



Задача по автоматический оценке ответа иностранного гражданина



Контекст

Проводится экзамен по русскому языку для иностранных граждан.

В ходе экзамена экзаменатор задаёт устно вопросы тестируемому. Тестируемый также устно на них отвечает. Весь диалог записывается в аудиофайл. После ответа экзаменатор выставляет балл тестируемому за каждый ответ.

Экзамен для одного человека состоит из 4 вопросов, которые оцениваются независимо.



Входные данные

Участникам чемпионата передаётся csv-файл. Файл содержит данные по 9 798 вопросам.

Состав данных:

1. Порядковый номер вопроса в экзамене одно человека (от 1 до 4)
2. Текст вопроса (допускается разметка html)
3. Ссылка на картинку из вопроса (есть не для всех вопросов)
4. Оценка, которую экзаменатор поставил за ответ
5. Транскрибация ответа из аудиозаписи. Транскрибация содержит как реплики тестируемого, так и экзаменатора.
6. Ссылка на аудиофайл оригинальной записи ответа.



Требования к результату

Участникам чемпионата необходимо разработать инструмент, который сможет наиболее точно ставить оценки тестируемым без участия экзаменатора.

Инструмент должен принимать на вход аналогичный csv-файл, но с незаполненным столбцом оценки экзаменатора. На выходе необходимо отдать такой же csv-файл, но уже с заполненным столбцом оценок экзаменатора.

Участники могут использовать любые доступные инструменты и нейросети, включая LLM. Однако запрещено использовать проприетарные иностранные LLM. Все инструмент должны работать либо на территории РФ, либо они должны предоставлять возможность развернуть их на своём сервере.

Участники сами решают, какие именно данные брать в качестве исходных: аудио или уже готовую текстовую транскрибацию.



Требования к результату

Примерные критерии, по которым оцениваются ответы тестируемого:

1. Ошибки в отдельных словах и единичные несогласованности фраз не считаются ошибкой.
2. Акцент не оценивается, если речь позволяет понять смысл сказанного.
3. Тестируемый должен выполнить коммуникативную задачу (должен ответить на вопрос или добиться ответа на свой вопрос).
4. Предложения тестируемого должны быть преимущественно полными.

Диапазоны оценок:

1. Вопрос 1 - от 0 до 1 балла
2. Вопрос 2 - от 0 до 2 баллов
3. Вопрос 3 - от 0 до 1 балла
4. Вопрос 4 - от 0 до 2 баллов

Итоговый продукт необходимо передать в двух вариантах:

1. Ссылка на репозиторий с исходным кодом
2. Ссылка на развёрнутое веб-приложение, в которое можно загрузить файл с закрытым датасетом. После загрузки файла должен проводиться анализ и выдача файла с автоматически проставленными оценками. Требований к внешнему виду никаких нет. Закрытый датасет не публикуется и не передаётся участникам. Размер закрытого датасета: 2476 записей.



Оценка точности

Проверка точности созданного продукта будет проводиться на закрытой тестовой выборке (закрытая выборка не передаётся участникам).

Точность равна среднему арифметическому от разницы по модулю между оценкой экзаменатора и оценкой продукта по каждому вопросу.