

Технический регламент
конкурса отдельных заданий
«ПРО//ОЦЕНКИ»

2021г.

Оглавление

1. Общая информация	3
2. Термины и определения	3
3. Участники конкурса и условия участия	6
4. Циклы и этапы конкурса	6
5. Конкурсная задача и оценивание решений Участников	7
6. Отборочный этап	7
7. Этап финальных испытаний	8
8. Судейская коллегия	9
9. Права Оргкомитета	9
10. Участникам запрещается	10
11. Обеспечение безопасности и дополнительные условия	11

1. Общая информация

- 1.1. Настоящий Технический регламент определяет требования к процедуре проведения конкурса отдельных заданий «ПРО//ОЦЕНКИ» в целях реализации Национальной технологической инициативы (далее Конкурс), обеспечению и другим техническим параметрам Конкурса.
- 1.2. Технический регламент публикуется на официальном Сайте Конкурса (далее - Сайте) <http://ai.upgreat.one/>
- 1.3. Настоящий Технический регламент является документом, детализирующим Конкурсное задание Конкурса, опубликованное на Сайте Конкурса, по регламентам проведения испытаний и процедуре определения результата и победителя Конкурса. Настоящий документ не противоречит Конкурсному заданию, но решает задачу детализации Конкурсного задания до технического уровня; технические детали, указанные в настоящем документе, являются необходимым и достаточным описанием задачи Конкурса. Технический регламент определяет требования к процедуре проведения Испытаний Конкурса, Программным комплексам Участников, процедуре контроля и алгоритму определения уровня разработок команд.
- 1.4. Основная цель настоящего документа: обеспечение единого набора документации и регламента участия в Конкурсе, обеспечивающих прозрачность и объективность как для Участников, так и для сторонних наблюдателей.

2. Термины и определения

- 2.1. В настоящем Техническом регламенте помимо терминов, перечисленных ниже, используются в том числе, термины в значении, определенном в п. 1.10 Конкурсного задания Конкурса
- 2.2. **Алгоритм разметки** – алгоритм на основе искусственного интеллекта, используемый в Программном комплексе Участника для преобразования файла с неразмеченным текстом эссе в файл с размеченным текстом того же эссе.
- 2.3. **Алгоритмическая разметка** – разметка текста эссе, сгенерированная алгоритмом разметки в Программном комплексе Участника.
- 2.4. **БД, База данных эссе** – совокупность всех текстов эссе на русском языке, как размеченных, так и неразмеченных, представленных в виде файлов, организованная по определённым правилам.
- 2.5. **Выборка данных, датасет (ДС)** – набор текстовых файлов эссе, размеченных или неразмеченных.
- 2.6. **Выборка для обучения алгоритмов разметки, Обучающая выборка, Размеченный эталонный открытый ДС** – набор размеченных текстовых файлов эссе,

предоставляемый Участникам Конкурса на равных условиях, используемый Участниками для обучения своих алгоритмов разметки.

- 2.7. **Выборка для тестирования алгоритма разметки, Тестовая выборка, Размеченный эталонный закрытый ДС** – набор неразмеченных текстовых файлов эссе, предоставляемый Участникам Конкурса на равных условиях для независимого оценивания точности алгоритмов разметки в рамках Отборочного этапа. В Системе проверки решений используется выборка размеченных файлов тех же эссе, доступная только Оператору Конкурса и недоступная Участникам.
- 2.8. **Выборка для финального испытания, Финальная выборка, Финальный ДС** – набор неразмеченных текстовых файлов эссе, предоставляемый Участникам Конкурса на равных условиях в рамках этапа Финальных испытаний. Разметка финальной выборки экспертами осуществляется после того, как от всех Участников будут получены алгоритмические разметки всех файлов финальной выборки.
- 2.9. **Испытания, Финальные испытания** – ограниченный по времени период, начинающийся в назначенный Оргкомитетом день, когда Участники проводят с помощью разработанных ими Программных комплексов потоковую разметку текстов, содержащихся в файлах Выборки финального испытания.
- 2.10. **Конкурс** – открытый конкурс отдельного задания конкурс «ПРО//ОЦЕНКИ», направленный на решение научно-технологической проблемы в области искусственного интеллекта. Входит в Систему конкурсов поддержки команд.
- 2.11. **КЗ, Конкурсное задание** – основной документ, определяющий цели, задачи и порядок проведения Конкурса. Конкурсное задание утверждается конкурсной комиссией технологических конкурсов в целях реализации Национальной технологической инициативы.
- 2.12. **КО, Классификатор ошибок** – перечень типов ошибок и смысловых блоков которые могут использоваться в экспертных и алгоритмических разметках текстов эссе. Классификатор ошибок и язык разметки описаны в Приложении 3 к данному Техническому регламенту.
- 2.13. **Лидерборд** — список Участников, ранжированный исходя из методики оценивания
- 2.14. **ЛК, Личный кабинет** – раздел пользовательского веб-интерфейса Платформы, в котором пользователь может получить доступ к данным, аналитической информации и сервисам Платформы в соответствии с его ролью в Конкурсе и правами доступа.
- 2.15. **Оценка за эссе** – оценка в баллах, вычисляемая на основании разметки текста эссе согласно правилам проверки эссе данного типа (например, оценка за мини сочинение по обществознанию в рамках ЕГЭ). Оценка может быть вычислена как по экспертной разметке, так и по алгоритмической.
- 2.16. **Ошибка** – фрагмент текста эссе, обладающий обязательными параметрами начала, конца и типа ошибки согласно КО, а также необязательными параметрами подтипа ошибки согласно КО, комментария (для детализации подтипа ошибки без привязки к

тексту эссе), пояснения (для объяснения ошибки с привязкой к тексту эссе), исправления и тега.

- 2.17. **Параметры Конкурса** — числовые параметры, влияющие на значения автоматически вычисляемых критериев. Параметры Конкурса объявляются перед началом цикла и перед Финальными испытаниями, а также могут быть изменены решением Оргкомитета при согласовании с Судейской коллегией и Технической комиссией Конкурса.
- 2.18. **Платформа** – сайт в компьютерной сети Интернет, обеспечивающий доступ к данным, аналитической информации и сервисам для всех категорий пользователей, вовлечённых в участие в Конкурсе и проведение Конкурса. В частности, Участникам предоставляется возможность тестирования Алгоритмов разметки, Экспертам предоставляется возможность размечать тексты эссе в Программе Разметчике, и т.д. Платформа имеет интерфейс на русском и английском языках.
- 2.19. **Программа Разметчик** – программное обеспечение с вэб-интерфейсом на базе Платформы, которое позволяет проводить разметку текстовых файлов, выделяя и описывая ошибки и смысловые блоки согласно КО. Доступ к полной функциональности Разметчика имеют определённые Оргкомитетом категории пользователей. Участники имеют права на просмотр разметки, но не могут самостоятельно делать разметку в интерфейсе Разметчика.
- 2.20. **ПКУ, Программный комплекс Участника** – разработанный Командой Участника Конкурса комплекс программного и аппаратного обеспечения.
- 2.21. **ПСР, Программа сравнения решений** – программное обеспечение с вэб-интерфейсом на базе Платформы, позволяющее пользователю визуально сравнить два размеченных файла одного и того же эссе (например, алгоритмическую разметку с экспертной или две экспертных разметки от разных Экспертов).
- 2.22. **Размеченный файл, разметка файла** – текстовый файл эссе, размеченный в соответствии с правилами, описанными в данном Техническом регламенте.
- 2.23. **СПР, Система проверки решений** – автоматическая система проверки, которая принимает выборку файлов с алгоритмической разметкой, сравнивает их с соответствующей экспертной разметкой и выдаёт результат сравнения в форме аналитического отчёта. Результат проверки может быть визуально проконтролирован для любого файла из выборки в Программе сравнения решений.
- 2.24. **Смысловой блок** – фрагмент текста эссе, обладающий обязательными параметрами начала, конца и типа смыслового блока согласно КО. Смысловые блоки не являются ошибками, но их наличие может влиять на оценку за эссе.
- 2.25. **Участник, Участник Конкурса** — команда от 2 до 5 человек (российские или иностранные физические лица) чьи заявки на участие в Конкурсе прошли одобрение Оргкомитетом. Все требования положений Конкурсного задания распространяются на участника также, как и на команду.

- 2.26. **Файл, файл с текстом эссе** – размеченный или неразмеченный текстовый файл, имеющий расширение *.txt и кодировку UTF-8 без спецификаций.
- 2.27. **Цикл** – единый логический блок Конкурса, который состоит из этапов квалификации, финального испытания и подведения итогов.
- 2.28. **Эксперт, Специалист** – привлечённый профильный специалист общеобразовательного учреждения, который проверяет текст эссе и формирует в результате проверки размеченный файл.
- 2.29. **Экспертная разметка** – разметка текста эссе, сформированная экспертом в результате проверки вручную неразмеченного файла данного эссе. В Платформе реализован специальный сервис для автоматизации процесса экспертной разметки.
- 2.30. **Эссе, текст, текст эссе** – оцениваемая письменная работа учащегося в виде развернутого и аргументированного текста, написанная по заданной проблеме или теме (например, сочинение, написанное учеником 11 класса общеобразовательной школы в рамках единого государственного экзамена по русскому языку, литературе).

3. Участники конкурса и условия участия

- 3.1. Для участия в конкурсе Участник подаёт заявку путем заполнения электронной формы на Сайте конкурса <https://ai.upgreat.one>. Порядок формирования Команды Участника описан в п.4 КЗ.
- 3.2. В состав Команды может входить от 2 до 5 человек.

4. Циклы и этапы конкурса

- 4.1. Конкурс проводится в один цикл и включает в себя Отборочный этап (п. 6 КЗ), Финальный этап (п. 7 КЗ) и Подведение итогов (п. 8 КЗ).
- 4.2. Использование общих программных модулей (в случае, если команде предоставляется доступ к подобным модулям) является обязательным, поскольку обеспечивает равные условия для всех Участников Конкурса. Участники Конкурса не имеют право самостоятельно вносить изменения в общие программные модули. В случаях обнаружения ошибок в них Участник может подать в Оргкомитет по электронной почте в свободной форме заявку «О внесении изменений в общие программные модули», описав в ней обнаруженные ошибки и/или свои предложения по улучшению работы модулей.
- 4.3. Оргкомитет оставляет за собой право обновлять Обучающую и Тестовую выборки. Увеличение Обучающей и Тестовой выборок возможно в течение всего Конкурса. Корректировка Обучающей и Тестовой выборок возможна не позднее, чем за 10 дней до Финальных испытаний.

- 4.4. Оргкомитет оставляет за собой право обновлять исходные коды общих программных модулей, но не позднее, чем за 10 дней до Финальных испытаний. В случае обновления Оргкомитет оповещает всех Участников путём публикации на сайте Платформы, сообщая цель и суть произведённых модификаций

5. Конкурсная задача и оценивание решений Участников

- 5.1. Предметом Конкурса является выбор наилучшего решения для автоматического определения итоговой оценки (оценка за эссе в баллах по методике ЕГЭ) по текстам академических эссе и сочинений, с помощью программных комплексов, разработанных участниками Конкурса.
- 5.2. Лидерборд строится по Относительной точности предсказания оценки - чем оно выше, тем лучше. Пусть $K(X)$ и $K(Y)$ – две оценки за эссе (в баллах), вычисленные по разметкам X и Y соответственно. Определим парную точность разметки X относительно разметки Y при предсказании оценки за эссе:

$$m(X, Y) = \left(1 - \frac{|K(X) - K(Y)|}{\max K}\right) \cdot 100\%$$

где $\max K$ – максимальная возможная оценка за эссе данного типа.

Относительная точность предсказания оценки (ОТПО) определяется как отношение средней по всем эссе парной точности $m(A, \mathcal{E})$ алгоритмической разметки A относительно экспертной разметки \mathcal{E} к средней по всем эссе парной точности $m(\mathcal{E}, \mathcal{E}')$ экспертных разметок $\mathcal{E}, \mathcal{E}'$ друг относительно друга:

$$\text{ОТПО} = \frac{\text{avr}_{\text{эссе}} \text{avr}_{\mathcal{E}} m(A, \mathcal{E})}{\text{avr}_{\text{эссе}} \text{avr}_{\mathcal{E}, \mathcal{E}'} m(\mathcal{E}, \mathcal{E}')} \cdot 100\%.$$

По построению, парная точность $m(X, Y)$ аналогична метрике $M_1(X, Y)$, а *относительная точность предсказания оценки (ОТПО)* эквивалентна ОТАР, если в качестве метрики $M_1(X, Y)$ взять $m(X, Y)$, а остальные метрики $M_2 \dots M_7$ не использовать ($w_2 = \dots = w_7 = 0$).

- 5.3. Универсальный классификатор ошибок и методика расчета метрик представлена в приложении 1 и является частью технологического конкурса Национальной технологической инициативы Up Great «Взаимное обучение на русском языке» и «Взаимное обучение на английском языке»

6. Отборочный этап

- 6.1. Данный раздел дополняет п.6 КЗ.
- 6.2. Отборочный этап Конкурса заключается в определении команд, соответствующих двум критериям:

- корректно заполненная заявка;
- подтверждение Командой наличия работающего Программного комплекса и готовности участия в Испытаниях.

7. Этап финальных испытаний

7.1. Данный раздел дополняет п.7 КЗ.

7.2. В этапе Финальных испытаний могут участвовать все Участники, успешно прошедшие Отборочный этап.

7.3. В день проведения Финальных испытаний открывается доступ к Платформе.

7.4. В момент начала проведения Испытаний участнику становится доступен архив содержащий:

- файлы эссе;
- файл-задание в формате json.

7.5. Участник должен обработать полученные данные, внести результат обработки в поля “answer” файла-задания и загрузить файл-задание на Платформу. Разрешается загружать не более 3 вариантов файла-задания. Учитывается только последний вариант.

7.6. Структура файла-задания:

```
[
  {
    "id": 80001,
    "answer": -1
  },
  {
    "id": 80001,
    "answer": -1
  }
  ....
]
```

7.7. В поле answer участник должен записать оценку за эссе (в баллах), формат поля - число. По умолчанию установлено значение -1.

7.8. В поле answer допускается вносить только одну оценку

7.9. Испытания ограничиваются объемом времени в 15 минут.

7.10. Количество эссе не менее 1000 файлов.

7.11. Порядок определения премий победителю и призёрам описан в п.10 КЗ.

7.12. Порядок перезапуска Испытаний:

7.12.1. Испытания могут быть перезапущены в случае массового технического сбоя по

решению Оргкомитета, на основании соответствующих протоколов Судейской Коллегии.

7.12.2. Перезапуск может быть назначен на следующий день, о чём сообщается отдельно на официальном сайте.

7.12.3. Файлы на перезапуск новые. Уже отданные не учитываются.

7.12.4. Результаты по отданным файлам не учитываются.

8. Судейская коллегия

8.1. На Отборочном этапе:

8.1.1. Судейская коллегия проверяет степень готовности ПКУ к проведению Финального испытания

8.1.2. Оргкомитет выносит окончательные решения о допуске Участников к Финальному испытанию.

8.2. На этапе Финальных испытаний

8.2.1. Судейская коллегия проверяет автоматически формируемые отчеты о решениях Участников

8.2.2. Оргкомитет выносит решения по итогам Финального испытания.

9. Права Оргкомитета

9.1. Оргкомитет оставляет за собой право обновлять Обучающую и Тестовую выборки. Увеличение Обучающей и Тестовой выборок возможно в течение всего Конкурса. Корректировка Обучающей и Тестовой выборок возможна не позднее, чем за 10 дней до Финальных испытаний.

9.2. Оргкомитет оставляет за собой право обновлять исходные коды общих программных модулей, но не позднее, чем за 10 дней до Финальных испытаний. В случае обновления Оргкомитет оповещает всех Участников путём публикации на сайте Платформы, сообщая цель и суть произведённых модификаций.

9.3. Оргкомитет имеет право приостановить Финальные испытания, например, для выяснения причины технического сбоя на стороне Платформы и его оперативного устранения.

9.4. Оргкомитет имеет право проверить решение команды очно.

9.5. Оргкомитет имеет право запросить у команды дополнительные данные, обеспечивающие прозрачность и объективность судейства.

- 9.6. Оргкомитет имеет право аннулировать решения команды в случае нарушения настоящего Технического регламента и/или Конкурсного задания.
- 9.7. Оргкомитет имеет право публиковать разметки Участников после Финальных испытаний на Сайте конкурса с целью обеспечения прозрачности в определении победителей и призёров.
- 9.8. Оргкомитет имеет право аннулировать заявку, отстранить Команду и/или Участника от участия в Конкурсе, если выясняются обстоятельства недобросовестного поведения, подлога или нарушения положений Конкурсного задания или Технического регламента. В том числе отстранение от дальнейшего участия в Конкурсе может производиться в следующих случаях: выявлены действия команды, направленные на получение, распространение, использование информации об ошибках или технических особенностях Системы принятия решений, Сайта конкурса, Технического регламента, Процедурного регламента, Конкурсного задания и иных документов, регламентирующих проведение Конкурса равно как технических решений, используемых в ходе его проведения, которые могут быть использованы для получения несправедливого преимущества в ходе Конкурса, в том числе искажения результатов испытаний, решений Жюри, доступа к информации о разработках других участников, равно как и не доведение такой информации, в случае, если она стала известна Команде, до сведения Оргкомитета.

10. Участникам запрещается

- 10.1. Участнику запрещается осуществлять экспертную разметку Тестовой и Финальной выборки вручную, с помощью краудсорсинга или любыми иными способами с привлечением труда людей.
- 10.2. Запрещается привлекать труд людей для извлечения каких-либо дополнительных данных из текстов эссе в неразмеченных выборках – Тестовой и Финальной. Анализ и разметка этих текстов должны производиться Участником исключительно алгоритмически.
- 10.3. Запрещается самостоятельно модифицировать общие программные модули.
- 10.4. Запрещается на этапе Финальных испытаний отправлять файл решения более 3 раз. Все последующие отправки будут проигнорированы, в качестве окончательного варианта будет рассматриваться последняя отправка.
- 10.5. Запрещается на этапе Финальных испытаний отправлять вместо файла решения Конкурса какую-либо иную информацию. Такие отправки будут проигнорированы.
- 10.6. В случае нарушения перечисленных запретов Оргкомитет по запросу Судейской коллегии имеет право отстранить Команду от участия в Конкурсе

11. Обеспечение безопасности и дополнительные условия

- 11.1. Деятельность Участников в рамках Конкурса должна соответствовать действующим на территории Российской Федерации природоохранным нормам, требованиям техники безопасности.
- 11.2. В случае проведения очных мероприятий Оргкомитет Конкурса предоставляет Участникам правила по технике безопасности и охране окружающей среды, с которыми все члены команды Участника должны ознакомиться под подпись и соблюдать.
- 11.3. Все технические инструкции по работе Платформы, а также параметры и способы работы Платформы являются предметом отдельных документов, которые публикуются на официальном сайте Конкурсов, или на Платформе или ссылки на них ведут с Сайта или Платформы. Указанные документы не влияют на сутевую составляющую Конкурсного задания и Технического регламента, но описывают технические параметры функциональности Платформы и одинаково доступны всем Командам, в том числе для соблюдения равных условий участия Команд.