

Сравнение различных LLM без ограничений для ответа.

Сравнение моделей будет происходить по следующим метрикам: величина контекстного окна, количество параметров, очки MT-bench (набор стандартных текстовых пакетов, используемых для качества машинного перевода систем), очки MMLU (бенчмарк для измерения знаний, приобретенных во время предварительного обучения, путем оценки моделей исключительно в условиях нулевого и кратковременного тестирования).

Сравниваются 4 популярных LLM из семейства llama, а также gemma из семейства gemma. В качестве “эталонных” данных используется информация о ChatGPT-4. Данные о результатах бенчмарков получены с [данного сайта](#).

Таблица 1 – Сравнение LLM моделей семейства llama и ChatGPT-4.

Модель	Токены	Параметры	MT-bench	MMLU
llama2	4K	7B	6.27	45.8
mistral	4K	7B	7.6	>55.4
gemma	4K	9B	-	64.3
vicuna	16K	7B	6.22	49.8
openchat	4K	7B	7.81	64.3
ChatGPT-4	8K	1.73T	9.18	>86.4

Также стоит упомянуть, что vicuna и openchat полноценно поддерживают русский язык.

Далее представлен практический тест моделей, основанный на данных, полученных страниц МОЭМВ вики, полученных из парсера. Будет использоваться следующий датасет данных и вопросов:

1. info – Регистрация первокурсников – 1523 символов – “Как зарегистрироваться на сайте моэмв?”

2. info – Научно-технический семинар 2024 (бывш. конференция ППС) – 3579 символов – “Какие сроки сдачи у научной статьи?”

3. 3 курс – Искусственные нейронные сети – 3735 символов – “Какие темы будут изучаться на занятиях?”

1. llama2

Тест 1.

Добро пожаловать на кафедру МОЭВМ! Для того, чтобы комфортно себя чувствовать во время обучения и своевременно получать нужную информацию, а также выполнять и сдавать задания по некоторым дисциплинам, необходимо выполнить следующие действия по регистрации:

1. Завести аккаунт на Discord, указав свои реальные имя и фамилию, установить фото в качестве аватара, скачать клиент.

2. Завести аккаунт на GitHub, указав свои реальные имя и фамилию в поле Name, установить фото в качестве аватара.

3. Завести аккаунт на <https://e.moevm.info>, указав свои реальные имя и фамилию в полях Имя и Фамилия, установить фото в качестве аватара.

4. Заполнить свои данными на <https://forms.gle/SDDN3ss4o5cnYsop7>.

Основная страница кафедры: [<http://moevm.info>](http://moevm.info)

VK: [<https://vk.com/moevm_leti>](https://vk.com/moevm_leti)

Email для связи с администрацией: [<ask\(at\)moevm.info>](mailto:ask(at)moevm.info).

Прочитайте справку по синтаксису Wiki и попробуйте свои силы без ущерба для проекта в песочнице.

Общие сведения и рекомендации по работе с Wiki:

* Правила организации информации в Wiki;

* Регистрация первокурсников – МОЭВМ Вики.

Надеюсь, это поможет вам успешно зарегистрироваться на кафедре!

Тест 2.

Научно-технический семинар кафедры МО ЭВМ 2024 (бывш. конференция ППС) пройдет в течение нескольких месяцев, и сроки сдачи научной статьи могут отличаться в зависимости от конкретного года и программы мероприятия.

В 2017–2024 годах для подачи материалов конференции мы используем систему EasyChair, в которую необходимо сначала зарегистрироваться, а затем проследовать по ссылке указанной ниже для подачи статьи.

Основные даты научного семинара для 2024 года выглядят следующим образом:

- * Начало приема статей: 15 декабря 2023 года;
- * Окончание приема статей: 16 января 2024 года (включительно);
- * Предоставление рецензии (результатов рецензирования статьи программным комитетом) : 24 января 2024 года;
- * Дни конференции: 1–2 февраля 2024 года – выступления участников с докладами;
- * Прием исправленных версий статей: до московского времени 6 февраля 2024 года.

Важно помнить, что сроки могут изменяться в зависимости от конкретных условий и программы мероприятия, поэтому рекомендуется проверять на сайте конференции заднимыслями для последних изменений.

Тест 3.

Вот некоторые из тем, которые могут быть рассмотрены на занятиях по искусственным нейронным сетям:

1. Introduction to Deep Learning:

- * What is deep learning?
- * History of deep learning

- * Types of deep learning models (e.g., feedforward neural networks, recurrent neural networks, convolutional neural networks)

2. Artificial Neural Networks:

- * Introduction to artificial neural networks
- * Types of artificial neural networks (e.g., perceptron, multilayer perceptron, backpropagation)
- * Training and testing artificial neural networks

3. Convolutional Neural Networks:

- * Introduction to convolutional neural networks
- * Architecture of convolutional neural networks
- * Applications of convolutional neural networks (e.g., image classification, object detection)

4. Recurrent Neural Networks:

- * Introduction to recurrent neural networks
- * Types of recurrent neural networks (e.g., simple RNNs, LSTMs, GRUs)
- * Applications of recurrent neural networks (e.g., language modeling, speech recognition)

5. Autoencoders:

- * Introduction to autoencoders
- * Types of autoencoders (e.g., feedforward autoencoders, recurrent autoencoders)
- * Applications of autoencoders (e.g., image compression, anomaly detection)

6. Generative Adversarial Networks:

- * Introduction to generative adversarial networks
- * Architecture of generative adversarial networks
- * Applications of generative adversarial networks (e.g., image generation, video prediction)

7. Natural Language Processing:

- * Introduction to natural language processing
- * Types of natural language processing tasks (e.g., text classification, sentiment analysis, machine translation)
- * Applications of natural language processing (e.g., chatbots, language translation software)

8. Recommendation Systems:

- * Introduction to recommendation systems

- * Types of recommendation systems (e.g., content-based filtering, collaborative filtering)

- * Applications of recommendation systems (e.g., movie recommendations, product recommendations)

9. Time Series Analysis:

- * Introduction to time series analysis

- * Types of time series analysis (e.g., ARIMA, LSTM)

- * Applications of time series analysis (e.g., stock market prediction, weather forecasting)

10. Transfer Learning and Fine-tuning:

- * Introduction to transfer learning and fine-tuning

- * How to use pre-trained models for a specific task

- * Applications of transfer learning and fine-tuning (e.g., image classification, natural language processing)

These are just some examples of the topics that may be covered on the course. The exact curriculum will depend on the instructor and the level of the students.

2. mistral

Тест 1.

Чтобы зарегистрироваться как первокурсник на кафедре МОЭВМ, необходимо выполнить следующие шаги:

1. Если у вас еще нет аккаунта, создайте его на платформе Discord по адресу <<https://discord.com>>. Укажите ваше реальное имя и фамилию, установите фотографию в качестве аватара и скачайте клиент.

2. Если у вас еще нет аккаунта, создайте его на GitHub по адресу <<https://github.com>>. Укажите ваше реальное имя и фамилию в поле "Name", установите фотографию в качестве аватара.

3. Если у вас еще нет аккаунта, зарегистрируйтесь на сайте кафедры по адресу <<https://e.moevm.info>>. Укажите ваше реальное имя и фамилию в полях "Имя" и "Фамилия", установите фотографию в качестве аватара.

4. Заполните свои данными формой по адресу <<https://forms.gle/SDDN3ss4o5cnYsop7>>.

Ресурсы кафедры:

- * Основная страница: <<http://moevm.info>>
- * ВКонтакте: <https://vk.com/moevm_leti>
- * Email для связи с администрацией:
[ask@moevm.info] (mailto:ask@moevm.info)
- * МОЭВМ Вики: <<https://start.moevm.wiki/>>

Для получения дополнительной информации о синтаксисе и работе с Wiki, посетите страницу "Справка по синтаксису" и "Песочница". Ознакомьтесь также со "Справочником по организации информации в Wiki", где приведены правила для регистрации на moeiv.

Тест 2.

Сроки сдачи статей для Научно-технического семинара кафедры МОЭВМ 2024 (бывшей конференции ППС) следующие:

1. Начало приема статей - 15 декабря.
2. Окончание приема статей - 16 января (включительно).
3. Предоставление рецензий (результатов рецензирования статей программным комитетом) - 24 января.
4. Дни конференции - 1-2 февраля.
5. Прием исправленных версий статей - до Московского времени 6 февраля.

Подробная информация о мероприятии и требованиям к материалам можно найти в тексте вашего сообщения и на официальном сайте конференции. Если у вас есть вопросы, связанные с участием в конференции, необходимо писать по почте ask@moevm.info.

Тест 3.

На занятиях будут изучаться следующие темы:

1. Искусственные нейронные сети
2. Рейтинговая система
3. Лекции

4. Практические занятия
5. Лабораторные работы
6. Индивидуальные задания
7. Оценка (рейтинг)

При этом в практических занятиях будет рассматриваться:

1. Основы Python
2. Создание простой нейронной сети с использованием библиотеки Keras
3. Представление данных и библиотека NumPy
4. Операции с тензорами в библиотеке Keras
5. Оптимизация нейронных сетей в библиотеке Keras
6. Процесс решения задач с применением нейронных сетей в библиотеке Keras
7. Мониторинг моделей глубокого обучения средствами библиотеки Keras
8. Ансамблирование моделей нейронных сетей с использованием библиотеки Keras

В лабораторных работах будут рассматриваться следующие задачи:

1. Многоклассовая классификация цветов
2. Бинарная классификация отраженных сигналов радара
3. Регрессионная модель изменения цен на дома в Бостоне
4. Распознавание рукописных символов
5. Распознавание объектов на фотографиях
6. Прогноз успеха фильмов по обзорам
7. Классификация обзоров фильмов
8. Генерация текста на основе "Алисы в стране чудес"

Для работы с этими задачами необходимы следующие библиотеки: NumPy, Pandas, Matplotlib, Scikit-Learn и Keras. Для установки рекомендуется использовать Python 3.7 или более позднюю версию. Библиотеки можно установить с помощью pip. Для подключения библиотеки Keras в коде необходимо импортировать keras (или tf.keras) и (или

tf.keras.layers). Кроме того, рекомендуется установить TensorFlow, так как он включает модуль Keras

3. gemma

Тест 1.

****Как зарегистрироваться на моеvm?****

Чтобы зарегистрироваться на моеvm, необходимо выполнить следующие действия:

1. ****Создать (или использовать) аккаунт на Discord:****
 - Завести аккаунт, если у вас его нет.
 - Указать ваше реальное имя и фамилию.
 - settlersкое фото в качестве аватара.
 - скачайте приложение Discord.
2. ****Создать (или использовать) аккаунт на GitHub:****
 - Завести аккаунт, если у вас его нет.
 - Указать ваше реальное имя и фамилию в поле Name.
 - settlersкое фото в качестве аватара.
3. ****Создать (или использовать) аккаунт на e.моеvm.info:****
 - Завести аккаунт, если у вас его нет.
 - Указать ваше реальное имя и фамилию в полях Имя и Фамилия.
 - settlersкое фото в качестве аватара.
4. ****Заполнить форму:****
 - Заполните форму с данными, которые были указаны при регистрации на Discord, GitHub и e.моеvm.info.

****Дополнительные ресурсы:****

- [Основная страница кафедры] (<http://моеvm.info>)

- [VK] (vk.com/moevm_leti)
- [Email для связи с администрацией] (ask(at)moevm.info)
- [Вики Вики start] (wiki.moevm.info/wiki/start)
- [Вики Ознакомиться с синтаксисом] (wiki.moevm.info/wiki/syntax)
- [Вики Попробовать свои силы без ущерба для проекта] (wiki.moevm.info/wiki/sandbox)
- [Общие сведения и рекомендации по работе с Wiki] (wiki.moevm.info/wiki/rules)

Тест 2.

****Научно-технически́ семинар кафедры МО ЭВМ 2024 (бывш. конференция ППС) ****

Ежегодно, в весеннем семестре на кафедре МО ЭВМ, проводит место научно-технически́ семинар, который является продолжением традиций прошлой конференции ППС.

****Основная информация:****

* ****Название:** Научно-технически́ семинар студентов и преподавателе́ каф. МО ЭВМ 2024**

* ****Дата:** 1 - 2 февраля 2024 года**

* ****Format:** ОНЛАИ́Н**

* ****Язык:** Русский**

* ****Количество участников:** Не ограничено**

* ****Сроки:****

* Начало приема статье́: 15 декабря 2023 года

* Окончание приема статье́: 16 января 2024 года

* Предоставление рецензии́: 24 января 2024 года

****Участие в семинаре дает студентам следующие возможности:****

* Получение дополнительных баллов для поступления в магистратуру

* Подготовиться к выступлению на конференции более высокого уровня

- * Опубликовать статью в сборнике
- * Обсудить результаты своей научной работы
- * Получить качественную обратную связь от коллег и рецензентов
- * Научиться, чем занимаются коллеги
- * Порепетировать предстоящую защиту выпускной работы

****Требования к материалам:****

- * Статьи (в формате PDF) не менее 3 и не более 12 страниц
- * Аннотация (от 150 слов)
- * Формат: PDF
- * Шрифт: не менее 18
- * Обязательные слайды:
 - * Титульный (тема, фамилия и имя докладчика, организация)
 - * Цель и задачи
 - * Постановка задачи (детальное описание, что именно вы делали и какие свойствами должно обладать решение)
 - * Заключение (что конкретно вы сделали, какие направления для дальнейших решений)

Тест 3.

К сожалению, текст не содержит информацию о темах, которые будут изучаться на занятиях, поэтому я не могу отвечать на этот вопрос.

4. vicuna

Тест 1.

Для регистрации на кафедре МОЭВМ необходимо выполнить следующие действия:

1. Завести аккаунт на платформе Discord, указав свои реальные имя и фамилию, а также скачать клиент.
2. Завести аккаунт на платформе GitHub, указав свои реальные имя и фамилию в поле Name, а также установить фото в качестве аватара.

3. Завести аккаунт на сайте `e.moevm.info`, указав свои реальные имя и фамилию в полях Имя и Фамилия, а также установить фото в качестве аватара.

4. Заполнить данными формой на Google Forms: <https://forms.gle/SDDN3ss4o5cnYsop7>

После выполнения всех этих действий необходимо связаться с администрацией кафедры по email `ask(at)moevm.info`, указав свои данные в поле Name.

Тест 2.

Некоторые журналы требуют, чтобы научные статьи были опубликованы в течение определенного периода времени после их подачи. Это может быть от нескольких месяцев до нескольких лет, в зависимости от журнала и тематики исследования. Важно следить за сроками, указанными издательством журнала, чтобы избежать просрочки публикации статьи.

Тест 3.

Темами, которые могут быть изучены на занятиях по курсу "Искусственные нейронные сети" в рамках магистерской программы "Прикладное машинное обучение", могут включать следующие:

1. Основы искусственных нейронных сетей и их применение в компьютерной визуализации;
2. Теория и практика глубокого обучения с использованием библиотеки Keras;
3. Операции с тензорами в библиотеке Keras;
4. Оптимизация нейронных сетей в библиотеке Keras;
5. Процесс решения задач с применением нейронных сетей в библиотеке Keras;
6. Ансамблирование моделей нейронных сетей с использованием библиотеки Keras;
7. Классификация обзоров фильмов и распознавание рукописных символов;
8. Генерация текста на основе "Алисы в стране чудес".

Кроме того, могут быть изучены различные методы машинного обучения, такие как регрессионные модели, кластер

5. openchat

Тест 1.

Для регистрации на кафедре МОЭВМ, выполните следующие шаги:

1. Заведите (если нет) аккаунт на <https://discord.com> и указать ваши реальные имя и фамилию, установить ваше фото в качестве аватара, скачать клиент.

2. Заведите (если нет) аккаунт на <https://github.com>, указать ваши реальные имя и фамилию в поле Name, установить ваше фото в качестве аватара.

3. Заведите (если нет) аккаунт на <https://e.moevm.info>, указать ваши реальные имя и фамилию в полях Имя и Фамилия, установить ваше фото в качестве аватара.

4. Заполните своими данными <https://forms.gle/SDDN3ss4o5cnYsop7>.

После выполнения этих шагов, вы сможете комфортно себя чувствовать в течение всего периода обучения и своевременно получать нужную информацию, а также выполнять и сдавать задания по некоторым дисциплинам.

Тест 2.

Научно-технический семинар кафедры МО ЭВМ 2024 (бывш. конференция ППС)

Ежегодно, в весеннем семестре на кафедре МО ЭВМ проводится обсуждение научных успехов студентов и преподавателей в формате небольшой конференции. В прошлом эта активность была привязана к ВУЗовской конференции ППС. В этом году мероприятие получило название "Научно-технический семинар студентов и преподавателей каф. МО ЭВМ 2024". Семинар пройдет в режиме ОНЛАЙН.

С 2017 года для подачи материалов конференции мы используем систему EasyChair, в которой необходимо сначала зарегистрироваться, а затем проследовать по ссылке указанной ниже для подачи статей. Регистрация и подача статей на НТС-2024

Лучшие статьи будут рекомендованы к расширению для публикации в журналах, а также будет оказана помощь в подготовке данной публикации. Требования к материалам

На конференцию принимаются статьи (в формате PDF), не менее 3 и не более 12 страниц, оформленные в шаблоне IEEE. Объем аннотации – от 150 слов.

Для получения положительного впечатления можно использовать:

- * Скриншоты / видео / интерактивное демо ваших разработок,
- * Ссылки на репозитории с исходным кодом,
- * Ссылки на опубликованное приложение.

Как выступить с докладом по научной работе? (видео-урок)

Тест 3.

На занятиях будут изучаться следующие темы:

1. Искусственные нейронные сети (ИНС) – основы и применения.
2. Рейтинговая система – описание и правила.
3. Лекции по различным аспектам глубокого обучения, включая архитектуры ИНС, алгоритмы оптимизации, методы обработки данных и тензоров.
4. Практические занятия – создание и тренировка простых нейронных сетей с использованием библиотеки Keras.
5. Лабораторные работы – решение различных задач машинного обучения с применением ИНС, включая классификацию, регрессию и обработку изображений.
6. Индивидуальные задания – самостоятельное решение задач машинного обучения с использованием ИНС.

7. Реи́тинг – оценка знаний студентов по пройденным темам и задачам.

8. Список источников – перечень литературы и ресурсов для самостоятельного изучения материалов.

Схема занятий:

1. Лекция 1 – Введение в Искусственные ней́ронные сети.
2. Лекция 2 – Архитектуры ИНС и их применения.
3. Лекция 3 – Оптимизация ИНС и алгоритмы обучения.
4. Лекция 4 – Представление данных и библиотека NumPy.
5. Лекция 5 – Операции с тензорами в библиотеке Keras.
6. Лекция 6 – Оптимизация ней́ронных сетей́ в библиотеке Keras.
7. Лекция 7 – Процесс решения задач с применением ней́ронных сетей́ в библиотеке Keras.
8. Лекция 8 – Мониторинг моделей́ глубокого обучения средствами библиотеки Keras.
9. Лекция 9 – Ансамблирование моделей́ ней́ронных сетей́ с использованием библиотеки Keras.
10. Практические занятия – создание и тренировка простых ней́ронных сетей.
11. Лабораторные работы – решение различных задач машинного обучения с применением ИНС.
12. Индивидуальные задания – самостоятельное решение задач машинного обучения с использованием ИНС.
13. Реи́тинг – оценка знаний студентов по пройденным темам и задачам.

6. Итог

Тест 1.

Для регистрации на МОЭВМ необходимо выполнить следующие шаги:

1. Создать аккаунт на Discord (если не существует) и указать реальные имя, фамилию, а также установить фото к аватар.

2. Создать аккаунт на GitHub (если не существует) и указать реальные имя, фамилию в поле Name, а также установить фото как аватар.

3. Создать аккаунт на e.moevm.info (если не существует) и указать реальные имя, фамилию в полях Имя и Фамилия, также установить фото как аватар.

4. Заполнить своими данными форму <https://forms.gle/SDDN3ss4o5cnYsop7>.

Также рекомендуется ознакомиться с ресурсами кафедры, включая основную страницу, VK, email

администрацией и Wiki МОЭВМ. Для начала работы на Wiki можно ознакомиться с справкой по синтаксису

попробовать свои силы в песочнице без ущерба для проекта. Общие сведения и рекомендации по работе с Wiki можно найти на странице Правила организации информации в Wiki.

Тест 2.

Сроки сдачи материалов для Научно-технического семинара кафедры МО ЭВМ 2024 следующие:

- * Начало приема статей – 15 декабря
- * Окончание приема статей – 16 января (включительно)
- * Предоставление рецензий – 24 января
- * Прием исправленных версий статей – до 6 февраля Московского времени

Тест 3.

Согласно представленному тексту, на занятиях будут изучаться следующие темы:

1. Искусственные нейронные сети (Artificial Neural Networks);
2. Создание простой нейронной сети с использованием библиотеки Keras;
3. Представление данных и библиотека NumPy;
4. Операции с тензорами в библиотеке Keras;

5. Оптимизация нейронных сетей в библиотеке Keras;
6. Процесс решения задач с применением нейронных сетей в библиотеке Keras;
7. Мониторинг моделей глубокого обучения средствами библиотеки Keras;
8. Ансамблирование моделей нейронных сетей с использованием библиотеки Keras.

Также будут изучаться практические задачи и лабораторные работы, связанные с глубоким обучением и машинным обучением с помощью Scikit-Learn и TensorFlow.

Выводы.

По субъективному мнению лучшими в каждом тесте оказались следующие модели:

- Тест 1 – openchat
- Тест 2 – mistral
- Тест 3 – mistral \ llama3

Хуже всего показала себя модель gemma, которая не смогла адекватно ответить ни на один из вопросов.

Mistral показал неплохие результаты по всем тестам, по сравнению с другими моделями, но при этом допустил некоторые грамматические ошибки при переводе текста на русский язык, например: “библиотеке”.

Llama-2 не смог проанализировать текст из третьего теста и дал ответ, основанный только на своих данных, при этом на английском. Первые же два ответа сформулированы хуже, чем у mistral.

Visuna – аналогичная проблема, как и у Llama-2, но уже со вторым тестом. При ответе на 3 тест модель смешала лекции и лабораторные работы и дала неполный ответ.

Openchat – самый лаконичный ответ на первый тест и провал по второму и третьему.

Llama3 – самая современная модель из представленных, не сильно опережает своего конкурента mistral в этих конкретных тестах. mistral в аналогичных тестах выдала результат похожий или более информативный.

Исходя из проведенных тестов, а также из данных о параметрах модели (см. Таблицу 1), наиболее предпочтительным для проекта выглядит модель Mistral, чье преимущество заключается также в меньшей загрузке видеокарты при тестировании по сравнению с Llama3.