

Submission guidelines. Technical details

Data structure

General format for all tasks:

```
{"0": [{"type": ..., "content": ...}, ...], "1": [...], ...}
```

Examples

Input.json

```
{"0": [{"type": "text", "content": "Ответь на вопрос по изображению. На чём едет человек справа на картинке?"}, {"type": "image", "content": "images/00000.jpg"}, {"type": "text", "content": "варианты: a) на велосипеде; b) на лошади; c) на машине; d) на самокате"}], "1": [{"type": "text", "content": "Дай описание изображения."}, {"type": "image", "content": "images/00005.jpg"}], "2": [{"type": "text", "content": "Сгенерируй на изображении."}, {"type": "text", "content": "Пара человек сидит в тени. Велосипед, оставленный на велопарковке, стоит на солнце."}], ...}
```

Ground_truth.json

```
{"0": [{"type": "text", "content": ["а на велосипеде", "а", "на велосипеде"]}], "1": [{"type": "text", "content": ["две птицы летают над водой", "птицы летают", "два сокола летят над озером"]}], "2": [{"type": "image", "content": "gt_images/image_2.jpg"}], ...}
```

Predictions.json

```
{"0": [{"type": "text", "content": "a")}, {"type": "image", "content": "images/image_0.jpg"}], "1": [{"type": "text", "content": "две большие птицы летают над озером."}], "2": [{"type": "image", "content": "images/image_2.jpg"}], ...}
```

Data structure

TextQA

Input:

```
{"0": [{"type": "text", "content": "Мужская сборная команда Норвегии по биатлону в рамках этапа Кубка мира в немецком Оберхофе выиграла эстафетную гонку. Вторыми стали французы, а бронзу получила немецкая команда. ..."}, {"type": "text", "content": "На сколько секунд сборная России опередила ближайших соперниц?"}, {"type": "text", "content": "а) на 0,3 секунды; б) на секунду; в) на 0,5 секунд"}], }
```

```
"1": [{"type": "text", "content": "Определи правильный ответ по тексту."}, {"type": "text", "content": "Улица Авиаконструктора Сухого (название с 2004 года) — улица в Северном административном округе города Москвы на территории Хорошёвского района. ..."}, {"type": "text", "content": "Какая улица пролегает от Проектируемого проезда № 6161 до пересечения с юго-восточной частью взлётно-посадочной полосы Ходынского Поля?"}], }
```

Ground Truth:

```
{"0": {"type": "text", "content": ["а", "на 0,3 секунды", "а на 0,3 секунды"]},  
"1": {"type": "text", "content": ["Авиаконструктора Сухого"]}, ... }
```

Solution format:

```
{"0": [{"type": "text", "content": "_answer to question 0_"}, ...]}
```

MathQA

Input:

```
{"0": [{"type": "text", "content": "Решите линейное уравнение."}, {"type": "text", "content": "6 умножить на у – 54 это -72, для у."}, {"type": "text", "content": "a) -4; б) -20; в) -5; г) -3"}], ... }
```

Ground Truth:

```
{"0": {"type": "text", "content": ["d", "-3", "d -3"]}, ... }
```

Solution format:

```
{"0": [{"type": "text", "content": "_answer to question 0_"}, ...]}
```

Data structure

ImageGeneration

Input:

```
{ "0": [ {"type": "text", "content": "Сгенерируй изображение по тексту."},  
 {"type": "text", "content": "Торт со светло-бежевой глазурью, украшенный  
 фигуркой медведя, держащий связку воздушных шариков."}], ... }
```

Ground Truth:

```
{ "0": [ {"type": "image", "content": "output_images/00000.jpg"}], ... }
```

Solution format:

```
{"0": [ {"type": "image", "content": "images/_name of image 0_"}, ... ]}
```

ImageCaptioning

Input:

```
{ "0": [ {"type": "text", "content": "Что представлено на картинке?"}, {"type":  
 "image", "content": "images/00000.jpg"}], ... }
```

Ground Truth:

```
{ "0": [ {"type": "text", "content": ["Торт со светло-бежевой глазурью,  
 украшенный фигуркой медведя, держащий связку воздушных шариков.",  
 "Детский торт, покрытый светлой глазурью с украшением в виде  
 медвежонка."]}], ... }
```

Solution format:

```
{"0": [ {"type": "text", "content": "_caption for image 0_"}, ... ], ... }
```



images/00000.jpg

Data structure

VisualQA

Input:

```
{ "0": [ {"type": "text", "content": "Ответь на вопрос по картинке."}, {"type": "text", "content": "Какого цвета плакат на ограждении?"}, {"type": "image", "content": "images/00001.jpg"}], ... }
```

Ground Truth:

```
{ "0": [ {"type": "text", "content": ["желтого", "ярко-желтого"]}], ... }
```

Solution format:

```
{"0": [ {"type": "text", "content": "_answer to question 0_"}, ...], ...}
```



images/00001.jpg

TRitW

Input:

```
{ "0": [ {"type": "text", "content": "Какой текст изображен на картинке?"}, {"type": "image", "content": "images/00002.jpg"}], ... }
```

Ground Truth:

```
{ "0": [ {"type": "text", "content": ["наука"]}], ... }
```

Solution format:

```
{"0": [ {"type": "text", "content": "_text in image 0_"}, ...], ...}
```



images/00002.jpg

Evaluation metrics

TextQA

$$S_{TextQA} = F1 = 2 \cdot \frac{Recall \cdot Precision}{Recall + Precision}$$

$$Precision = \frac{|tokens_{common}|}{|tokens_{predicted}|}, Recall = \frac{|tokens_{common}|}{|tokens_{gt}|}$$

where $|tokens_{common}|$ is the number of common tokens for the predicted and the true answer;
 $|tokens_{predicted}|$ is the number of tokens in the predicted answer;
 $|tokens_{gt}|$ is the number of tokens in the true answer.

MathQA

$$S_{MathQA} = Exact\ Match$$

This metric shall receive the value of 1, if all tokens in the predicted answer are in complete accord with tokens in the true answer; otherwise the metric value shall be $EM = 0$.

Evaluation metrics

ImageGeneration

FID is the Frechet distance between two multidimensional Gaussian distributions:

$N(\mu, \Sigma)$ – a distribution of features for images created by the generator, and

$N(\mu_\omega, \Sigma_\omega)$ – a distribution of features for real images used to train the network.

$$FID = |\mu - \mu_\omega|^2 + \text{tr} \left(\Sigma + \Sigma_\omega - 2(\Sigma \Sigma_\omega)^{\frac{1}{2}} \right)$$

CLIP score is a metric that allows evaluating how much the visual presentation corresponds to the textual description.

$$\text{similarity} = \frac{x_1 \cdot x_2}{\max(\|x_1\|_2 \cdot \|x_2\|_2, \epsilon)}$$

where x_1 is the textual embedding,

x_2 is the picture embedding,

ϵ is the arbitrarily small positive number entered to avoid the division by 0.

$$S_{ImageGeneration} = \frac{1}{2} \left(CLIP_{score} + \frac{200 - \min(200, X)}{200} \right)$$

Evaluation metrics

ImageCaptioning

$$METEOR = \frac{10PR}{R+9P}$$

$$P = \frac{m}{w_t},$$

where m is the number of n-grams in the machine translation that were found in the reference one, w_t is the number of n-grams in the machine translation.

$$R = \frac{m}{w_r},$$

where w_r is the number of n-grams in the reference translation.

$$S_{ImageCaptioning} = \frac{1}{2} \cdot (METEOR + CLIP_{score})$$

VQA

$$S_{VisualQA} = METEOR$$

Some modifications:

1. If answer is number, the smallest number shall be selected from a couple of predicted and reference results and divided by the biggest one. Therefore, if the numbers match, the metric for this pair equals to 1; otherwise, the metric shall be calculated proportionally.
2. Numerals shall be translated from the text format to the numeric one: "three" - 3.

Evaluation metrics

TRitW

$$S_{TRitW} = 1 - NED = 1 - \frac{D(s_i, \hat{s}_i)}{\max(l_i, \hat{l}_i)}$$

where $D(\cdot)$ indicates a Levenshtein distance between the predicted text s_i and the ground truth one \hat{s}_i ; l_i and \hat{l}_i denote the length of s_i and \hat{s}_i , respectively.

Hidden tasks

For each hidden sub-task the corresponding hidden metric S_i^h is calculated. $S_k \in [0,1]$, where 0 is the worst, and 1 is the best metric's value, $k \in \{\text{Hidden1}, \text{Hidden2}, \text{Hidden3}, \text{Hidden4}, \text{Hidden5}, \text{Hidden6}\}$.

Evaluation metrics

Integral Metric

$$S = \frac{1}{15} \sum_{j \in [open_t]} S_j + \frac{1}{10} \sum_{k \in [hidden_t]} S_k$$

Threshold for the open tasks:

- TextQA: 0.2
- MathQA: 0.25
- ImageGeneration: 0.21
- ImageCaptioning: 0.15
- VisualQA: 0.24
- TRitW: 0.13

General rules for metric calculation

- For textual tasks the answer shall be read from the **first** dictionary, where the "**type**" key corresponds to the "**text**" value. If the model generates an image for sub-tasks with textual output, a dictionary for the image (`{"type": "image", "content": ...}`) **shall be ignored** when calculating metrics.
- In most cases for sub-tasks with textual output, a correct answers file shall contain **several correct answer options** for each input example. When calculating metrics for tasks with textual output, a prediction generated by the model shall be compared to each answer option, and **the maximum calculated metric value shall be saved**.
- Before calculation of metrics, **the post-processing of generated predictions is conducted**: reducing to lower case, deleting punctuation characters, reducing words to normal form.
- For image tasks the answer shall be read from the first dictionary, where the "**type**" key corresponds to the "**image**" value. If the model generates some text in addition to the image, a dictionary for the textual output (`{"type": "text", "content": ...}`) **shall be ignored** when calculating metrics.

Errors in metrics

- Error: File is not found. Check predictions.json file and that the predictions are saved in the output folder.
- Error: Unable to read json file. - file contains invalid structure.
- Error: A dictionary is expected. - json must contain a dictionary structure.
- Error: Not all inputs contain predictions! - for each sample a response must be generated.
- Error: {image_name} path doesn't exist. - image path is incorrect or image didn't save.
- Error: unable to open image {image_name}. - image is not opening using Image.open()
- ERROR_CUSTOM_SUBMISSION: {n} final metrics was not calculated.

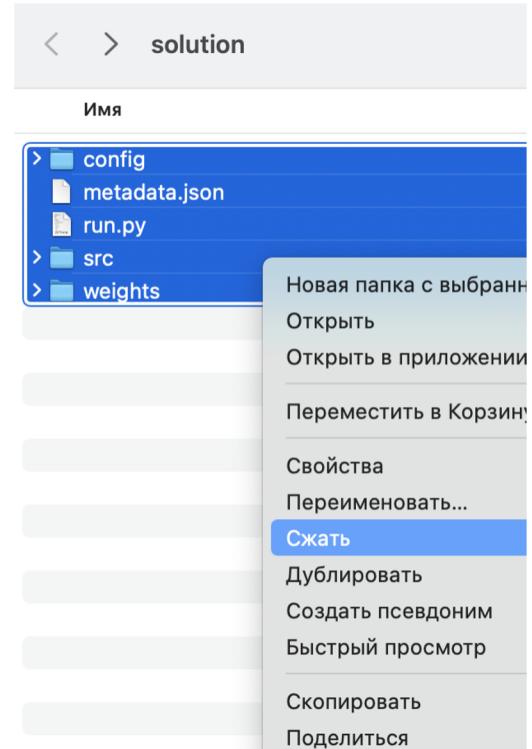
Container contents

The archive root must contain ***metadata.json*** file with the following contents:

```
{  
    "image": "cr.msk.sbercloud.ru/aijcontest2022_official/fusion:0.0.1",  
    "entrypoint": "python3 run.py"  
}
```

config/fusion_inference.yaml

```
data:  
    input: '../input/input.json'  
    pred_output: './output/predictions.json'  
    pred_images_output: './output/images/'  
model:  
    rudolph:  
        name: '2.7B'  
        fp16: true  
        device: 'cuda'  
        cache_dir: '/tmp/rudolph'  
        pretrained: False  
    rudolph_weight: './weights/light_6tasks_rudolph-2_7b-v4.pt'  
    tokenizer_weights: './weights/bpe.model'  
vae:  
    dwt: false  
    device: 'cuda'  
vae_weights: './weights/vqgan.gumbel8-sber.model.ckpt'
```

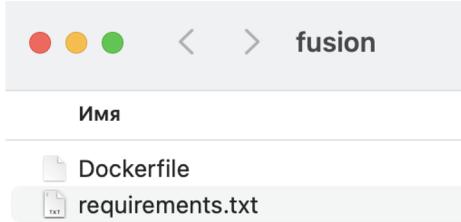


archive creating

Code for inference and train https://github.com/ai-forever/fbc2_aij2022

Custom docker images

Create folder with Dockerfile and requirements.txt



requirements.txt

```
einops==0.3.2
transformers==4.10.2
catalyst==21.8
rudalle
catalyst
bitsandbytes
pandas
wandb
ruclip~=0.0.2
razdel
jiver
numpy~=1.21.6
Pillow~=9.2.0
more_itertools
```

Dockerfile

```
FROM cr.msk.sbercloud.ru/aicloud-base-images/horovod-cuda11.1-pt1.9.1-a100:0.0.28
USER root
WORKDIR /app
ARG DEBIAN_FRONTEND=noninteractive
COPY requirements.txt .
RUN pip install -r requirements.txt
USER jovyan
WORKDIR /home/jovyan
```

Install Docker Desktop and open Terminal

```
Username: aijcontest
Password: AIJcontest22
```

```
> docker login cr.msk.sbercloud.ru
> cd fusion/
> docker build --tag myimage:0.0.1 .
> docker tag myimage:0.0.1 cr.msk.sbercloud.ru/aijcontest2022/teamtest:0.0.1
> docker push cr.msk.sbercloud.ru/aijcontest2022/teamtest:0.0.1
```

Limitations

A container with the solution will be run under the following conditions:

- 100 Gb RAM
- 3 vCPU
- 1 GPU Tesla A100 (80 Gb)
- time to execute the solution: 5 hours
- the solution does not have access to internet resources;
- maximum size of packed and unpacked archive with the solution: 15 Gb
- maximum size of the Docker image used: 25 Gb
- no more than 3 valid solutions for evaluation during one day.

We provide the participants with an opportunity to get access to computing resources of Christofari to train their models. The resources are limited. To get access, it is necessary to send a request to Christofari_AIContest_2022@sberbank.ru describing how exactly you are planning to use the computing resources.

DS Works

Регистрация

Принять пользовательское соглашение [»](#)

Отправить

Техподдержка Чат в Телеграм

1) user registration



@testfbc2022
С нами с 15.10.2022

Электронная почта
testfbc2022@gmail.com...

Начинающий
Грейд

Место 0 Скор 0

Победы в чемпионатах

0 0 0

Настройки

Личные данные

Имя

Фамилия

Отчество

Дата рождения

Номер телефона

Биография

Страна

Город

ВУЗ

Место работы

Отправить

2) fill in personal information



AIJ <contest>

FusionBrain Challenge 2.0

2 200 000 ₽ - Призовой фонд

10 окт. 2022 г. - 11 нояб. 2022 г.

Участвовать

Обзор Данные Турнирная таблица Комментарии

Описание Правила Призовой фонд

FusionBrain Challenge 2.0

3) competition selection

Обзор Данные Турнирная таблица Команда Комментарии

Регистрация команды

Вы можете собрать команду из нескольких человек или участвовать в одиночку.

Название команды FusionBrain

Участники

testfbc2022 testfbc2022@gmail.com

Введите email и нажмите Enter

Добавить

Зарегистрировать команду

4) team registration

DS Works

FusionBrain Challenge 2.0

2 200 000 ₽ - Призовой фонд

⌚ 10 окт. 2022 г. - 11 нояб. 2022 г.

Участвую

Обзор Данные Турнирная таблица Команда Загрузить решение Комментарии

Отправка решения

Загрузите файл

Чтобы успешно загрузить файл на сервер, используйте формат ZIP. У вас есть ограниченное число попыток каждые 24 часа. Одна попытка — один файл. Невалидный файл за попытку не считается.



Перетащите файлы сюда
или выберите

sample_submission.zip, 4.86 Gb

3 %

Добавьте описание

Кратко опишите ваше решение

Отправить

До следующей попытки
00:00:00

Осталось попыток
3

5) solution download

Отправить

До следующей попытки
00:00:00

Осталось попыток
2

История загрузок

Выберите 3 результата в качестве итогового решения для участия в рейтинге. В случае выбора меньшего кол-ва результатов система выберет лучший результат из имеющихся.

Название	Дата	Final_Score	TextQA	MathQA	ImageGeneration	Статус
<input type="checkbox"/> sample_submission.zip	15.10.2022, 17:59					Main predict

6) check solution status

Отправить

До следующей попытки
00:00:00

Осталось попыток
2

История загрузок

Выберите 3 результата в качестве итогового решения для участия в рейтинге. В случае выбора меньшего кол-ва результатов система выберет лучший результат из имеющихся.

Название	Дата	Final_Score	TextQA	MathQA	ImageGeneration	Статус
<input type="checkbox"/> sample_submission.zip	15.10.2022, 17:59	0.24327	0.25302	0.3125	0.28438	Complete

7) get scores

Pay attention

- No internet access, all weights should be archived.
- Remove all prints, tqdm for inference; **otherwise solution will hang.**
- Be sure to wait until the solution is fully loaded on the contest platform, otherwise an error “Failed transfer” occurs.
- Custom docker image should be downloaded in aijcontest2022.
- The "time" column in the leaderboard is the time of the main prediction stage.