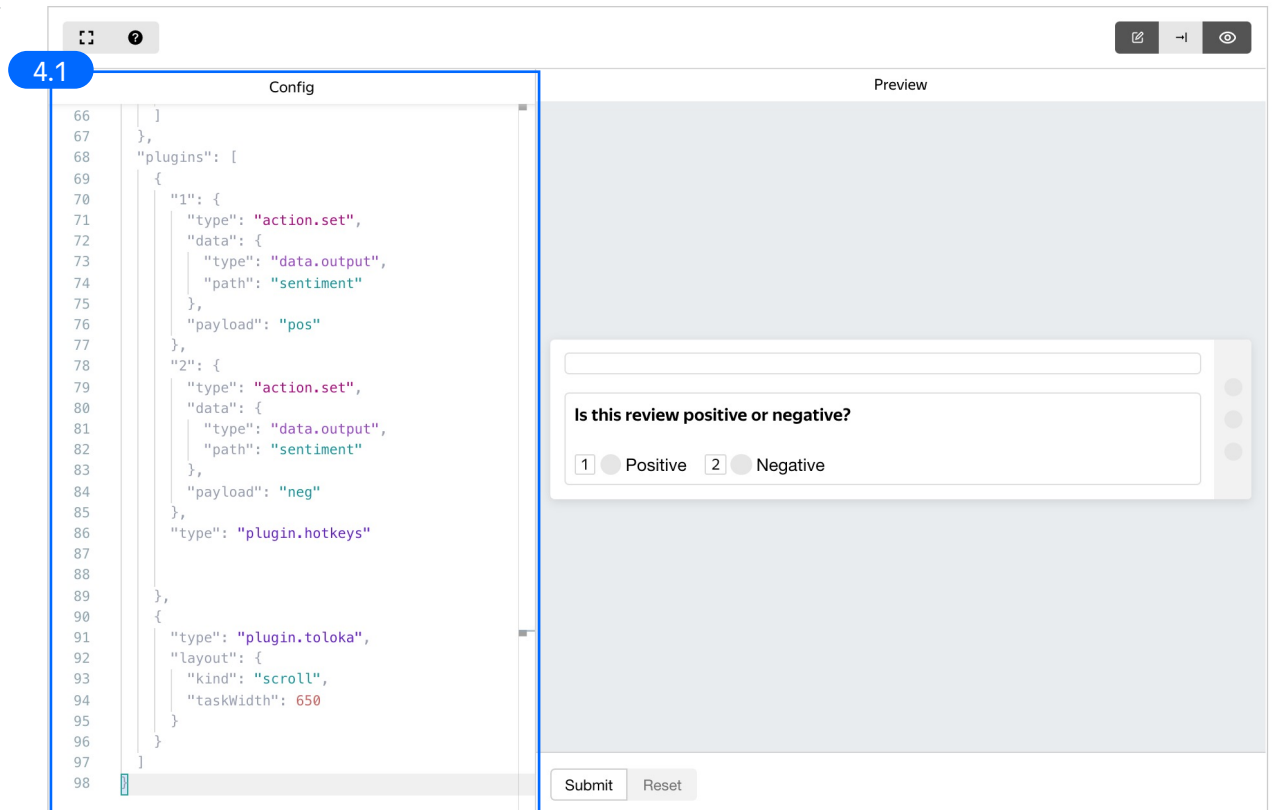
4.2. Make sure the specifications look like this:

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.

Data specification ?

Input data	Output data	‹ ›
review (string) ●	**sentiment** (string) ●	

Add field   Add field

5. Write comprehensive instructions. Be sure to add examples for ambiguous cases.

Click **Save.**

Get more tips on designing instructions in our Knowledge Base.

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.

In the task you will have to read customer reviews and define whether they are positive or negative

- **Positive**. Choose this option if the review reflects a customer's first-hand **good** experience with the product recommending to purchase it. For your convenience, you can also use the short-cut by pressing "1".

- **Negative**. Choose this option if the review reflects a customer's first-hand **poor** experience with the product recommending **not** to purchase it. For your convenience, you can also use the short-cut by pressing "2".

5

Save

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

4 Translations

ℹ 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

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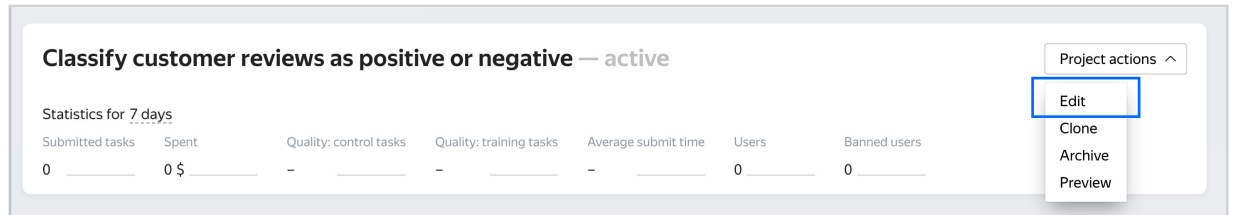
7. Click **Finish** to save the project.

Edit project

Back to the old interface   Cancel   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.

Classify customer reviews as positive or negative — active

Project actions ∧

Edit
Clone
Archive
Preview

Statistics for 7 days

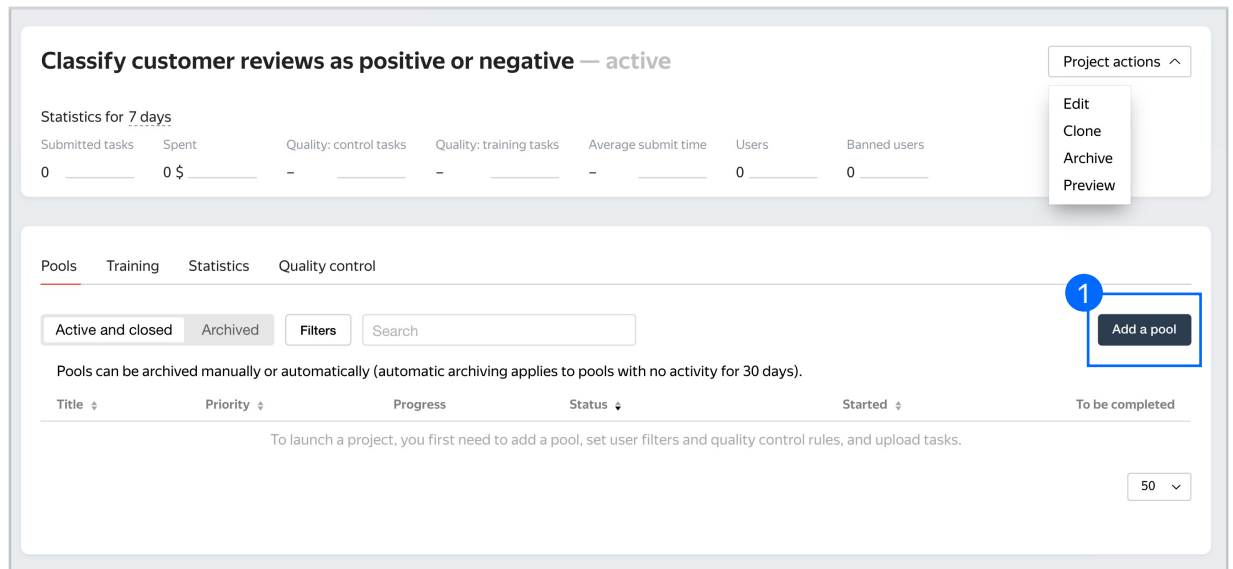| Submitted tasks	Spent	Quality: control tasks	Quality: training tasks	Average submit time	Users	Banned users
0	0 $	–	–	–	0	0

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.)

Classify customer reviews as positive or negative — active

Project actions ∧

Edit
Clone
Archive
Preview

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

2

POOL NAME (VISIBLE ONLY TO YOU) ❓ | Classify customer reviews as positive or negative ✕

☑ Use project description

PUBLIC DESCRIPTION ❓ | Decide whether a review is positive or negative

☐ 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
3.1 ✓ Other

PRICE IN US DOLLARS ❓ | FEE ❓ 0.005

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

   Binary 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.

3.2 PRICE IN US DOLLARS ❓ | 0.01 ✕ | FEE ❓ 0.005

➕ Dynamic pricing

3.3. [Filter](#) performers who can access the task. Choose "**No**" in the **Adult content** block. Click **Add filter** to choose the **Languages** and **Client** option 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

3.3

ADULT CONTENT ⓘ | No

Add filter ⌄ | Create a skill

3.4. 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.

Add filter ⌄ | Create a skill

3.4 FORMER PROFILE

Languages ⌄ | = | English ✕ 🗑 +

AND

CALCULATED DATA

Client ⌄ | = | Toloka web version ⌄ 🗑 +

OR | = | Toloka for mobile ⌄ 🗑

Calculated data
Browser
Bugfix version of the operating system.
Client
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
Type of client application

⁂ Toloka

**3.5.** Create a skill that will be used as a filter.

A skill is a performer's attribute, which is a number between 0 and 100. Skills usually reflect performer quality. They can also be set manually as a custom characteristic.

Read more about skills in the Requester's Guide.
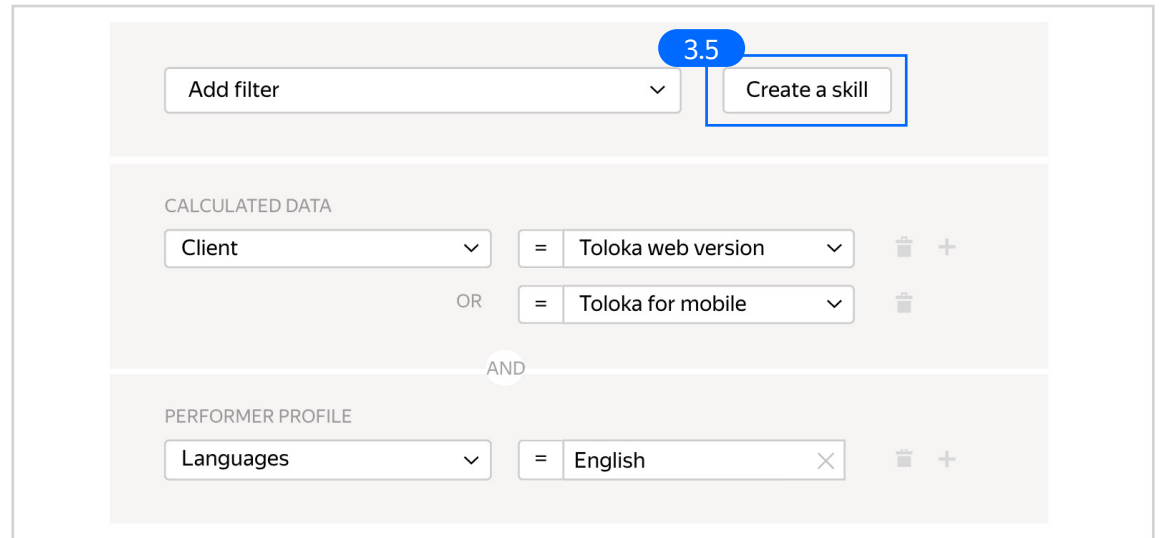
3.6. Enter a title and choose whether the skill will be public.

If your project will run for a long time, public skills that reflect quality are a good idea because they allow performers to track their own progress.
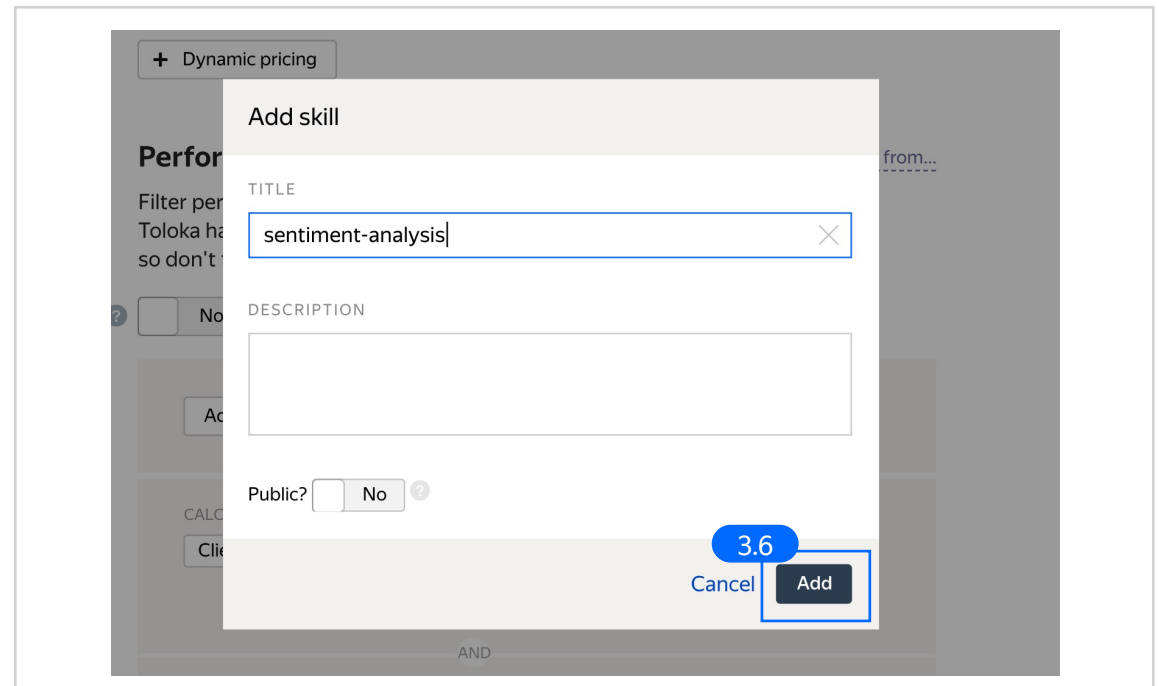
3.7. Add a filter based on the skill you just created. It will grant access only to those performers who are either new to the task (and have no quality attribute yet) or have decent quality based on the results of golden sets.

See section 3.11. for details on how to set up quality calculation.

You can use the same skill as a filter in other pools, keeping a project-based quality history for your performers.

CALCULATED DATA

Client | = | Toloka web version

OR | = | Toloka for mobile

AND

PERFORMER PROFILE

Languages | = | English

AND

3.7

SKILLS

sentiment-analysis | = | Missing

OR | ≥ | 90

3.8. Set up Quality control. Ban performers who give incorrect responses to control tasks. Click **+ Add Quality Control Rule**

Since the answers to such tasks 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 task settings in the Requester's Guide.

Quality control

Add rules to get more accurate responses.
All rules work independently.

NON-AUTOMATIC ACCEPTANCE ❓ | No

REVIEW PERIOD IN DAYS | 21

CAPTCHA FREQUENCY ❓ | Low

3.8

⊕ Add Quality Control Rule

3.9. Set **Captchas** to appear at low frequency. Suspend performers who aren't accurate enough – set **"suspend in the pool for 1 day."**

Captcha is a great tool to check whether performers are paying attention.

Read more about different [quality control rules](#) in our Knowledge Base.

CAPTCHA ⑦

Recent captchas to use   10 ✕

3.9
If   number of responses ⌄   ≥   4   ✕   +
and   correct responses (%) ⌄   <   75 ✕   +
then   suspend ⌄   in pool ⌄
        1 ✕   day ⌄
        captcha   ✕

3.10. Set up the [Fast responses rule.](#) It allows you to ban performers who submit tasks at a suspiciously high speed.

FAST RESPONSES ⑦

Recent task suites to use   items

Minimum time per task suite   20 ✕

3.10
If   number of responses ⌄   >   4   ✕   +
and   number of fast responses ⌄ >   2   ✕   +
then   suspend ⌄   in pool ⌄
        1 ✕   day ⌄
        fast responces   ✕

3.11. Set up individual quality calculation based on control tasks. Each performer's results will be recorded in the skill you created earlier.

Read more about configurating this rule in our Requester's Guide.

CONTROL TASKS ?
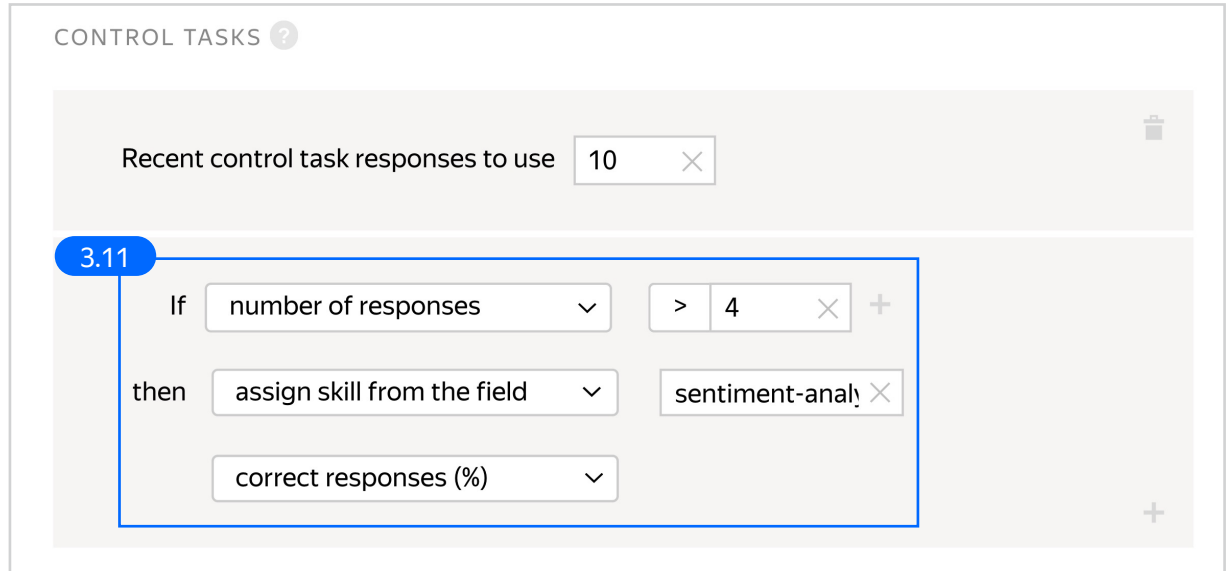
Recent control task responses to use   10   ✕

3.11

If   number of responses   ⌄   >   4   ✕   +

then   assign skill from the field   ⌄   sentiment-analy   ✕

correct responses (%)   ⌄

3.12. Set up the Submitted responses rule. Use it to add variety to answers so that they aren't biased toward only a few productive performers.

SUBMITTED RESPONSES ?
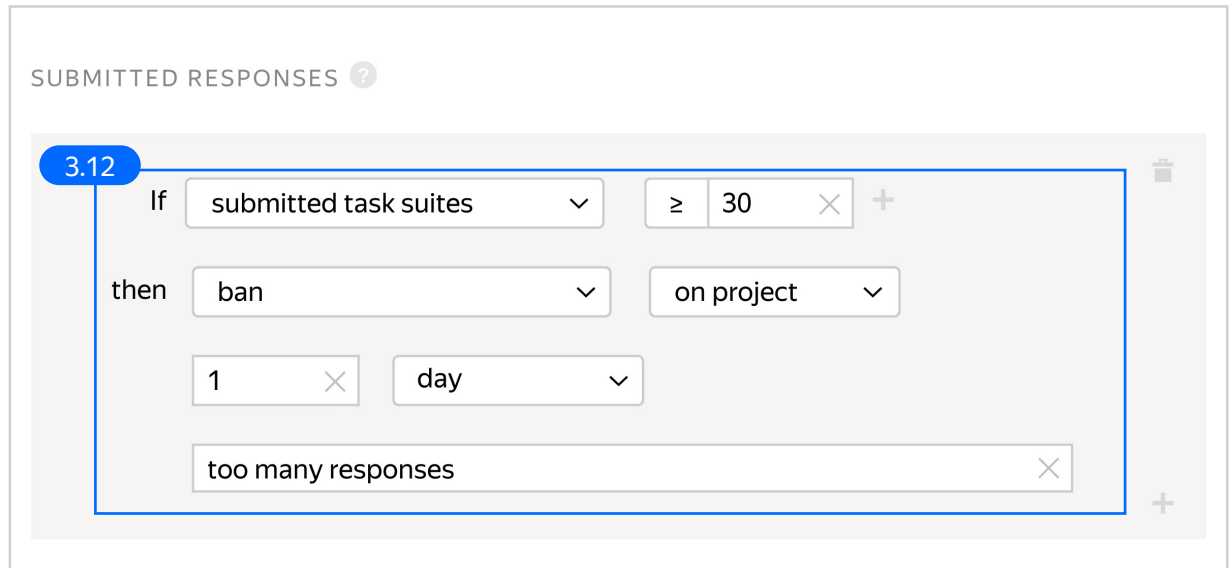
3.12

If   submitted task suites   ⌄   ≥   30   ✕   +

then   ban   ⌄   on project   ⌄

1   ✕   day   ⌄

too many responses   ✕

3.13. Overlap. This is the number of performers 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.

3.13

OVERLAP ? | 3 ✕

DYNAMIC OVERLAP ? | Off

3.14. 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.

3.14

4636 👤 ─────●─────────────────── 3708 👤

Speed | All 90% 80% 70% 60% 50% 40% 30% 20% 10% | Quality

80% top-rated performers were selected.
The task is available to **3708** active users.

3.15. 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.

Parameters

3.15

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

4

Cancel   Save

Prepare and upload a file with tasks

1. Prepare a TSV file with tasks as shown in our [example](#). [Origin](#)
License: CC BY-NC-SA 4.0

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.

Classify customer reviews as positive or negative — 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

0 | 0

File upload settings ?

Tasks per page

| By empty row | Set manually | Smart mixing |

Main tasks | 9 ✕
Training tasks | 0
Control tasks | 1 ✕

[Show advanced settings](#)

[Sample file for uploading tasks](#) | Close | **2.1** Upload

3. **Create control tasks**.

Click **Edit → Create control tasks.**

Classify customer reviews as positive or negative — 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

0 task pages
0 training tasks
230 tasks
0 control tasks

0 %
Completed 0

0

0

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.

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 **230** | Control tasks **0** | Training tasks **0**

Create control tasks | Create training tasks | Download

ID ⇕	Overlap ⇕	Responses from performers ⇕	Last updated ⇕
...2884603f	3	0	07/24/2021 12:11:13 AM
...2884604f	3	0	07/24/2021 12:11:13 AM
...2884604d	3	0	07/24/2021 12:11:13 AM

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, and click **Save and go to next**.

Repeat until you have enough control tasks. Create 30 tasks this way.

You need to create 30 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.
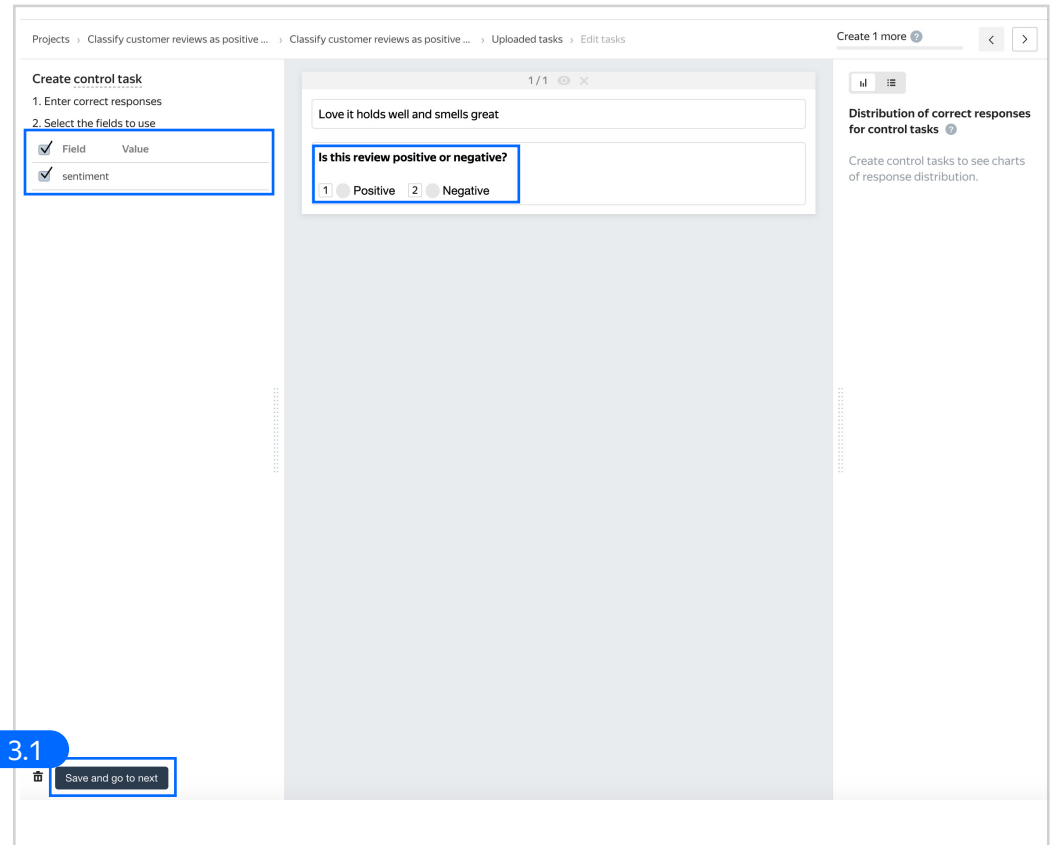
Tip. Make sure to include different variations of correct responses in equal amounts.

Go to **Uploaded tasks → Control tasks → Distribution of correct responses for control tasks** tab.
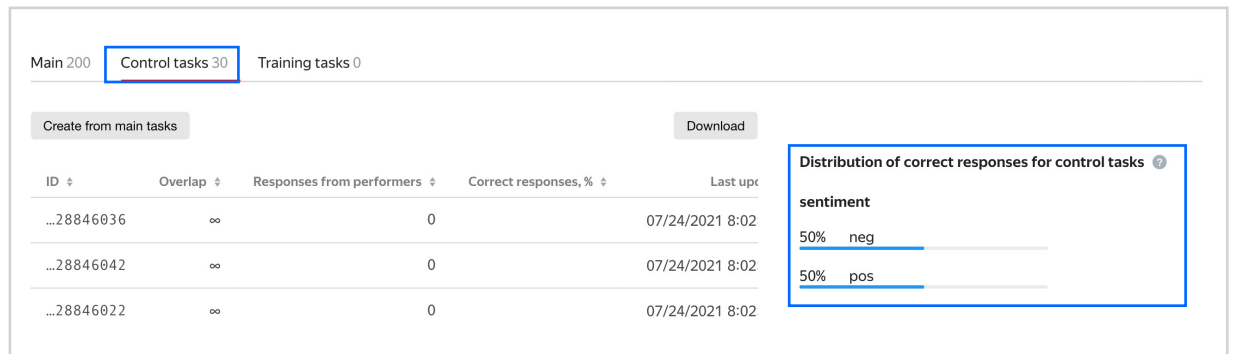
3.2. To check the number of control tasks, go to the **Pool page.**

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.

5. Start the pool.

5

Classify customer reviews as positive or negative —
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

69 task pages
200 tasks

0 training tasks
30 control tasks

0 %
Completed 0

0   69

Receiving responses

1. **Wait until the pool is completed.** Refresh the pool 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.

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.

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

Operations

PROJECT
Classify customer reviews as positiv... ⌄

POOL
Classify customer reviews as positiv... ⌄

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	100%	Success	**4** Download

10 ⌄

Appendix

Interface code
Step 4.1.

```
{
"view": {
"type": "view.list",
"items": [
{
"type": "view.group",
"content": {
"type": "view.text",
"content": {
"type": "data.input",
"path": "review"
}
}
},
{
"type": "view.group",
"label": "Is this review positive or negative?",
"content": {
"type": "view.list",
"items": [
{
"type": "view.text"
},
{
"type": "view.list",
"direction": "horizontal",
"items": [
{
"type": "field.radio-group",
"options": [
{
"label": "Positive",
"value": "pos"
}
],
"data": {
"type": "data.output",
"path": "sentiment"
}
},
{
"type": "field.radio-group",
"options": [
{
"label": "Negative",
"value": "neg"
}
],
"data": {
"type": "data.output",
"path": "sentiment"
}
}
],
"validation": {
"type": "condition.required",
"data": {
"type": "data.output",
"path": "sentiment"
}
}
}
]
}
}
]
},
"plugins": [
{
"1": {
"type": "action.set",
"data": {
"type": "data.output",
"path": "sentiment"
},
"payload": "pos"
},
"2": {
"type": "action.set",
"data": {
"type": "data.output",
"path": "sentiment"
},
"payload": "neg"
},
"type": "plugin.hotkeys"
},
{
"type": "plugin.toloka",
"layout": {
"kind": "scroll",
"taskWidth": 650
}
}
]
}
```

Toloka