

**Individual Terms and Conditions for the “VLDB’21 Crowd Science Challenge”
(hereinafter referred to as the “Contest”)**

1. The Contest is organized by **Yandex Services AG** (the “Organizer”), a company registered under the jurisdiction of Switzerland having its registered address at **Werftstrasse 4, 6005 Luzern**, registration number CHE-229.170.782, represented by **Dr. Herbert Trachsler**, **Authorized representative**.

Participants in the Contest (the “Participant”) must be a capable individual acting on their own behalf who has reached the age of eighteen and has signed up for the participation as described in clause 6 of these conditions. The name, e-mail address and other required data must be entered accurately. If a participant no longer wishes to take part in the Contest, he or she can inform the Organizer by e-mail to the following address: vldb21crowdchallenge@crowdscience.ai

Participant represents and warrants that neither it, nor any of their respective Affiliates, directors, employees, or other related parties is not listed on any sanctions list by the United States, European Union (“EU”), Switzerland, the United Nations Security Council, or any other relevant government bodies which prohibit or may prohibit Participant's participation in the Contest ("Sanctions"). The Organizer reserves the right to disqualify without any justification and to exclude from the Contest with immediate effect, or to forfeit the Prizes of any participant, should there be any Sanctions or update of the Sanctions against Participant, or should Yandex reasonably determine that it cannot perform its obligations under this Contest due to Sanctions-related prohibitions. Participant shall hold Yandex harmless against all liabilities.

Employees of the Organizer and / or affiliated companies, other persons involved in organizing the Contest, as well as their family members cannot participate in the Contest.

2. The goal of the Contest is to support the research on crowdsourcing and related areas.

3. The Contest is hosted on the Internet: <https://competitions.codalab.org/competitions/30835>

4. Key dates of the Contest (end of day UTC):

4.1 Training data release: April 15.

4.2 Competition starts (test data release): May 5.

4.3 Promo codes to use on the Toloka platform are distributed: May 25

4.3 Submission deadline: June 18.

5. By participating in the Contest, the Participant confirms that they have read and understood the terms of the Contest (the “Terms and Conditions”) available at [https://github.com/Toloka/VLDB2021 Crowd Science Challenge/blob/main/Terms.pdf](https://github.com/Toloka/VLDB2021_Crowd_Science_Challenge/blob/main/Terms.pdf)

Participation in this Contest implies the acceptance of these Terms and Conditions. The Organizer reserves the right to modify these without giving reasons. Changes will be communicated on the website and will not be sent to participants by e-mail.

The Contest is governed by Swiss law. For any dispute that may arise, the place of jurisdiction is Lucerne.

By taking part in the Contest, the Participant grants the Organizer and its affiliates the non-exclusive and unrestricted right to use, reproduce, publish and make available the predictions made on the test set and train set without remuneration, even after the Contest has ended.

The Organizer and its affiliates undertake to handle the personal data collected within the framework of the Contest with due care. The data collected in connection with participation in the Contest (e.g., first name, surname, nickname, e-mail address, bank details for Winners) will be collected, processed and used exclusively by the Organizer for the purposes of organizing the Contest, including the placing on a public leaderboard and awarding of Prizes, as well as in the manner specified in point 8.1 below.

The Organizer reserves the right to disqualify without any justification and to exclude from the Contest with immediate effect, or to forfeit the Prizes of any participant who manipulates or attempts to manipulate the Contest and/or violates the conditions of participation and/or unfairly attempts to influence the Contest.

The Organizer reserves the right to interrupt or terminate the Contest early without prior notice and without stating reasons, in particular if the normal course of the Contest cannot be guaranteed for technical or legal reasons. In such cases, participants may not assert any claims against the Organizer.

The Organizer shall not be liable to the Participant for any loss or damage in any form whatsoever arising from or in connection with the Contest, the processing of data and the Prize.

Participation in the contest is free of charge and does not give rise to any obligation to purchase.

6. The nature of the Contest and selection of Winners.

6.1. To participate in the Contest, an individual must create a profile on <https://competitions.codalab.org> and click "Participate" on the Contest page <https://competitions.codalab.org/competitions/30835>. Participants can participate in the Contest individually or in Teams of up to 5 members. To form a Team, Participants need to create a separate account for the Team and participate in the Contest using this account. Each participant can either participate individually or be a member of one Team only. In what follows, we use the term Team to refer to both a group of Participants or a single Participant (Team of 1).

6.2. Throughout the duration of the Contest, Teams will work with data provided by the Organizer. The data consists of two parts: test set and train set. Below we discuss the construction of both of these sets:

Train Set Each entry to the Train set consists of the ground truth text and 7 annotations collected on the Toloka crowdsourcing platform. The annotations are obtained by first creating an audio recording of the text and then asking performers on the crowdsourcing platform to annotate this recording.

Test Set The Test set is constructed in the same manner as the Train set, but the ground truth texts are not released to the participants (they are available to the Organizer).

6.3 The goal of each Team is to develop a method to reconstruct the ground truth texts from the provided annotations. By the end of the Contest, Teams need to submit their predictions of the ground truth texts which will be evaluated on the test set.

6.4 To compare submissions made by different teams, we use the WER metric (word error rate, see definition on Wikipedia https://en.wikipedia.org/wiki/Word_error_rate). Specifically, for each audio in the test set, we compute the WER distance between the ground truth text and the prediction submitted by the team. We then average these distances across all audios in the test set and the resulting quantity becomes the score of the team. The smaller the score, the better.

6.5 Throughout the Contest, Teams can submit their preliminary predictions to be evaluated on a public part of the test set. These preliminary evaluations will be maintained on a public leaderboard visible to all Participants of the Contest.

6.6 The Winning Teams of the Contest are determined by evaluating the quality of final predictions made by each Team against the private part of the test set. Teams that achieve the top 3 best scores on these metrics become the Winners of the Contest. In case of a tie between two or more team, the teams that made their submission earlier will be ranked higher.

6.7 At the middle of the Contest (25th of May 2021) the Organizer will take the top 20 teams from the public leaderboard available on the Contest page and award these Teams with promo codes to use on the Toloka crowdsourcing platform. Specifically, the Organizer will grant these Teams one promo code of a value of USD \$100 to collect data on the Toloka crowdsourcing platform. The promo codes are distributed over email. For the sake of clarity, the total quantity of promo codes is 20. This promo codes can be used at the discretion of the Teams, however, we recommend to use these promo codes to collect data on the Toloka crowdsourcing platform with a goal of improving the solutions of the Teams. Any legal recourse is excluded. Promocodes are not paid-out in cash. Promocodes may not be transferred to third parties.

7. Prizes.

7.1 The Winning Teams receive cash prizes in the following amount:

- 1st place. \$3,000
- 2nd place. \$2,000
- 3rd place. \$1,000

If a Team consists of multiple Participants, the price will be divided equally across all members of the Team.

7.2. There are a total of 3 Prizes to be won in the Contest.

7.3. The Prizes are distributed by the Organizer.

8. Announcement of the Winners and distribution of the Prize.

8.1. The Organizer announces the Winners by emailing them using the email address associated with their account. The list of Winners is also disclosed on the Contest website at the following link <https://crowdsience.ai/challenges/vldb21>

8.2. The Prizes are distributed in equal installments proportionally to the number of Winning teams' members. Winning teams will be asked to fill out a short form to specify the credentials of the members.

8.3. Winners are responsible to pay all applicable taxes.