***}Креативные индустрии***

***Тема задания заочного отборочного этапа конкурса***

*Создание алгоритма, который генерирует цельное видеопроизведение (видеоистория или музыкальный клип) по текстовой строке и заданному жанру. Видеопроизведение должно иметь художественную ценность и включать в себя синхронизированные видео-, голосовую и музыкальную дорожки.*

***Преамбула***

*Долгое время бытовало мнение, что алгоритмы не способны создавать произведения искусства. Сегодня же технологии искусственного интеллекта набирают все бо́льшую популярность в креативных индустриях: с их помощью пишут музыку и тексты, рисуют картины.*

*Мы хотим предложить вам пойти дальше и которое может создать сложное цельное творческое произведение, синхронизировав работу различных алгоритмов. Для решения этой задачи вам предстоит столкнуться с генеративными (в том числе мультимодальными) моделями, разобраться в особенностях их работы и объединить в единый алгоритм. Подобного рода решения могут значительно облегчить или даже полностью автоматизировать трудоёмкий процесс создания творческих произведений и их визуализации.*

***Задание заочного отборочного этапа конкурса***

*В результате работы вы должны получить решение, способное по запросу пользователя сгенерировать цельное видеопроизведение, имеющее художественную ценность. Это может быть:*

*· видеоистория с озвучкой текста, музыкальным оформлением и видеорядом, иллюстрирующим содержание истории;*

*· музыкальный видеоклип — музыкальный трек (песня со словами) с дополняющим его видеорядом.*

*Запрос пользователя должен включать в себя название будущего произведения и жанр:*

*· возможные жанры для видеоистории: басня, былина, сказка, повесть, комедия, трагедия, драма, ужасы;*

*· возможные жанры для музыкального клипа: рэп, рок, поп, джаз, электро, ретро, опера, мюзикл.*

*Вам предстоит создать алгоритмы для работы со следующими задачами, которые впоследствии будут синхронизированы и объединены в итоговое решение:*

*· генерация текста;*

*· озвучка текста;*

*· создание музыки;*

*· генерация видеоряда.*

*Генерация текста может производиться на любом языке. За создание мультиязычных систем предусмотрены дополнительные баллы (см. Критерии оценки).*

***Бейзлайн***

*Вам предоставлено базовое решение от разработчиков задачи по генерации истории, а также несколько альтернативных технологий в виде дополнительных ноутбуков в Google Colab. Все материалы размещены на GitHub соревнования:*

*https://github.com/sberbank-ai/DigiTeller*

***Формат решения и форма представления результатов выполнения задания заочного отборочного этапа конкурса***

*Результаты выполнения конкурсного задания должны быть представлены в следующем виде:*

*● Код решения (желательно в виде ссылки на ноутбук в Google Colab) и обученные модели, если они были обучены участником в виде, например, сохранённых чекпойнтов.*

*● Ссылка на примеры результатов генерации (от 3 до 5 примеров).*

*● Видеопрезентация кода. Видеопрезентация должна быть представлена в виде ролика продолжительностью не более 5 минут. Видео должно быть построено в формате питча и содержать презентацию наработанных навыков, обоснования технических решений команды.*

***Критерии оценки задания заочного отборочного этапа конкурса***

*Финальная оценка будет складываться из оценок за следующие элементы алгоритма.*

***ВАЖНО:*** *В некоторых пунктах присутствуют ключевые подпункты (подчеркнуты линией). Без выполнения этих пунктов за весь блок, к которому относится пункт, ставится 0 баллов.*

***1. Модель генерации текста по названию и жанру***

*● Соответствие сгенерированного текста заданному названию и жанру, связность текста — 0–25 баллов.*

*Соответствие жанру оценивается только в случае формата видеоистории. Соответствие жанру музыкального клипа оценивается в ходе оценки модели по генерации песни.*

*● Объём связного текста (в прочтении вашей моделью text2speech или какой-либо другой моделью, см. ниже пункт 2 и Примечание):*

*– меньше 2 минут — 0 баллов;*

*– от 2 минут (включительно) до 3 минут — 2 балла;*

*– от 3 минут (включительно) до 4 минут — 5 баллов;*

*– от 4 минут (включительно) до 5 минут — 7,5 балла;*

*– более 5 минут — 10 баллов.*

*● Поддержка двух языков — 10 баллов, трёх и более — 15 баллов.*

***Максимальное количество баллов за данную часть — 50 баллов.***

***2. Модель(-и) генерации голоса и музыки***

*В данном пункте возможны 2 сценария:*

*1) генерация фоновой музыкальной дорожки и голосовой дорожки, в которой озвучен сгенерированный ранее текст;*

*2) генерация полноценной песни, где музыкальная и голосовая дорожки должны быть полностью синхронизированы между собой.*

*Для каждого из сценариев предусмотрена своя система оценки и максимальное количество баллов. При этом можно работать над обоими сценариями и получить сумму баллов.*

***Сценарий 1.***

*2.1.1 Модель генерации фоновой музыки*

*● Качество сгенерированной музыки — 0–10 баллов.*

*● Передача музыкой настроения рассказа в каждый момент времени — 0–10 баллов.*

*2.1.2 Модель text2speech*

*● Качество сгенерированного голоса, соответствие тексту — 0–10 баллов.*

*● Генерация мультиголоса (генерация голосов для каждого персонажа, если в рассказе их несколько) — 0–5 баллов.*

*● Поддержка двух языков — 10 баллов, трёх и более — 15 баллов.*

***Максимальное количество баллов за сценарий — 50 баллов.***

***Сценарий 2.***

*2.2.1 Генерация песни*

*· Качественное звучание голосовой дорожки и её соответствие тексту — 0–10 баллов.*

*· Качественное звучание музыкальной дорожки и её синхронность с голосовой дорожкой — 0–35 баллов.*

*· Генерация мультиголоса в случае, если в песне присутствуют диалоги — 0–10 баллов.*

*· Соответствие музыкальной и голосовой дорожек исходному жанру — 0–10 баллов.*

*· Поддержка двух языков — 10 баллов, трёх и более — 15 баллов.*

***Максимальное количество баллов за сценарий — 90 баллов.***

***3. Модель(-и) генерации видео***

*В данном пункте опять возможны 2 сценария:*

*1) генерация картин по строчкам текста и их соединение (например, интерполяцией или морфингом);*

*2) генерация видео целиком.*

*Для каждого из сценариев своя система оценки и максимальное количество баллов. При этом можно работать над обоими сценариями и получить сумму баллов.*

*3.1 Генерация картин по строчкам текста и их соединение:*

*● Соответствие в каждый момент времени кадра и смысла текста — 0–20 баллов.*

*● Соответствие стиля кадра стилю повествования — 0–10 баллов.*

*● Качественное соединение сгенерированных кадров друг с другом в единое видео — 0–10 баллов.*

*● Разрешение сгенерированного клипа:*

*– 512х512 — 0 баллов;*

*– 1024х1024 — 2,5 балла;*

*– 2048x2048 — 5 баллов;*

*– 4096x4096 — 7,5 балла;*

*– 8192x8192 — 10 баллов.*

***Максимальное количество баллов за сценарий — 50 баллов.***

*3.2 Генерация видео целиком\*:*

*● Соответствие в каждый момент времени визуального смысла кадра из видео и смысла текста, жанра и стиля музыкальной и звуковой дорожек — 0–110 баллов.*

*● Разрешение сгенерированного клипа:*

*– 512х512 — 0 баллов;*

*– 1024х1024 — 2,5 балла;*

*– 2048x2048 — 5 баллов;*

*– 4096x4096 — 7,5 балла;*

*– 8192x8192 — 10 баллов.*

*\* Формат видеоклипов может быть самым разным. Например, можно генерировать человека (или виртуального персонажа), который будет “пропевать” эту песню и двигаться под неё (на виртуальной сцене). Для этого можно использовать следующие идеи:*

*a) “пропевать” песню самому (то есть снимать видео со своим участием). Затем использовать FaceSwap (BodySwap), и в получившемся видео заменять себя на этого персонажа, сохраняя движения (см.* [*обзор методов FaceSwap*](https://www.notion.so/FaceSwap-350382dc415d4dff9fe282500a524e13)*).*

*б) генерировать (или задавать вручную) в каждый момент времени положения персонажа в пространстве (так называемый “скелет”). А затем генерировать видеореалистичную версию этого персонажа (см.* [*пример такой генерации*](https://people.mpi-inf.mpg.de/~mhaberma/projects/2021-ddc/)*).*

*При этом на заднем фоне можно пускать слайдшоу сгенерированных картинок или абстрактных образов.*

*Принимаются любые оригинальные идеи (идея выше дана лишь как пример).*

***Максимальное количество баллов за сценарий — 120 баллов.***

*Оценку будет выставлять экспертное жюри.* ***Максимальная возможная оценка — 360 баллов****.*

***Бонус:***

*Публикуйте свои результаты в социальных сетях с хештегами #SberDigiTale, #SberAI, #AIIJC, при количестве лайков более 100 дополнительно присуждаются 10 баллов, более 1000 лайков — 20 баллов.*

*Примечание:1)* ***Длительность финального видео не должна быть меньше 1 минуты. Решения меньше 1 минуты рассматриваться не будут (оценка — 0 баллов).***

***2) С учётом дополнительных баллов для популярных решений максимальное число баллов равно 380.***