



Языковые, мультимодальные и генеративные модели

Денис Димитров

Управляющий директор
по исследованию данных, Sber AI;
Руководитель научной группы Sber AI Research;
Научный консультант, AIRI

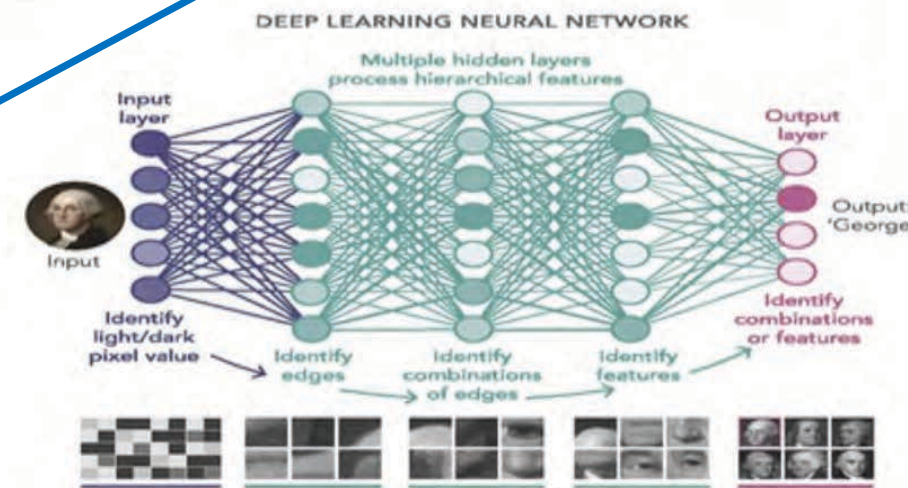
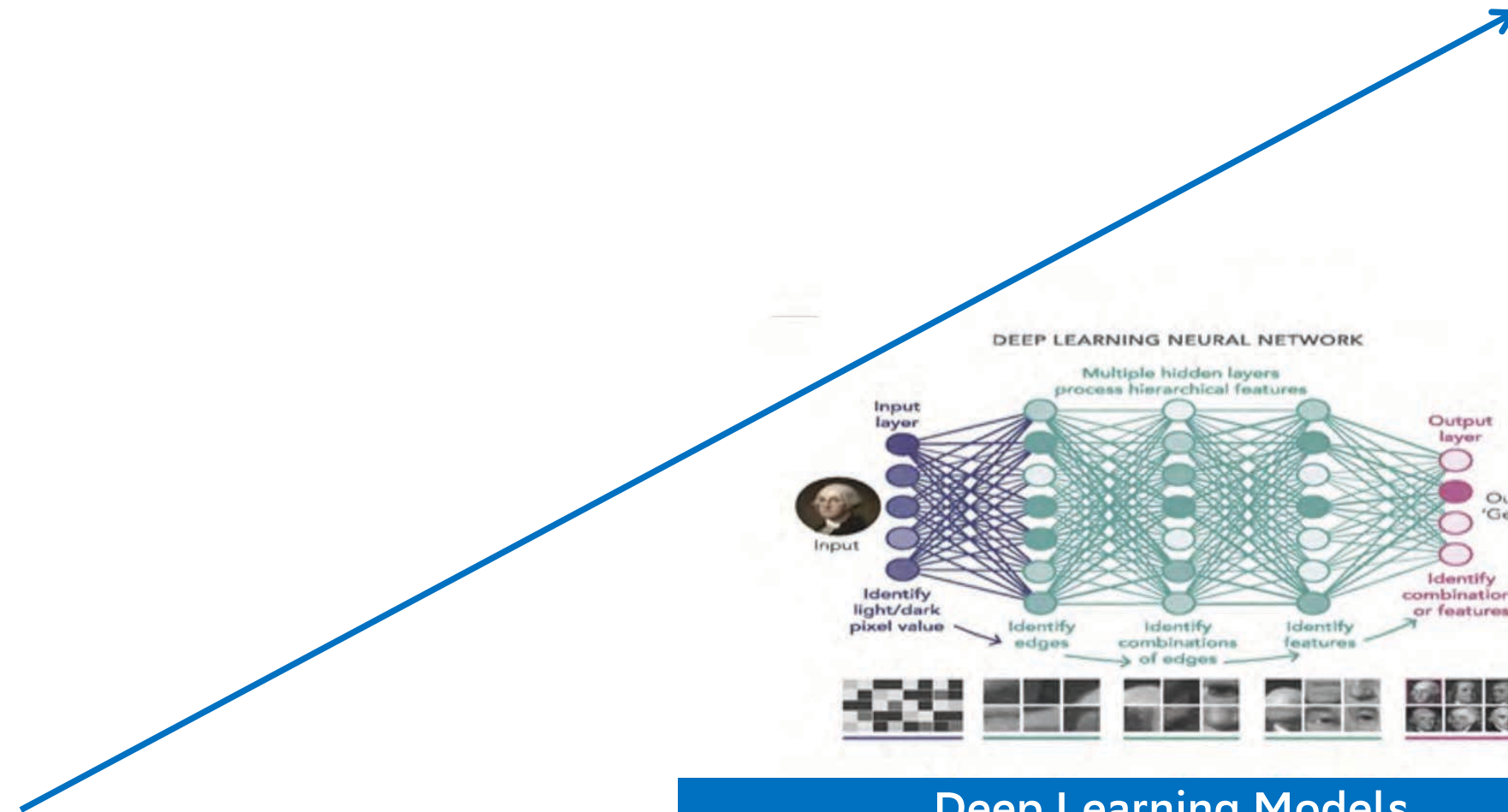
SBER
RE
RAI

01

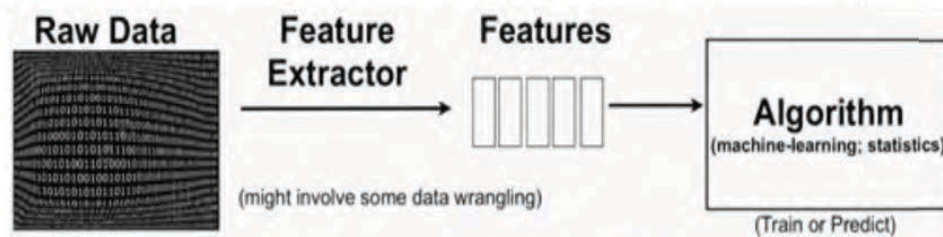
Введение

SEB
RE
RAI

Изменение парадигмы в AI



Deep Learning Models
(supervised learning, 2010–2017)



Statistical Models
(feature engineering, 1990–2010)

Masked Language Model

BERT A bird with a small head, yellow belly and short tail. → A bird with a small [red box], yellow [orange box] and short [green box].

Masked Image Models



Pre-trained Foundation Models
(self-supervised learning, 2017–now)



Единая мультимодальная и мультизадачная фундаментальная модель вместо набора узкоспециализированных унимодальных моделей

Искусственный интеллект

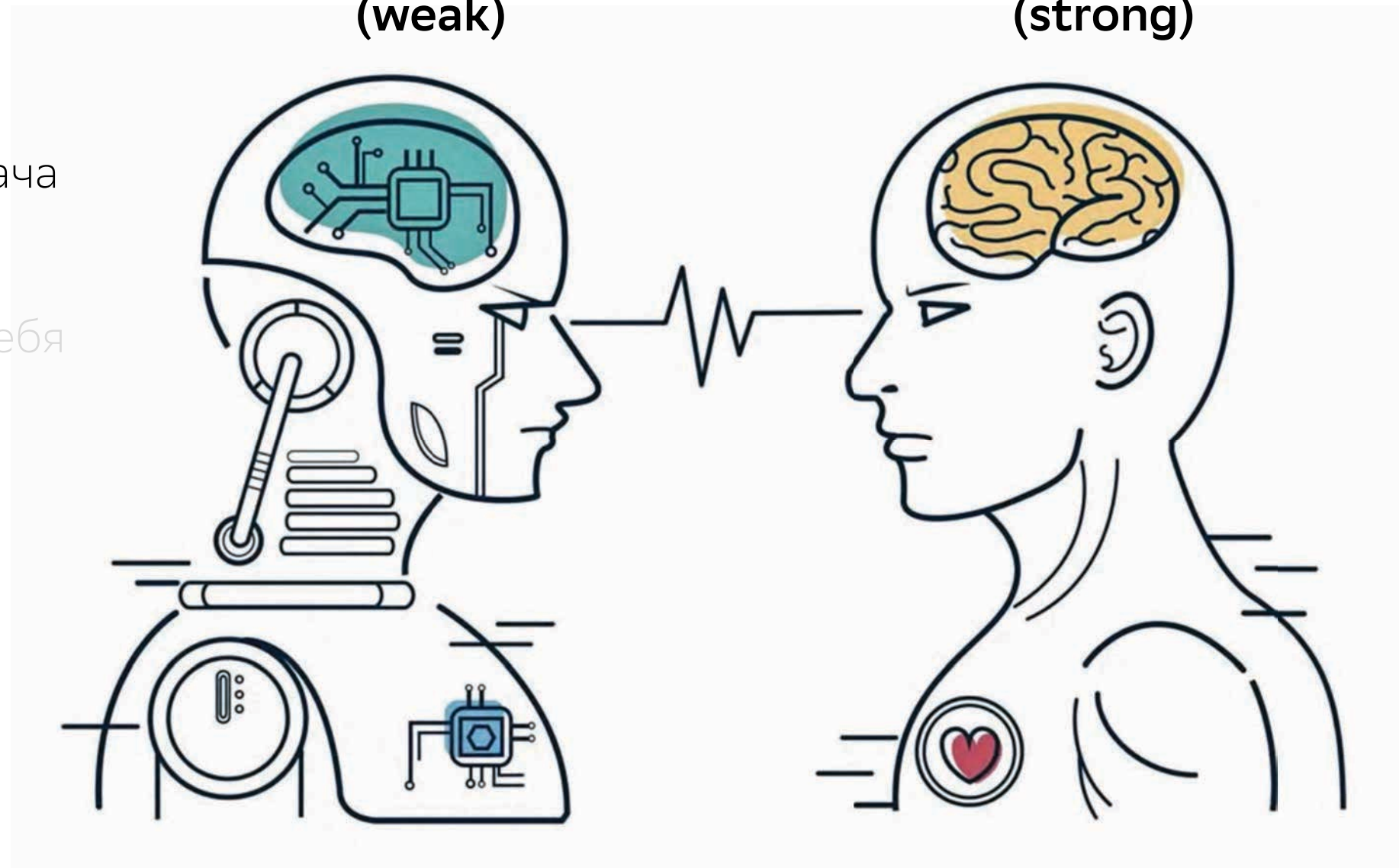


«Соответствующим образом запрограммированный компьютер с нужными входами и выходами и будет разумом, в том смысле, в котором человеческий разум — это разум.»

Дж. Серл. «Разумы, мозги и программы» 1980.

**Слабый
(weak)**

**Сильный
(strong)**



- Одна конкретная задача
- Обучен этой задаче
- Не осознает самого себя
- Не может мыслить независимо

- Решает множество задач с разными типами данных
- Легко адаптируется под новые задачи и домены
- Самосознание
- Саморефлексия
- Способность мыслить

Основные направления Generative AI:



Звуки

- Синтез голоса
- Генеративная музыка
- Пение



Изображения

- Текстуры для 3D
- Творчество
- Дизайн

Тексты

- Сюжеты
- Диалоги
- Программный код



3D

- Генерация 3D-объектов
- Оцифровка реальных объектов



Видео

- Музыкальные клипы
- Анимации
- Фильмы



02

Текст (языковые модели).

GigaChat

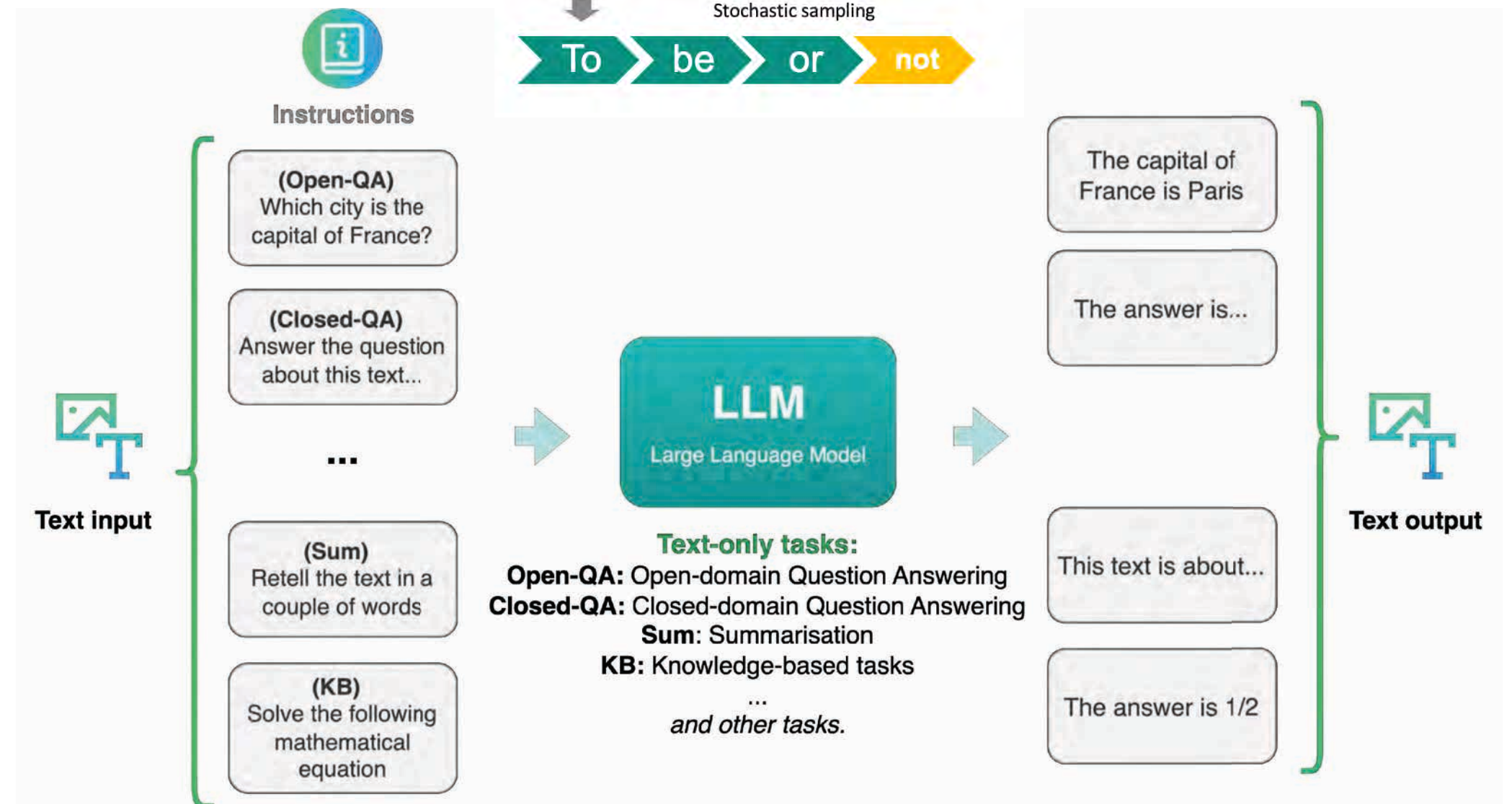
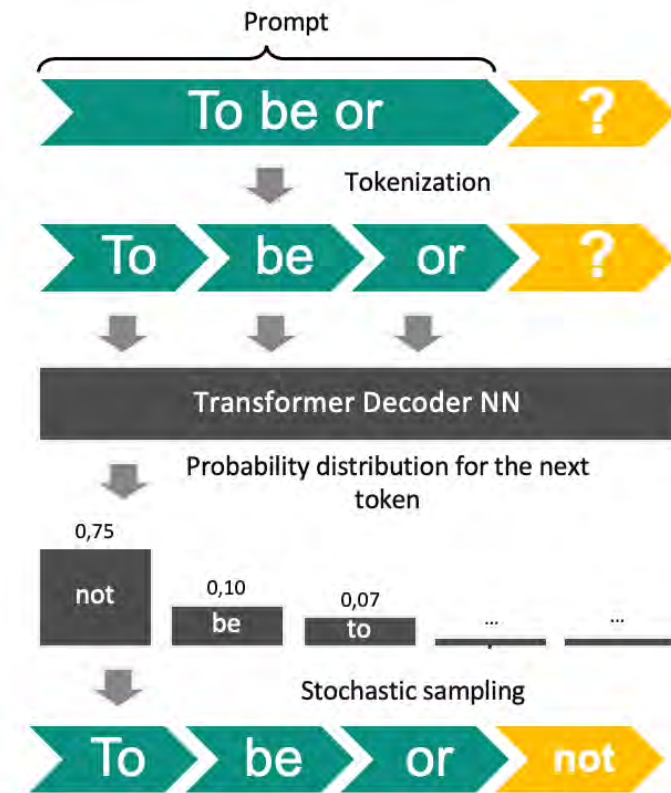
SEB
RE
RAI

Развитие LLM

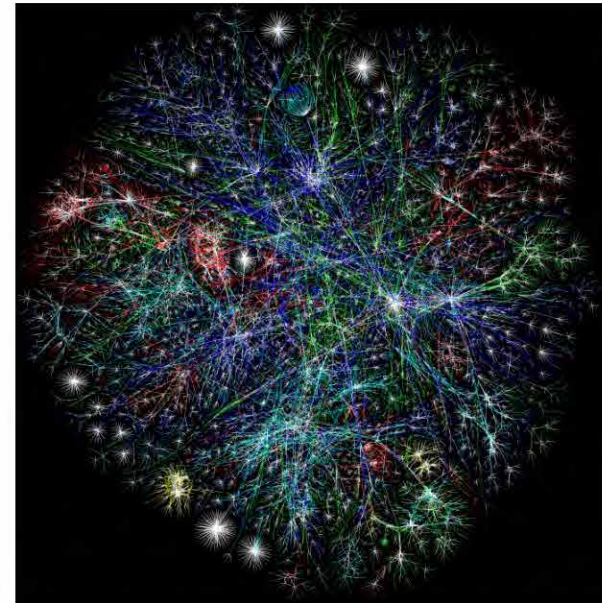


Почему большинство фундаментальных моделей работают с текстом?

- Текст – **основной и самый популярный источник знаний** человечества, который удобно передавать, редактировать, хранить
- Текст – **основа общения**
- Возможность общаться **ТЕКСТОМ** – основа любого AI-ассистента



Обучение LLM



Chunk of the internet,
~10TB of text



6,000 GPUs for 12 days, ~\$2M
~1e24 FLOPS



parameters.zip
~140GB file



- every ~year
- Stage 1: Pretraining**
1. Download ~10TB of text.
 2. Get a cluster of ~6,000 GPUs.
 3. Compress the text into a neural network, pay ~\$2M, wait ~12 days.
 4. Obtain **base model**.

*numbers for Llama 2 70B



<<USER>>
Can you write a short introduction about the relevance of the term "monopsony" in economics? Please use examples related to potential monopsonies in the labour market and cite relevant research.

<<ASSISTANT>>
"Monopsony" refers to a market structure where there is only one buyer for a particular good or service. In economics, this term is particularly relevant in the labor market, where a monopsony employer has significant power over the wages and working conditions of their employees. The presence of a monopsony can result in lower wages and reduced employment opportunities for workers, as the employer has little incentive to increase wages or provide better working conditions...

- every ~week
- Stage 2: Finetuning**
1. Write labeling instructions
 2. Hire people (or use [scale.ai](#)!), collect 100K high quality ideal Q&A responses, and/or comparisons.
 3. Finetune base model on this data, wait ~1 day.
 4. Obtain **assistant model**.
 5. Run a lot of evaluations.
 6. Deploy.
 7. Monitor, collect misbehaviors, go to step 1.

Текстовый диалоговый ассистент

Создание письма

Промпт

Напиши письмо моей собаке, чтобы она перестала грызть тапки

Результат

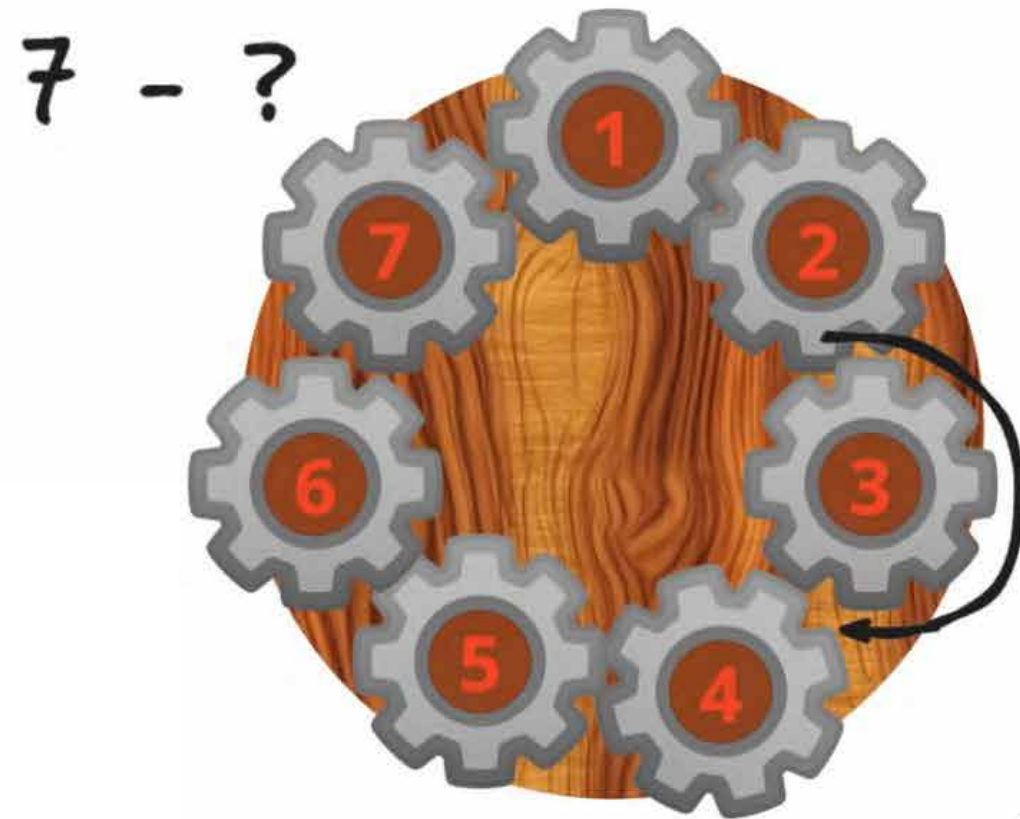
Привет, моя дорогая собака!

Я знаю, что ты любишь грызть тапки, но это может быть очень опасно для твоих зубов и здоровья. Я понимаю, что тебе скучно или хочется поиграть, но не стоит портить мебель. Пожалуйста, постарайся не грызть тапки хотя бы некоторое время. Если ты все же решишь продолжить, то лучше спрячься где-нибудь и погрызи что-то другое.

С любовью, [твое имя]

[Больше примеров →](#)

Сложные задачи



Потянешь?



Ян Лекун

ZH 7 axles are equally spaced around a circle. A gear is placed on each axle such that each gear is engaged with the gear to its left and the gear to its right. The gears are numbered 1 to 7 around the circle. If an attempt to rotate gear 3 clockwise was made, in which direction would gear 7 rotate?

Let's examine the rotation direction of each gear in sequence, starting with gear 3:

1. Gear 3 rotates clockwise (CW).
2. Since gears 3 and 4 are engaged, gear 4 rotates counterclockwise (CCW).
3. Gears 4 and 5 are engaged, so gear 5 rotates CW.
4. Gears 5 and 6 are engaged, so gear 6 rotates CCW.
5. Finally, gears 6 and 7 are engaged, so gear 7 rotates CW.

Therefore, gear 7 would rotate in the clockwise direction

AS you are thinking this wrong. I will give you a clue I did not say gear 3 rotates clockwise but we attempt to rotate it clockwise

Apologies for the misunderstanding. If we attempt to rotate gear 3 clockwise, considering that the gears are arranged in a circle and each gear is engaged with the gear to its left and the gear to its right, the following will happen:

1. We attempt to rotate gear 3 clockwise.
2. Gear 2 (engaged with gear 3) tries to rotate counterclockwise.
3. Gear 1 (engaged with gear 2) tries to rotate clockwise.
4. Gear 7 (engaged with gear 1) tries to rotate counterclockwise.
5. Gear 6 (engaged with gear 7) tries to rotate clockwise.
6. Gear 5 (engaged with gear 6) tries to rotate counterclockwise.
7. Gear 4 (engaged with gear 5) tries to rotate clockwise.

However, since gear 4 is also engaged with gear 3, gear 3's clockwise rotation would cause gear 4 to rotate counterclockwise. This creates a contradiction because we found that gear 4 tries to rotate clockwise due to the attempted rotation of gear 3.

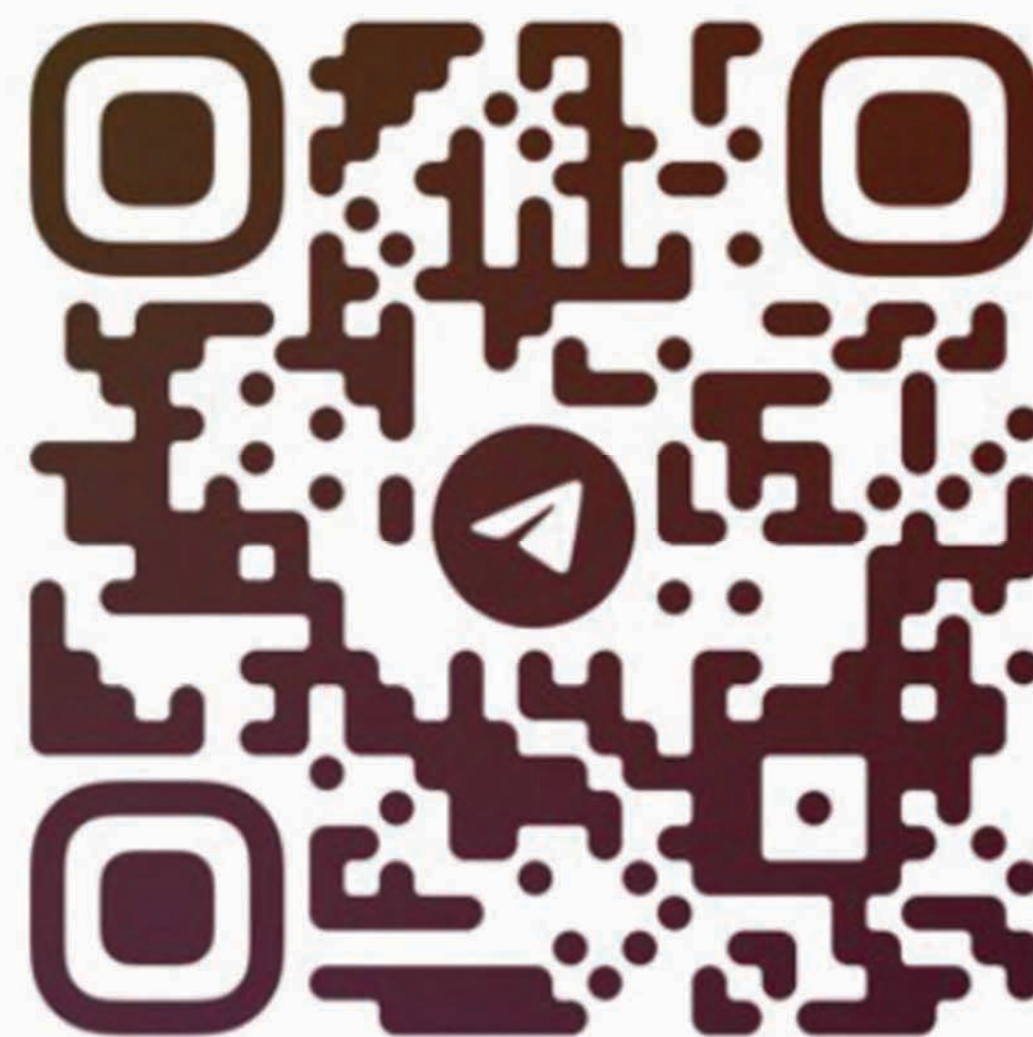
In this scenario, the system locks up because the gears cannot rotate without contradicting the direction imposed by the other gears. Thus, if someone attempts to rotate gear 3 clockwise, the entire system will not move, and gear 7 will not rotate.

Regenerate response

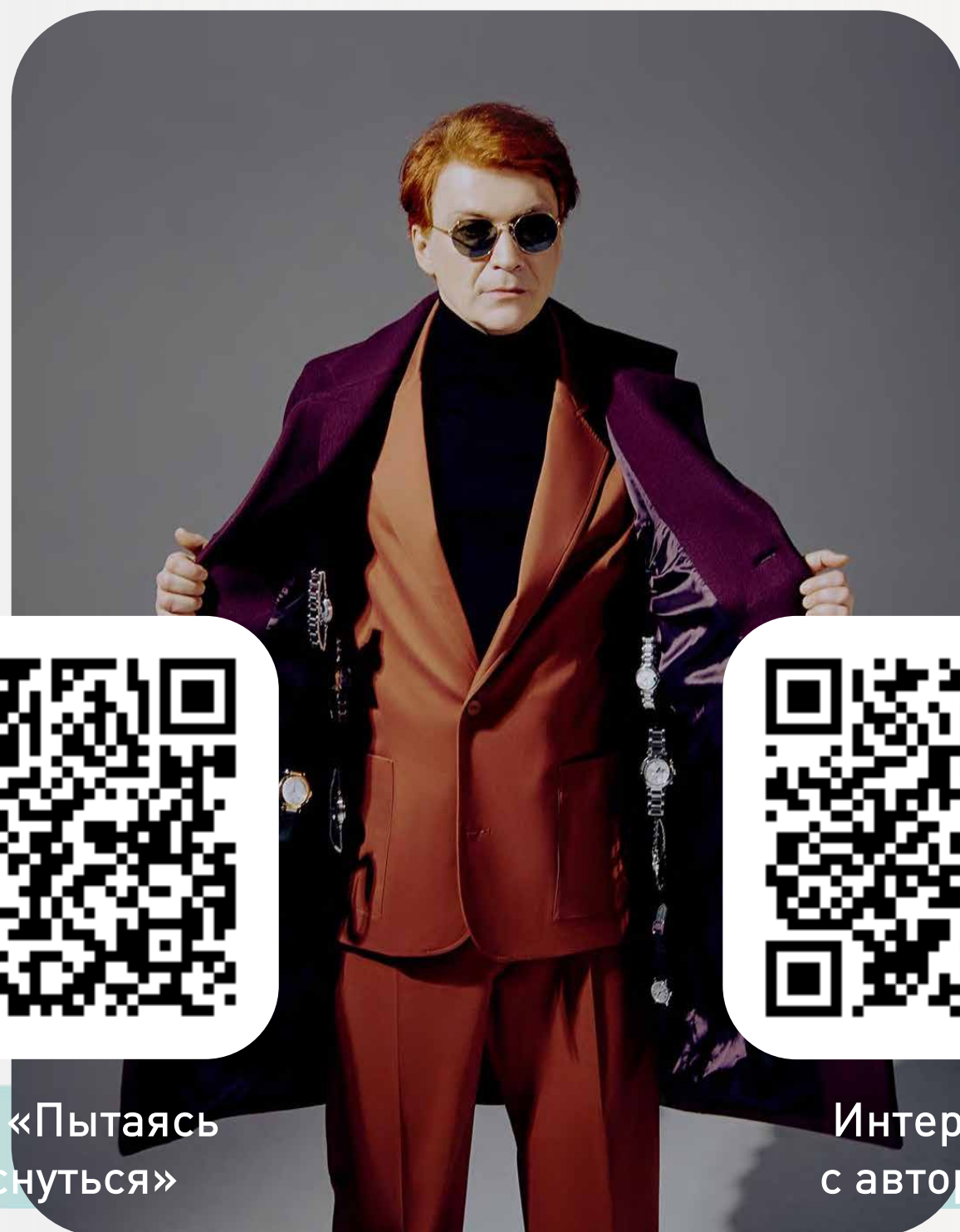
ГИГАЧАТ



- Ведет диалог
- Создает сюжеты и сценарии
- Пишет код
- Отвечает на вопросы
- Рисует картинки



@GIGACHAT_BOT



Книга «Пытаясь проснуться»



Интервью с авторами



Лето лучезарное в зелени стоит.
Солнышко янтарное над землёй летит.

В небе птицы весело радостно поют.
Соловьи разносят трель и цветы цветут.

Мамочка, родная, с юбилеем!
Стала ты у нас на год мудрее.
Пусть тебя согреет наша нежность,
И любви безмерной бесконечность.

Будь здоровой, милой и счастливой,
Доброй, нежной, ласковой, красивой!
Пусть улыбка светится в глазах,
И удача ждёт в твоих делах!



@VERSLIBRE_BOT

Иногда только текста недостаточно

— Еду в Петербург на два дня. Нашел фото с красивым дворцом, но не могу показать его тебе. Что это?

— К сожалению, без фотографии сложно определить, о каком дворце идет речь. В Санкт-Петербурге много красивых дворцов, включая Зимний дворец, Михайловский замок, Екатерининский дворец в Царском Селе и многие другие. Если вы сможете предоставить больше информации или описание дворца, я смогу помочь вам определить, о каком именно дворце идет речь.



03

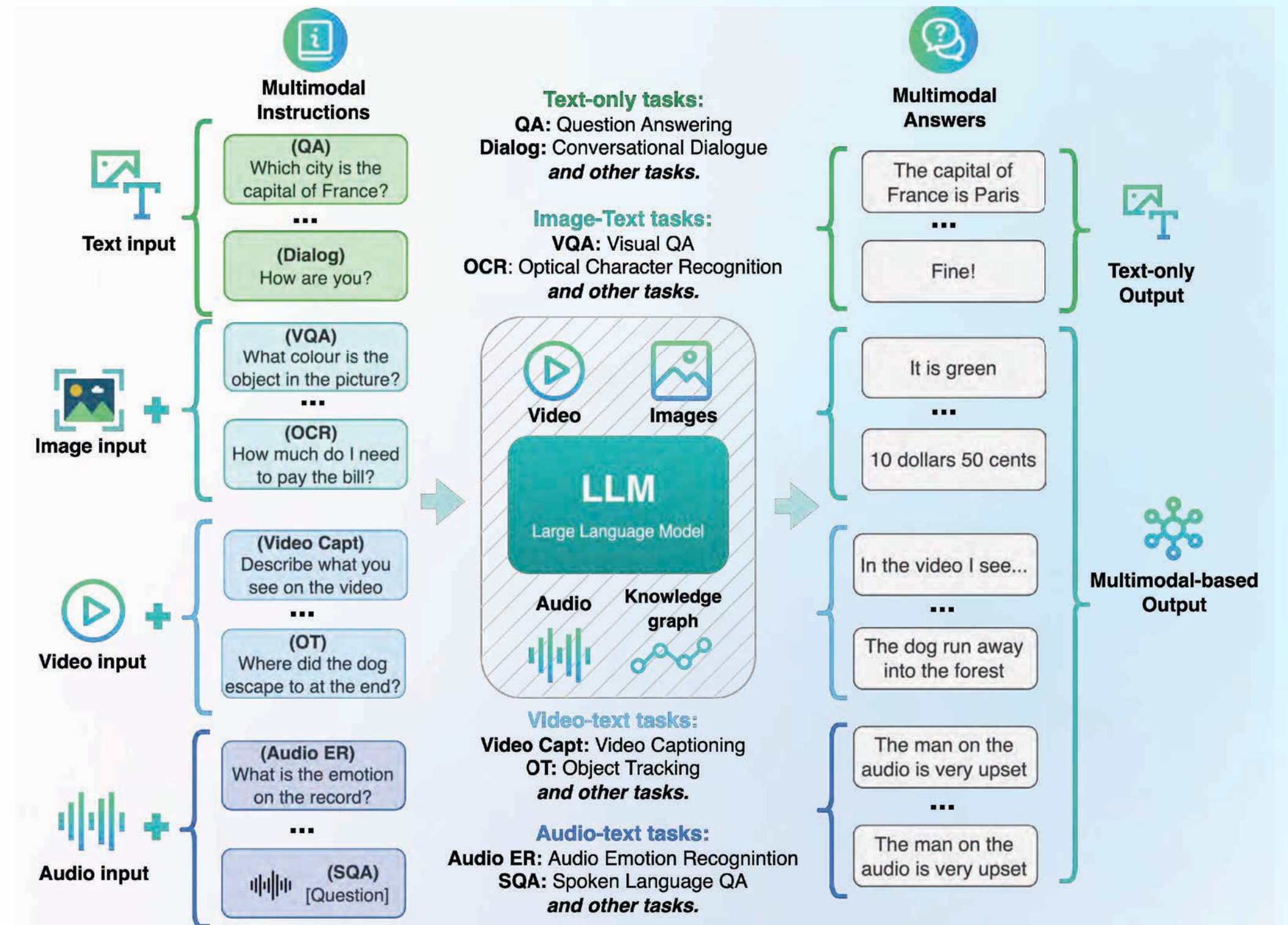
Мультимодальность
(языковые модели).
OmniFusion

SEB
RE
RAI

Language → Multimodality

Почему нужно нечто большее, чем только текст?

- Люди общаются между собой, используя **не только текст**
- Для создания качественного AI-ассистента нужно обучить его понимать **не только текст**
- **Текст + изображения + звук + видео + ...** — это наиболее реальный сценарий общения AI-ассистента с человеком



OmniFusion

Архитектура



Текст

Запрос:
Что здесь
изображено?

OmniFusion

Текст

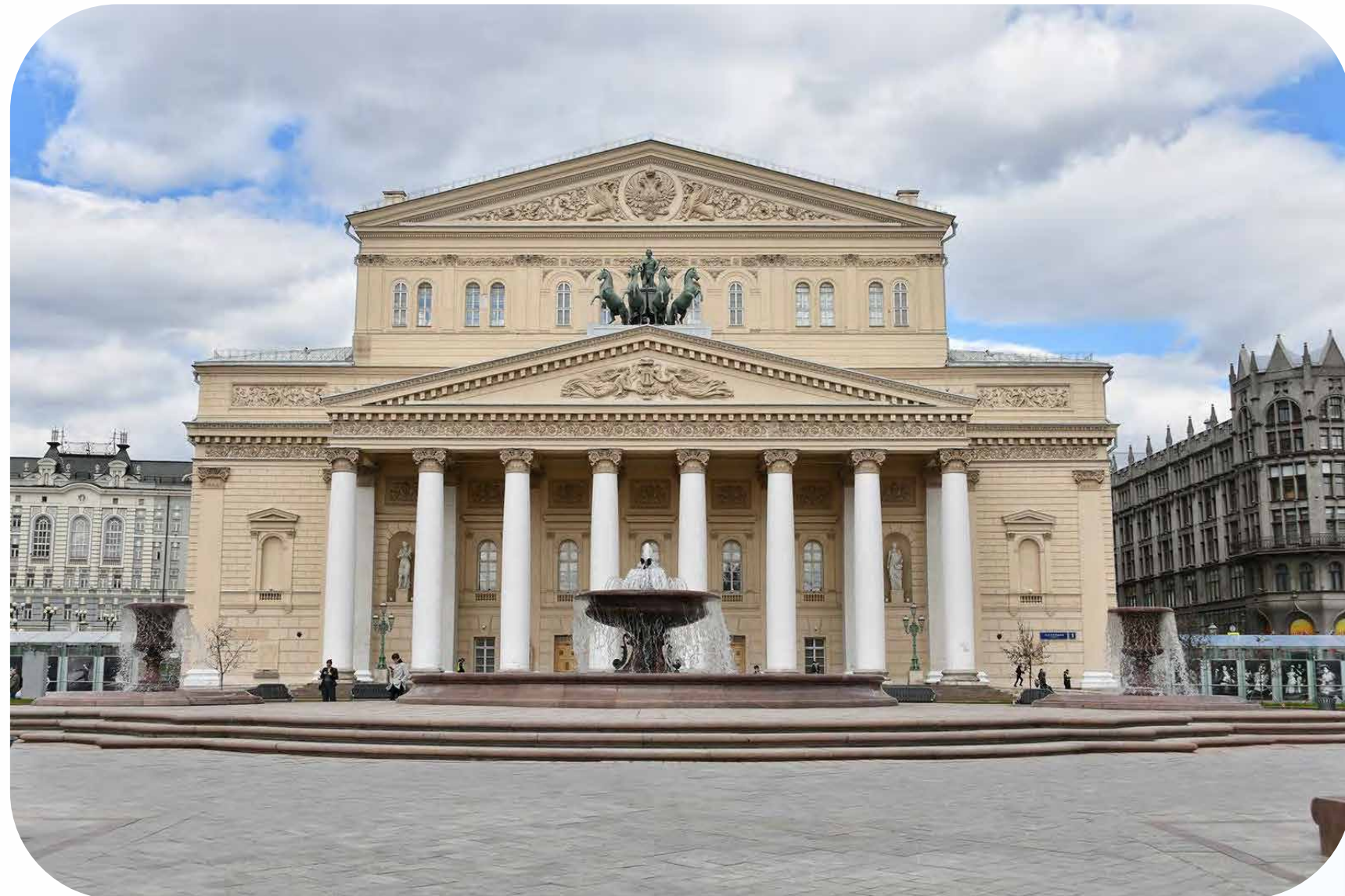
Ответ:
мраморная статуя
женщины с вуалью

Изображение



OmniFusion

Как помощник туриста



Что изображено на фото?



На снимке изображено большое, богато украшенное здание театра. Театр построен из камня, а его фасад украшен замысловатой резьбой, что придает ему величественный и впечатляющий вид

Как оно называется?

Большой Театр

В каком году был построен?

Большой Театр был построен в 1825 году

OmniFusion

Как помощник для спасения



Посмотри план эвакуации. Как мне покинуть здание?



Выход расположен с правой стороны этажа, и он отмечен зеленой стрелкой. Чтобы покинуть этаж, вам следует следовать за зеленой стрелкой и выйти из здания.

OmniFusion

Как помощник в кулинарии



Вот продукты, которые есть у меня в холодильнике.
Что посоветуешь приготовить?



Используя продукты из холодильника, вы можете приготовить разнообразные блюда на завтрак. Вот несколько идей:

1. Омлет с сыром
2. Омлет с ветчиной

Омлет с сыром кажется интересным вариантом.
Можешь подробно рассказать, как его готовить?

Конечно, есть подробные инструкции по приготовлению омлета с использованием ингредиентов из вашего холодильника:

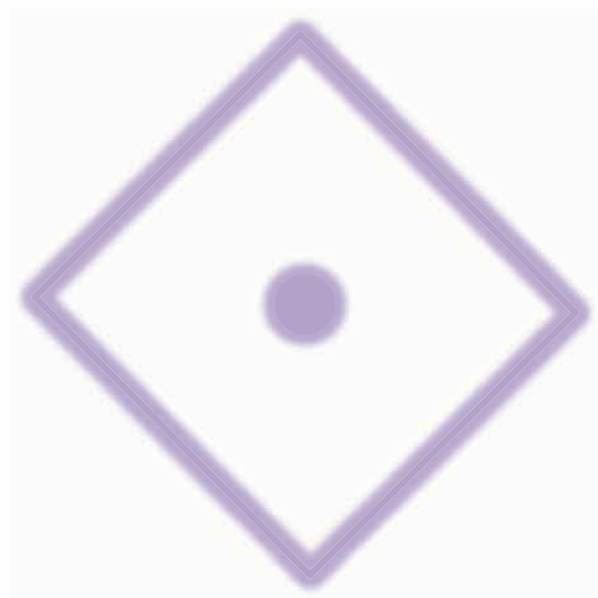
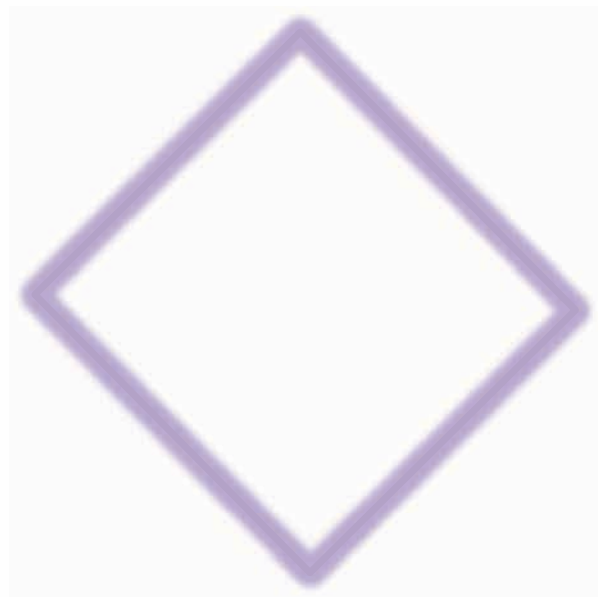
1. Разогрейте сковороду с антипригарным покрытием на среднем огне.
2. В небольшой миске взбейте вместе 2 яйца, 1 столовую ложку молока и щепотку соли и перца. Отложите.
3. Сбрызните сковороду оливковым маслом и перемешайте, чтобы оно покрыло дно.

...

MacBook Pro

OmniFusion

Умеет решать логические задачи



Вот три дополнительных изображения, расположенных в матрице 2x2. Посмотрите на два изображения в первом столбце, в левом верхнем углу и нижний левый. Затем найдите закономерность в первом столбце на основе двух левых изображений. Далее используем найденный шаблон и изображение в правом верхнем углу, чтобы сделать вывод о недостающей фигуре. Наконец, опишите, как будет выглядеть недостающее изображение?

Отсутствующее изображение в правом нижнем углу матрицы будет «звезда» с одной точкой

OmniFusion

Как помощник врача



Я ушиб ногу. Вот рентгеновский снимок. Поставь диагноз



На снимке изображен перелом

Какое мне подойдет лечение?

Лечение, скорее всего, будет включать иммобилизацию пальца гипсовой повязкой или шиной, чтобы он мог правильно зажить

OmniFusion

Распознавание изображений и описаний

Входные данные



Ван Гог. Красные виноградники в Арле. Монмажор

OmniFusion



Вопрос:

в каком стиле написана картина?

OmniFusion:

Картина написана в стиле импрессионизма.

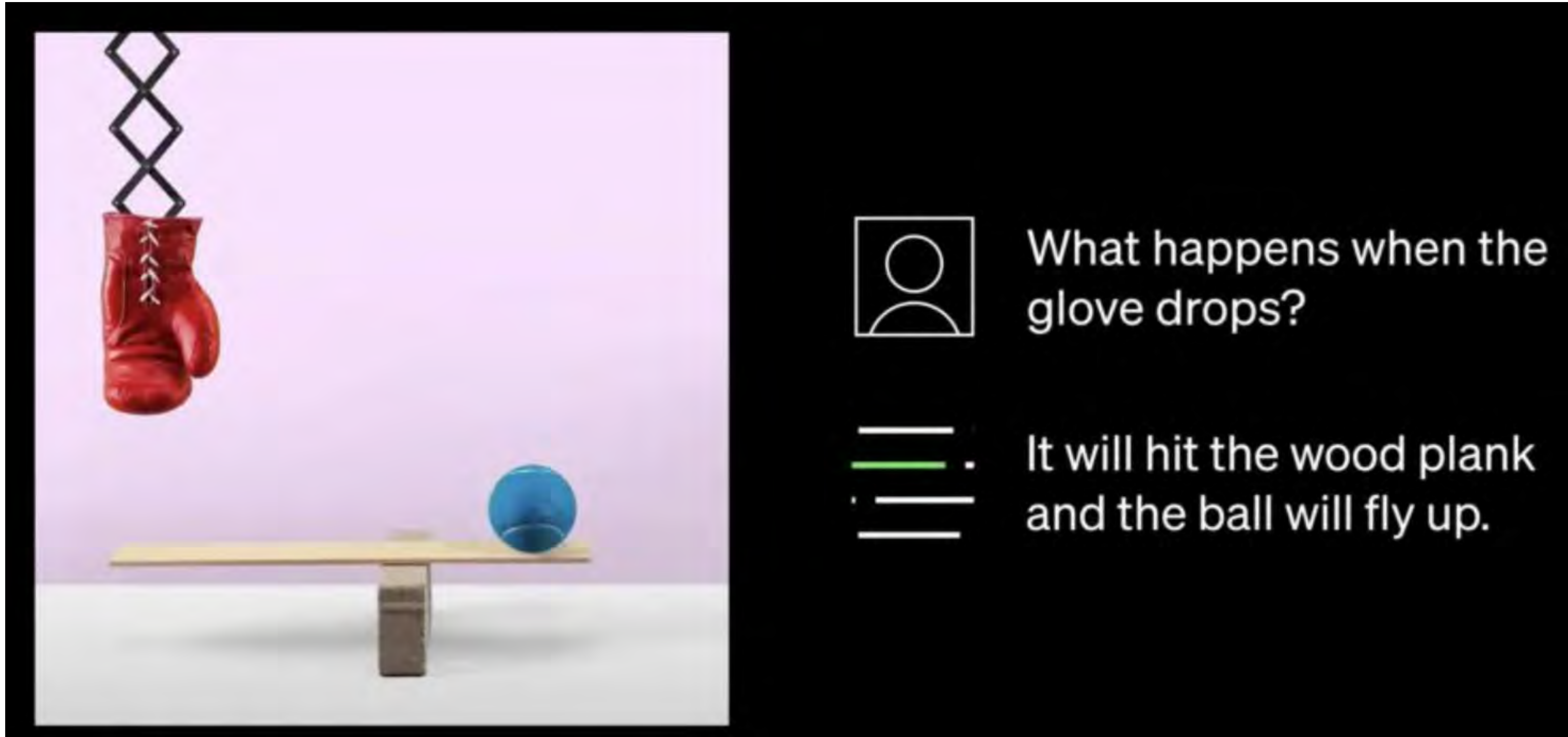
Вопрос:

в каком году она написана?

OmniFusion:

Картина была написана в 1888 году. Это дата, когда художник создал произведение искусства.

Сложные задачи



Kandinsky

Генеративные модели

Kandinsky: основные показатели и решенные задачи



11,9B

Параметры
Kandinsky 3.1

17,5B

Параметры
Kandinsky Video 1.0

101

Языки

+1M

Уникальных
пользователей за 4 дня

+15M

Уникальных
пользователей за год

+300M

Создано
изображений за год

Приложения: креативные компании
отдела маркетинга, SBOL, Imperial Porcelain
Factory, Gzhel, Kofemania, др.

Use:



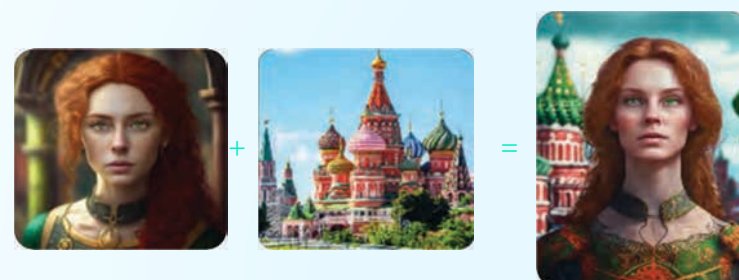
Создание изображений с помощью текста

"Red lighthouse at sunset in the sea" >



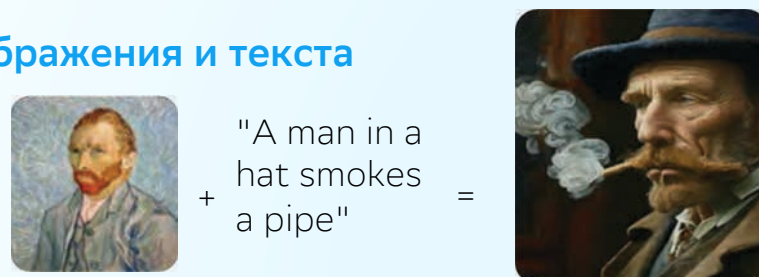
Смешивание изображений

Несколько
изображений

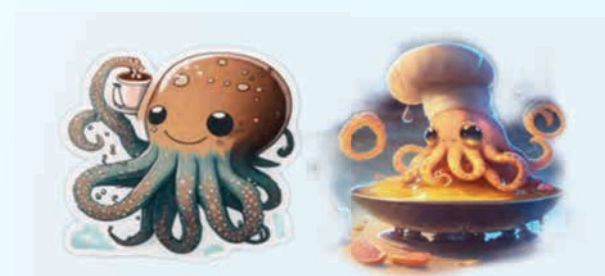


Смешивание изображения и текста

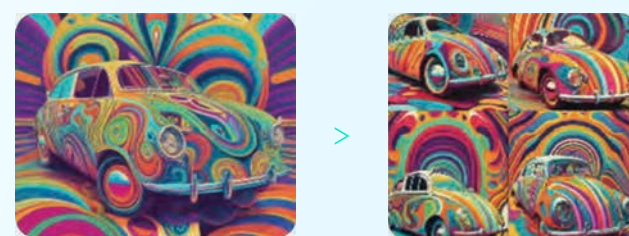
Изображение
в текст



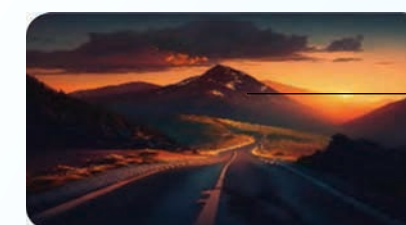
Создание стикеров и наклеек



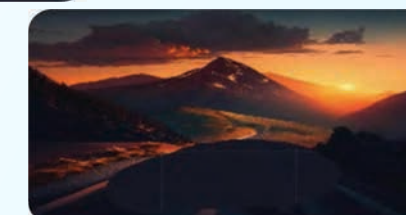
Вариации изображения



Редактирование изображений (inpainting)



"Пейзаж, горная дорога, закат"

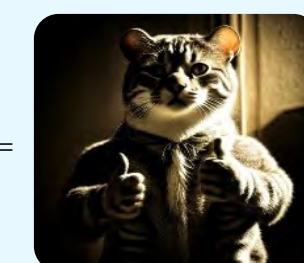


"Красный спортивный
автомобиль, стоящий на
дороге, яркие цвета, 4к,
высокая детализация"

Управляющая сеть



+ «Кот» =



Создание анимаций и видеороликов по тексту

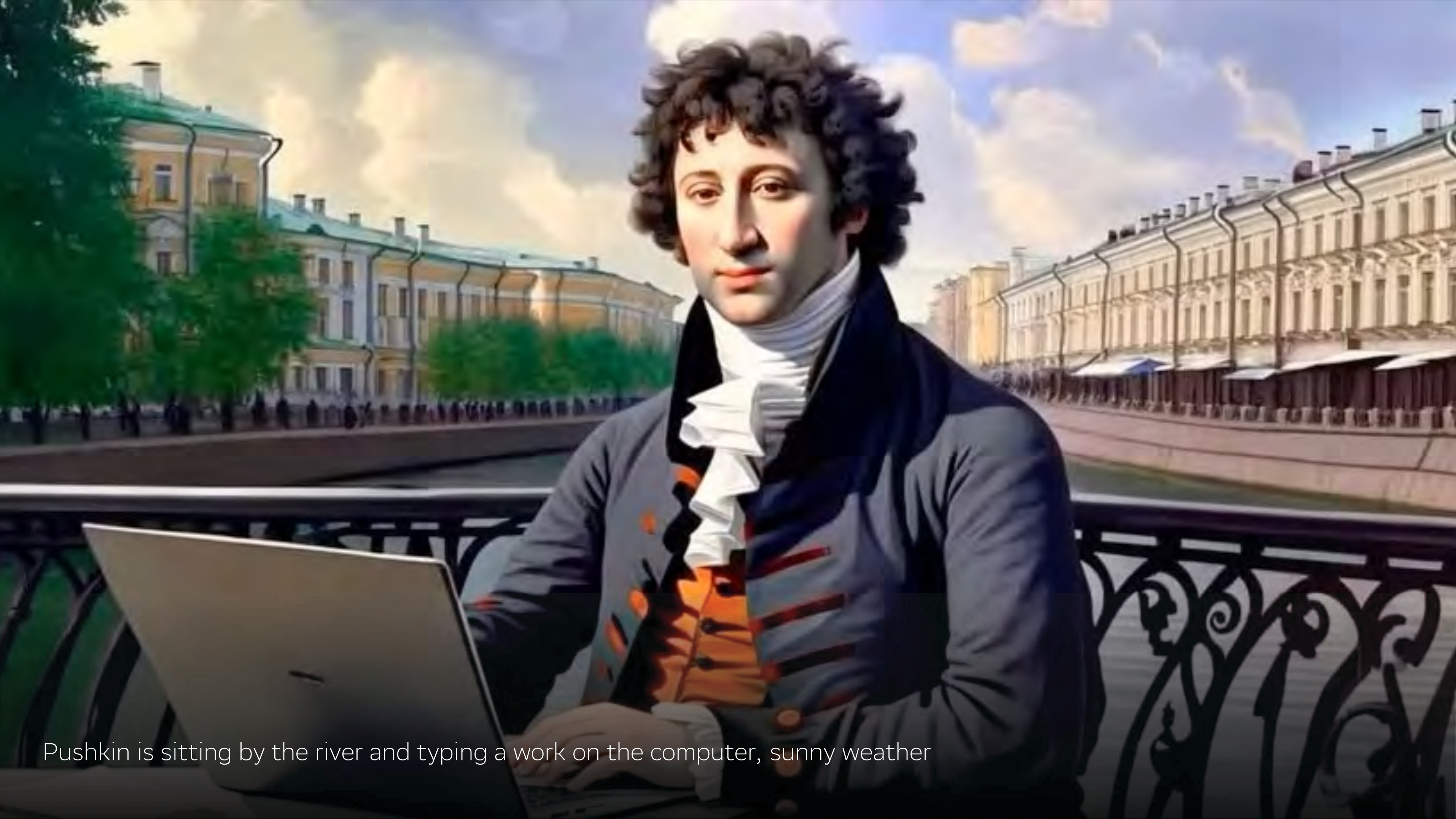




A young man riding a white horse, against a background of fields, nature all around him



A goldfish swims amidst a turbulent ocean, a moving world



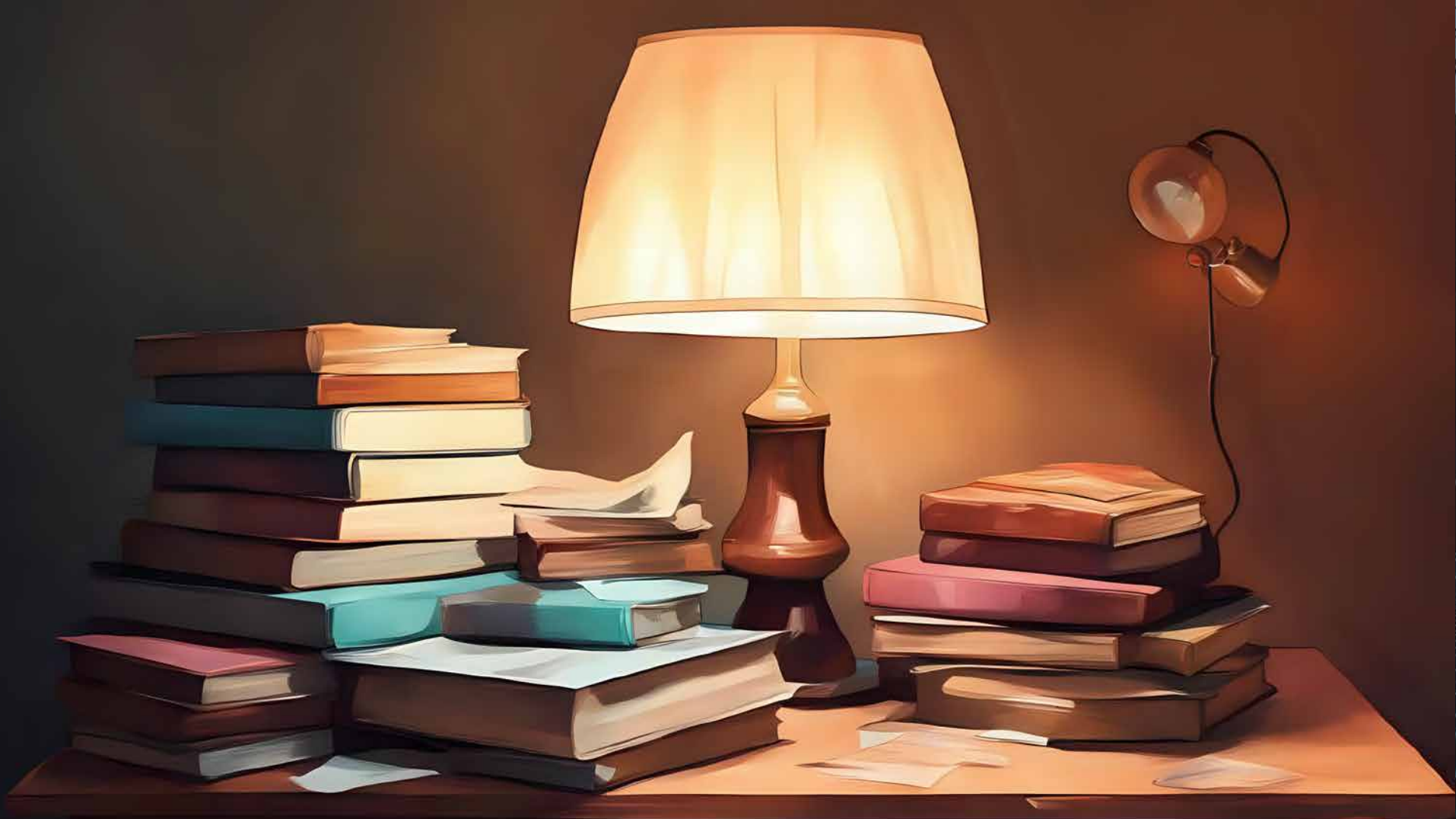
Pushkin is sitting by the river and typing a work on the computer, sunny weather



Swan with a crown on her head, fantastic lights twinkling, night time, surrounded by dark lake water



А.С. Пушкин. Москва 2024



Дорисовка изображения



**Дорисовка вне
границ изображения**

Дорисовка изображения

Дорисованный вариант

Оригинал



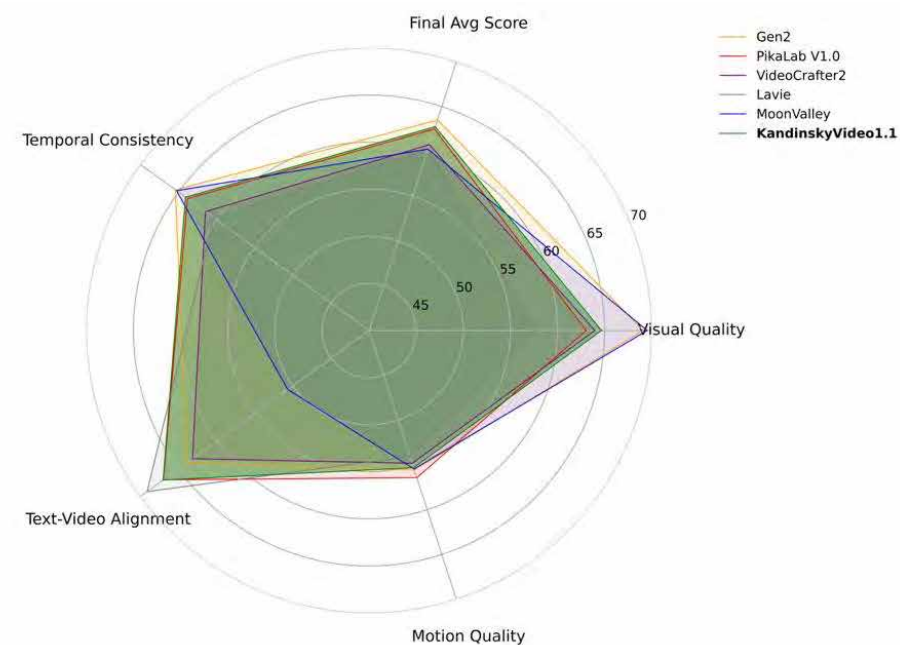
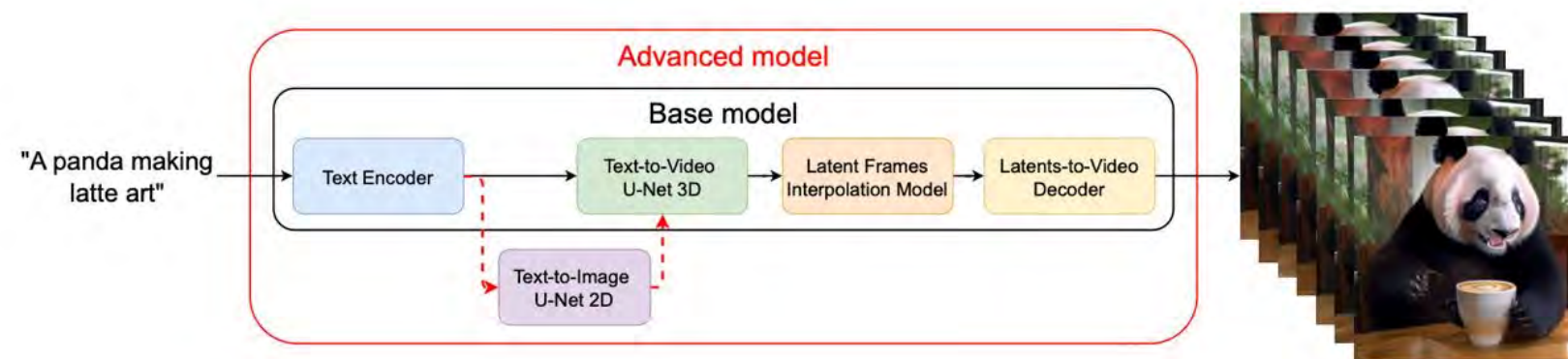
Vincent Willem van Gogh

Kandinsky



Kandinsky Video 1.1

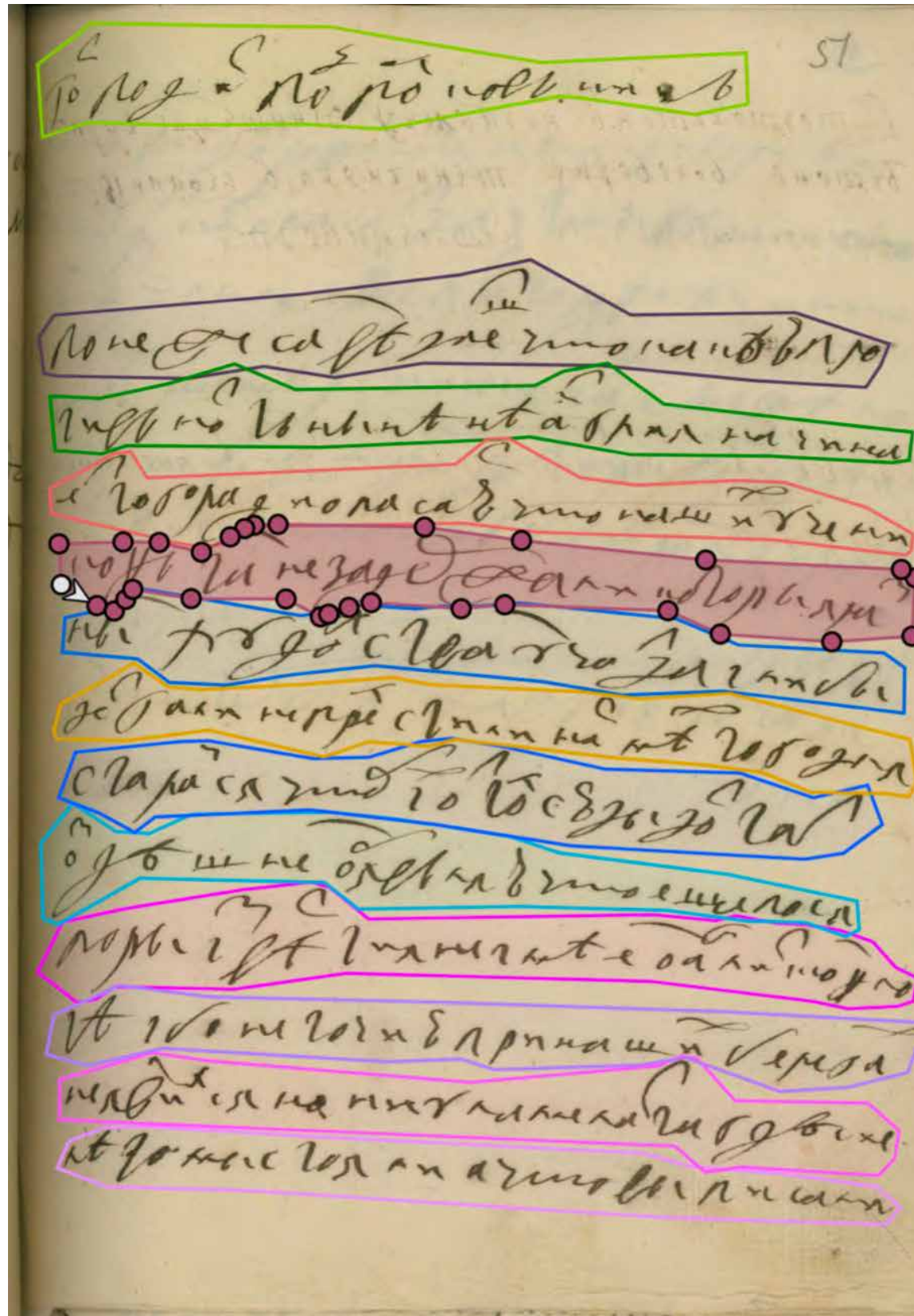
- Датасет для обучения ~5 млн пар «текст-видео»
- Генерация 5,5-секундного видео в разрешении 512x512 с FPS=30 занимает около 1,5 минут
- Можно выбирать FPS, динамичность видео (condition на motion score), соотношение сторон, начинать генерацию со стартового изображения
- ~20 000 уникальных пользователей (с момента релиза), 50 000 генераций



Автографы Петра Великого

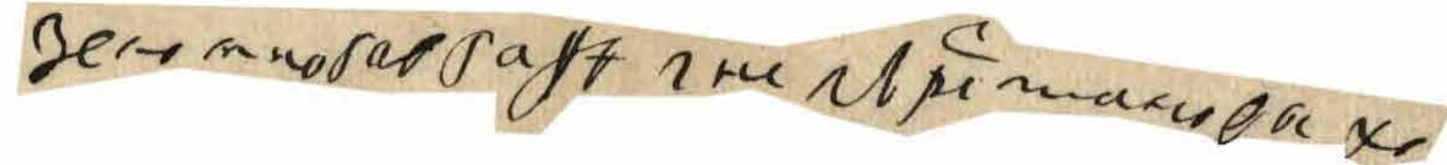
Чтение технологиями
искусственного интеллекта

Формирование датасета для обучения модели R AI



Пример размеченной страницы

the original text (1_1_10.jpg):



the translation (1_1_10.txt):

зело многа в гафѣ і не прѣстана выхо

Пример пары
сегментированная строка – её прочтение

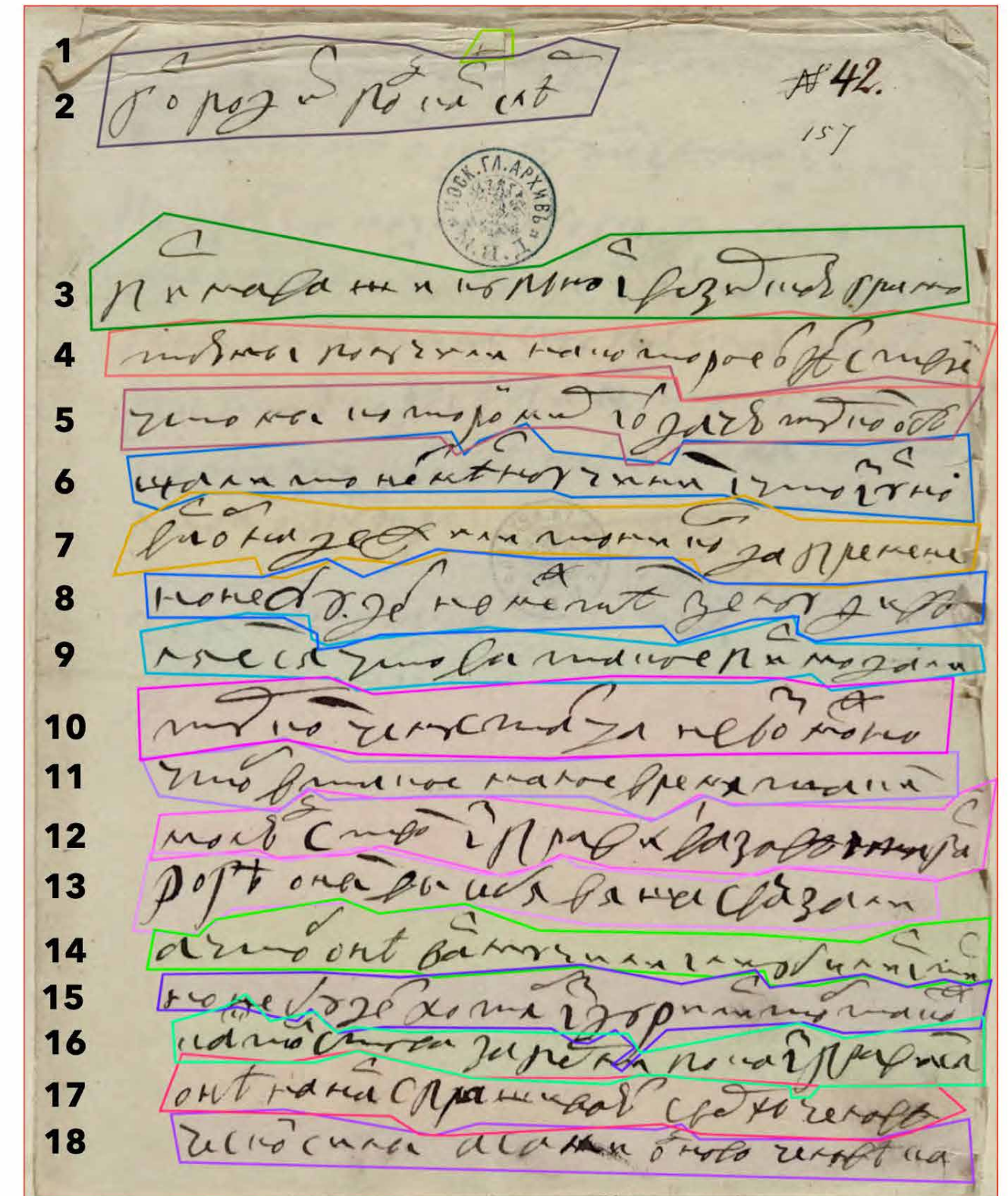
Всего было размечено:

9 656 строк

681 страница

Письмо Петра I —
П. П. Шафирову,
18 сентября 1711 г. //
РГАДА. Ф. 142.
Оп. 2.
Д. 7. Л. 157.

- 1 +
- 2 господин подканслѣр
- 3 писма ваши купно і с везирскою грамо
- 4 тою мы получили на которое отвѣтствуем
- 5 что мы которой мир і отдачу турком обѣ
- 6 щали то не тѣнно учиним і что ізусно
- 7 вас обнадежили то никогда премене
- 8 но не будет но меж тѣм зело удивѣ
- 9 ляемся что вы такое письмо дали
- 10 турком чему статца невозможно
- 11 чтоб в такое малое время таким
- 12 молюдством ісправит в азовѣ і таган
- 13 рогѣ оным вы себя вяще связали
- 14 а чтоб онѣ вас мучили іли убили істин
- 15 но не будет хотя б і здурилис то б також
- 16 как толстова заперли пока ісправимся
- 17 онѣ на нас спрашивают сверхъ челоуѣ
- 18 ческой силы а сами одново челоуѣка



Создание модели распознавания почерка Петра I

$$\text{CER} = \frac{\sum_{i=1}^n \text{dist}_c(\text{pred}_i, \text{true}_i)}{\sum_{i=1}^n \text{len}_c(\text{true}_i)}$$

$$\text{WER} = \frac{\sum_{i=1}^n \text{dist}_w(\text{pred}_i, \text{true}_i)}{\sum_{i=1}^n \text{len}_w(\text{true}_i)}$$

$$\text{StringAcc} = \frac{\sum_{i=1}^n [\text{pred}_i = \text{true}_i]}{n}$$

Качество распознавания
1-CER

№	Команды (62)	CER	WER	String Accuracy	Time	Последний	Входные данные
1	ocrv	2,386	13,853	63,215	201	1y	140
2	Vladislav Kramarenko	2,953	17,029	56,972	553	1y	183
ооо	azaza	2,997	14,816	61,55	292	1y	47
3	Magic City	3,374	19,579	52,601	54	1y	51
4	NotPeter	3,477	19,381	52,341	109	1y	4
5	Vitaly Kudelya	3,513	19,44	53,486	417	1y	44
6	Daniil Anastasyev	3,67	19,192	51,717	419	1y	5
7	Sergey Dulikov	3,811	21,008	49,792	468	1y	26
8	Irina Gorbunova	4,761	24,511	45,682	399	1y	32
9	Da Hong Pao	4,99	29,026	38,606	167	1y	28
10	Nikolai Sosedov	5,374	28,848	37,773	195	1y	1

Модель обучена на 6 196 строках

97.6%

Digital Пётр. Распознавание рукописей Петра I с помощью компьютерного зрения.

[Электронный ресурс]. Режим доступа: <https://www.fusionbrain.ai/digital-petr/>

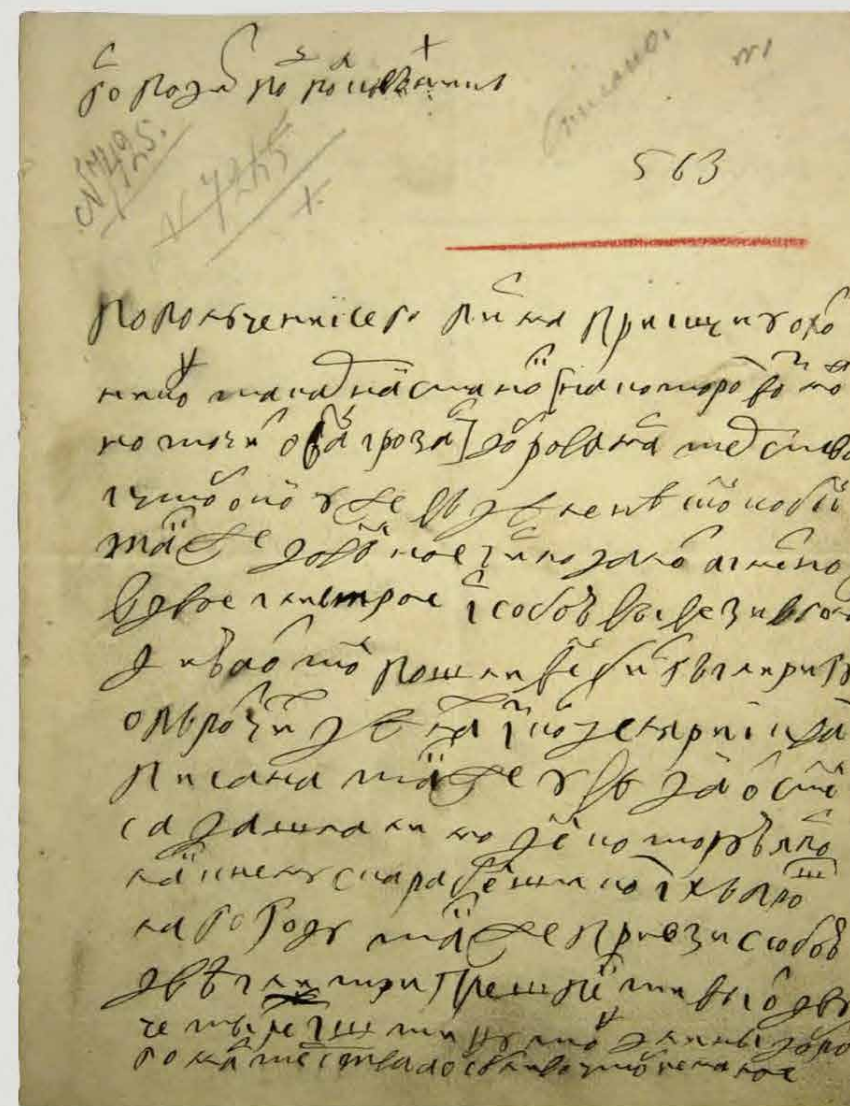
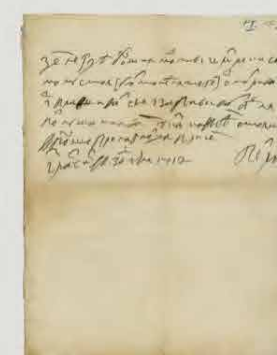
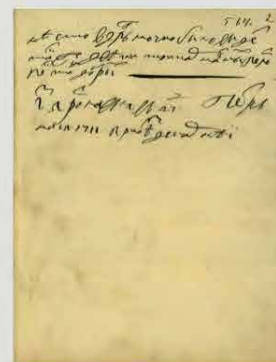
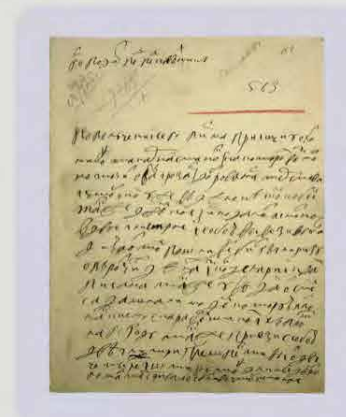


КАК ЭТО РАБОТАЕТ

О ПРОЕКТЕ

ПАРТНЁРЫ

ЗАГРУЗИТЬ ДОКУМЕНТ



Письмо Петра I из Ярославля Борису Ивановичу Куракину в Лондон, 27 мая 1711г.

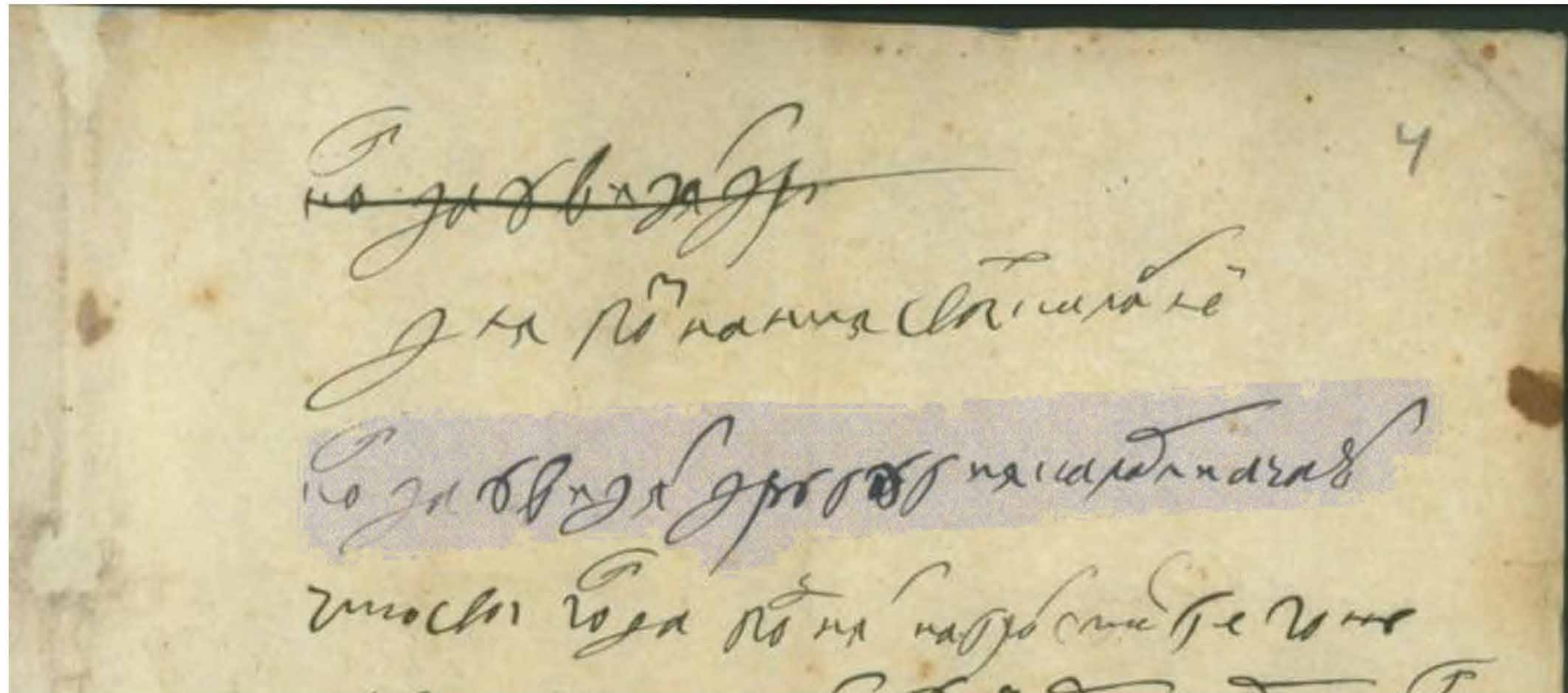
Господин подполковъникъ.

По полученіи сего письма приищи у охот-
никоф такарнай станок [на котором возмож-
но точит овал і розан] доброва мастерства,
і чтоб оной уже въ дѣле нѣсколко был,
также доволное число далот, а іменно,
вдвое іли втрое, і с собою вывези в Голан-
дию, а оттол пошли в Елбингъ іли Ригу.
О пѣротчих дѣлах із концеляриі к вам
писана. Также увѣдай от Стел-
са, дашла ли модел, которую я пос-
лал к нему с карабелшиком іхъ прош-
лаго году. Также прив(е)зи с собою
двѣ іли три прешпективы от дву,
четырех і шти футоф длины добро-
го мастерства, а особливо чтоб немалое...

Digital Пётр. Распознавание рукописей Петра I с помощью компьютерного зрения.

[Электронный ресурс]. Режим доступа: <https://www.fusionbrain.ai/digital-petr/>





=

2 для познания своих караблей

3 **когда увидят другогия карабли а чают**

4 что свои тогда поднят на гра станге тому

5 кто спрашивает бѣлой ординарной флаг

6 а тому кому отвѣчат красной ординарной

7 на том же мѣстѣ а потом тому кто отвѣ

8 чал поднят синей н бѣлой на форстенгѣ

9 а во просителю отвѣчат поднят красной орди

10 нарной на форстенгѣ

11 когда вофросят откуда карабль

12 отвѣчат из питербурха потом паки

13 спросит что лозонъ отвѣчат бог он

- 1 когда увидат дру
- 2 для познания своіх караблей
- 3 когда увидят другугия карабли а чают
- 4 что своі тогда поднят награ станге тому
- 5 хто спрашивает бѣлой ординарной флаг
- 6 а тому кому отвѣчат красной ординарной
- 7 на том же мѣстѣ а потом тому хто отвѣ
- 8 чал поднят синей н бѣлой на форстенгѣ
- 9 а во просителю отвѣчат поднят красной орди
- 10 нарной на форстенгѣ
- 11 когда вофросят откуды карабль
- 12 отвѣчат із питербурха потом паки
- 13 спросит что лозонь отвѣчат бог он

увидят

на грот-стенге

вопросителю

с нами

4

- 1 ~~когда увидят дру~~
- 2 для познания своіх караблей
- 3 когда увидят другугия карабли а чают
- 4 что своі тогда поднят награ станге тому
- 5 хто спрашивает бѣлой ординарной флаг
- 6 а тому кому отвѣчат красной ординарной
- 7 на том же мѣстѣ а потом тому хто отвѣ
- 8 чал поднят синей н бѣлой на форстенгѣ
- 9 а во просителю отвѣчат поднят красной орди
- 10 нарной на форстенгѣ
- 11 когда вофросят откуды карабль
- 12 отвѣчат із питербурха потом паки
- 13 спросит что лозонь отвѣчат бог он

Транскрибирование текста ИИ и человеком



Компьютерная программа

когда увидат дру
для познания своіх караблей
когда увидят другугия карабли а чают
что своі тогда поднят награ станге тому
кто спрашивает бѣлой ординарной флаг
а тому кому отвѣчат красной ординарной
на том же мѣстѣ а потом тому хто отвѣ-
чал поднят синей н бѣлой на форстенгѣ
а во просителю отвѣчат поднят красной орди-
нарной на форстенгѣ
когда вофросят откуда карабль
отвѣчат із питербурха потом паки
спросит что лозонь отвѣчат бог он

Человек

Когда увидят дру
Для познания своіх караблей.
Когда увидят другугия карабли, а чают
что своі, тогда поднят на гротстенге тому,
кто спрашивает, белой ординарной флаг.
А тому, кому отвѣчат, красной ординарной
на том же мѣстѣ. А потом тому, хто отвѣ-
чал, поднят синей н белой на форстенгѣ,
а вопросителю отвѣчат поднят красной орди-
нарной на фостенгѣ.
Когда вофросят, откуда карабль ,
отвѣчат із Питербурха. Потом паки
спросит, что лозонь, отвѣчат Бог с нами.

Спасибо



Денис Димитров

Управляющий директор
по исследованию данных, Sber AI,
Руководитель группы Generative AI
Научный консультант, AIRI

