

Полный Анализ Ответов Моделей на Вопросы по Изображениям

Автор: Андрей Устинов

1 ноября 2024 г.

Содержание

1	Введение	3
2	Обзор Платформ	3
2.1	NexaAI	3
2.2	Ollama	3
3	Тест скорости обычных LLM	3
3.1	Скорость Генерации Токенов	3
3.2	Интеграция с LoRA	3
4	Сравнение возможностей моделей	3
4.1	Изображение 1: 1.jpeg	4
4.1.1	Вопрос 1: What materials might these bottles be made of?	4
4.1.2	Вопрос 2: How could you differentiate between different types of products if labels were not visible?	5
4.1.3	Вопрос 3: What might be the environmental impact of these types of containers?	6
4.2	Изображение 2: 2.jpeg	8
4.2.1	Вопрос 1: What emotions do the people in the image seem to express?	8
4.2.2	Вопрос 2: What could be the relationship between the individuals?	9
4.3	Изображение 3: 3.jpeg	9
4.3.1	Вопрос 1: What might the puppy be feeling based on its expression?	9
4.3.2	Вопрос 2: How can you tell the puppy is comfortable in its environment?	10
4.3.3	Вопрос 3: What traits are typical for this breed of dog?	12
4.4	Изображение 4: 4.jpeg	13
4.4.1	Вопрос 1: What time of year could this image represent?	13
4.4.2	Вопрос 2: How might someone feel walking through this forest path?	14
4.4.3	Вопрос 3: What types of animals could live in this environment?	15
4.5	Изображение 5: 5.jpeg	16
4.5.1	Вопрос 1: What occasion might this table setting be prepared for?	16
4.5.2	Вопрос 2: What types of food are commonly served at gatherings like this?	18
4.5.3	Вопрос 3: How could the lighting in the image affect the atmosphere of the gathering?	19
5	Сводная Таблица Статистики	20

6	Анализ моделей TTS и STT	21
6.1	Краткое описание раздела	21
6.2	Методология	21
6.3	Оценка модели TTS <i>Bark</i>	21
6.3.1	Выводы по модели <i>Bark</i>	22
6.4	Оценка модели TTS <i>Bark-small</i>	22
6.4.1	Выводы	23
6.5	Дополнительная информация	23
6.6	Общий вывод по разделу	24
6.7	Общий вывод по разделу	24
7	Заключение	24
7.1	Платформы и их особенности	24
7.2	Сравнительный анализ	25
7.3	Перспективы	25

1 Введение

Исследование направлено на сравнительный анализ производительности и точности мультимодальных моделей искусственного интеллекта в задачах обработки визуально-текстовых запросов. В качестве объектов исследования выбраны платформы NexaAI и Ollama, в сравнении с базовой моделью ChatGPT-4V. Основные метрики оценки включают: латентность ответа, релевантность генерируемого контента и точность интерпретации визуального контекста. Методология тестирования основана на серии стандартизированных запросов с использованием контрольного набора изображений.

2 Обзор Платформ

2.1 NexaAI

NexaAI специализируется на компактных моделях с оптимальным использованием ресурсов

Простая установка, минимальная конфигурация, быстрый старт

Удобный Python API и документация для быстрой интеграции

Готовое UI-решение на Streamlit с гибкой настройкой

Быстрое развёртывание сервера

Модели для широкого спектра типов данных

Низкий порог входа, понятная архитектура, минимум зависимостей

2.2 Ollama

Плюс ollama в кроссплатформенности и простоте использования Высокая оптимизация

REST API для интеграции

Совместимость с библиотекой OpenAI

Контейнеризация моделей

Поддержка SOTA LLM и MLLM

Активное сообщество

3 Тест скорости обычных LLM

3.1 Скорость Генерации Токенов

- llama3.2-3b-instruct-q4 (NexaAI): 52.93 токен/сек
- Llama3.2-3B-Instruct:q4 (Ollama): 64.25 токен/сек

NexaAI демонстрирует высокую скорость генерации токенов, превосходя Ollama на 11.32 токен/сек.

3.2 Интеграция с LoRA

Обе платформы поддерживают интеграцию с LoRA (Low-Rank Adaptation)

4 Сравнение возможностей моделей

Для проведения детального анализа ответов моделей были выбраны пять различных изображений, каждое из которых сопровождалось набором вопросов.

4.1 Изображение 1: 1.jpeg



Рис. 1: Разнообразие бутылок для личного ухода на теплом фоне.

4.1.1 Вопрос 1: What materials might these bottles be made of?

Ответ ChatGPT-4o: These bottles are likely made of plastic, as this is a common material for personal care containers due to its durability and lightweight properties.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4:** The bottles in the image are made of glass.
- **nanoLLaVA:fp16:** As an AI language model, I don't have personal preferences or opinions. However, based on the image, it's possible that the bottles could be made of a variety of materials, including glass, plastic, metal, or ceramic. The specific material would depend on the manufacturer's choice.
- **llava-phi-3-mini:q4:** 1. Glass 2. Plastic
- **llava-llama-3-8b-v1.1:q4:** 1. Glass 2. Plastic 3. Metal
- **llava-v1.6-mistral-7b:q4:** The bottles might be made of glass or plastic.
- **minicpm-v:** The bottles in the image appear to be made of plastic, which is a common material for personal care products due to its lightweight nature and ease of molding into various shapes. The smooth finish on some bottles suggests they might have been designed with an emphasis on hygiene and aesthetic appeal, typical characteristics of cosmetic packaging materials like polyethylene or polipropylene.

Вывод 1: Для первого вопроса изображения 1 модели продемонстрировали разнообразные ответы относительно материалов, из которых могут быть изготовлены бутылки. **ChatGPT-4o** корректно указал пластик как вероятный материал, что соответствует общепринятым практикам в производстве средств личного ухода. Модели **llava-v1.6-vicuna-7b:q4**, **llava-llama-3-8b-v1.1:q4** и **llava-v1.6-mistral-7b:q4** также упомянули стекло и пластик, что делает их ответы релевантными, хотя некоторые из них добавили металлы, что может быть менее характерно для данного контекста. Модель **nanoLLaVA:fp16** предоставила наиболее широкий спектр материалов, включая металлы и керамику, что

может быть избыточным для данного типа продукции. Модель **llava-phi-3-mini:q4** сосредоточилась на стекле и пластике, что является точным, но менее детализированным ответом по сравнению с **minicpm-v**, которая предоставила подробное описание с упоминанием конкретных типов пластмасс, таких как полиэтилен и полипропилен. Таким образом, **minicpm-v** показала наиболее полное и точное понимание вопроса, хотя и не единственная модель, предоставившая корректные ответы.

4.1.2 Вопрос 2: How could you differentiate between different types of products if labels were not visible?

Ответ ChatGPT-4o: The shape, color, and type of dispenser can give clues. For example, spray bottles may contain perfumes, while pump bottles are often used for lotions or cleansers.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4:** By examining the image, one can differentiate between different types of products based on their distinct features. For example, a product with a round shape and a hole in the center is likely to be a donut, while a product with a rectangular shape and a hole in the center is likely to be a piece of paper. Additionally, the color and texture of the products can also provide clues to their identity.
- **nanoLLaVA:fp16:** If labels were not visible, it would be difficult to differentiate between different types of products. For example, a product with a label may be a specific type of appliance, while a product with a label may be a specific brand or type of food. Additionally, if the product is not labeled, it may not be clear what it contains or how it works.
- **llava-phi-3-mini:q4:**
 1. Shape and size: Different products often have unique shapes and sizes that can help identify them.
 2. Material: Feeling the texture or weight of the product can give clues about its material.
 3. Smell: Some products have distinctive smells that can help identify them.
 4. Sound: Some products make unique sounds when they are used or moved.
 5. Taste: Some products have a distinct taste that can help identify them.
 6. Color: Different products often have different colors that can help identify them.
 7. Packaging: Even if the product itself is not labeled, the packaging may have information about the product.
 8. Shape and size: Different products often have unique shapes and sizes that can help identify them.
 9. Material: Feeling the texture or weight of the product can give clues about its material.
 10. Smell: Some products have distinctive smells that can help identify them.
 11. Sound: Some products make unique sounds when they are used or moved.
 12. Taste: Some products have a distinct taste that can help identify them.
 13. Color: Different products often have different
- **llava-llama-3-8b-v1.1:q4:**
 1. Visual cues: Look for differences in shape, size, color, and packaging. For example, a bottle of soda might have a distinct shape and color compared to a bottle of water.
 2. Brand logos: Some products have recognizable logos that can help identify the brand.
 3. Packaging: Different types of products often have unique packaging designs, such as a box for a packaged snack or a bag for a bagged snack.
 4. Context: Consider the context in which the product is being displayed. If it's in a grocery store, it might be a packaged product. If it's in a restaurant, it could be a prepared dish.
 5. Smell: In some cases, you might be able to identify a product by its smell. For example, a bag of bread might have a distinct aroma compared to a bag of chips.

- **llava-v1.6-mistral-7b:q4**: By analyzing the shape, size, color, and texture of the products.
- **minicpm-v**: If labels were not visible, differentiating between types of products could be challenging. However, there are several alternative methods to identify them:
 1. ****Shape and Size****: Different product categories often have distinct shapes and sizes. For example, shampoo bottles tend to be tall and slender, while lotion containers might be shorter with a wider base.
 2. ****Cap Design****: The design of the cap can also provide clues about what type of product it is. Some caps are designed specifically for certain products like spray

Вывод 2: Второй вопрос изображения 1 касался способов различения продуктов без видимых этикеток. **ChatGPT-4o** предложил основывать различие на форме, цвете и типе диспенсера, что является практическим и применимым подходом. Модель **llava-v1.6-vicuna-7b:q4** предоставила менее релевантный ответ, приводя примеры, не связанные с бутылками для ухода. Модель **nanoLLaVA:fp16** сосредоточилась на лейблах, что отклоняется от сути вопроса. Модели **llava-phi-3-mini:q4**, **llava-llama-3-8b-v1.1:q4**, **llava-v1.6-mistral-7b:q4** и **minicpm-v** предложили разнообразные методы, такие как форма, размер, материал и дизайн крышек, что делает их ответы более полными и релевантными. **minicpm-v** выделяется своей структурированностью и детализацией, предложив конкретные методы различения продуктов, что делает её ответ наиболее информативным и полезным в данном контексте.

4.1.3 Вопрос 3: What might be the environmental impact of these types of containers?

Ответ ChatGPT-4o: Plastic containers can contribute to pollution if not properly recycled. They take hundreds of years to decompose and may harm wildlife.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4**: The environmental impact of these types of containers, such as Styrofoam or polystyrene, can be significant. Styrofoam is made from non-renewable resources and is not biodegradable, which means it can persist in the environment for a long time. It is also a major contributor to landfills and can take hundreds of years to decompose. Additionally, when it is discarded improperly, it can contribute to pollution and harm wildlife.
- **nanoLLaVA:fp16**: As an AI language model, I do not have personal opinions or beliefs. However, based on scientific research, the environmental impact of these types of containers can vary depending on the type, size, and material used. For example, lightweight and durable containers made from recycled materials can have a lower environmental impact than those made from non-renewable materials. Additionally, containers made from biodegradable or compostable materials can reduce the amount of waste sent to landfills.
- **llava-phi-3-mini:q4**:
 1. Plastic bottles: These containers can contribute to plastic pollution and harm wildlife if not properly disposed of or recycled.
 2. Glass bottles: These containers can also contribute to plastic pollution if not properly disposed of or recycled. However, glass is a more environmentally friendly option as it is recyclable and does not produce microplastics.
 3. Aluminum cans: These containers are recyclable and can be reused multiple times, making them a more environmentally friendly option. However, the production of aluminum cans requires a significant amount of energy and resources.
 4. Paper cups: These containers are biodegradable and can be composted, making them

a more environmentally friendly option. However, the production of paper cups requires the use of trees and other natural resources.

5. Plastic bags: These containers can contribute to plastic pollution and harm wildlife if not properly disposed of or recycled.

6. Glass jars: These containers are recyclable and do not produce microplastics, making them a more environmentally friendly option. However, the production

- **llava-llama-3-8b-v1.1:q4**:
 1. Single-use plastic bottles: These bottles are often used for water or other beverages, but they can take hundreds of years to decompose. They can also end up in landfills, oceans, and other natural environments, where they can harm wildlife and contribute to pollution.
 2. Single-use plastic cups: Similar to single-use plastic bottles, these cups are often used for hot or cold beverages, but they can also take hundreds of years to decompose. They can contribute to pollution and harm wildlife if not properly disposed of.
 3. Single-use plastic straws: These straws are often used for drinking, but they can also take hundreds of years to decompose. They can contribute to pollution and harm wildlife if not properly disposed of.
 4. Single-use plastic forks and knives: These utensils are often used for eating, but they can also take hundreds of years to decompose. They can contribute to pollution and harm wildlife if not properly disposed of.
 5. Single-use plastic bags: These bags are often used for carrying groceries or other items, but they can also take hundreds of years to decompose. They can contribute to pollution and harm wildlife if not properly disposed of.
 6. Single-use plastic lids: These lids are often used for
- **llava-v1.6-mistral-7b:q4**: The environmental impact of these types of containers can be significant, as they contribute to pollution and waste accumulation. The use of single-use plastics, such as those found in takeout containers, can lead to plastic pollution in the environment, which can harm wildlife and ecosystems. Additionally, the production and disposal of these containers can consume resources and contribute to greenhouse gas emissions. To minimize the environmental impact, it is important to reduce the use of single-use plastics and opt for reusable or biodegradable alternatives.

If the containers have any additives like fragrances or dyes that are harmful to the environment during their disposal, this can further exacerbate ecological issues such as water pollution.

- **minicpm-v**: The environmental impact of these types of containers can be significant. If they are made from plastic, which is a non-biodegradable material, it could lead to pollution in landfills and oceans if not properly recycled or disposed of. The production process for plastics also has an environmental cost due to the use of fossil fuels and energy consumption.

If the containers have any additives like fragrances or dyes that are harmful to the environment during their disposal, this can further exacerbate ecological issues such as water pollution.

Вывод 3: Третий вопрос изображения 1 относился к экологическому воздействию контейнеров. **ChatGPT-4o** кратко отметил проблему загрязнения пластиком и его долговечность в природе. Модель **llava-v1.6-vicuna-7b:q4** некорректно упомянула Styrofoam, что может не соответствовать типу рассматриваемых контейнеров. Модель **nanoLLaVA:fp16** предоставила общий обзор, затронув различные материалы и их влияние, что делает её ответ релевантным, но менее специфичным. Модели **llava-phi-3-mini:q4** и **llava-llama-3-8b-v1.1:q4** детализировали экологические проблемы различных материалов, включая

пластик, стекло и алюминий, что делает их ответы более информативными. Модель **llava-v1.6-mistral-7b:q4** также охватила значимые аспекты экологического воздействия. **minicpm-v** предоставила наиболее полный и детализированный ответ, охватывая разнообразные аспекты воздействия, включая добавки и их влияние на экологию. Таким образом, **minicpm-v** показала наивысшую степень полноты и точности в сравнении с другими моделями.

4.2 Изображение 2: 2.jpeg



Рис. 2: Пара наслаждается общением на свежем воздухе в солнечный день.

4.2.1 Вопрос 1: What emotions do the people in the image seem to express?

Ответ ChatGPT-4o: The individuals appear to be happy and relaxed, possibly enjoying each other's presence and the pleasant outdoor environment.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4:** The people in the image appear to be happy and joyful.
- **nanoLLaVA:fp16:** As an AI language model, I do not have access to the specific image you are referring to. However, based on general human emotions, it can be inferred that the people in the image are likely experiencing a range of emotions. Some people may be feeling happy, while others may be feeling sad or upset. The emotions can also vary depending on the context and the individual's personal experiences.
- **llava-phi-3-mini:q4:** 1. Happiness 2. Excitement 3. Nostalgia
- **llava-llama-3-8b-v1.1:q4:** The people in the image appear to be expressing a mix of emotions, including happiness, excitement, and anticipation. They are gathered together, smiling and posing for a photo, which suggests a positive and joyful atmosphere.
- **llava-v1.6-mistral-7b:q4:** The people in the image seem to express happiness and joy.
- **minicpm-v:** The people in the image appear to express emotions of joy, happiness, and affection. The man is smiling broadly with a playful expression as he carries the woman on his back, while she has her head tilted upwards, eyes closed, and mouth open in what seems like laughter or delight. Their body language suggests they are sharing a light-hearted moment together.

Вывод: Модели **llava-v1.6-vicuna-7b:q4**, **llava-llama-3-8b-v1.1:q4**, **llava-v1.6-mistral-7b:q4** и **minicpm-v** предоставили наиболее точные и релевантные ответы, описывая позитивные эмоции, такие как счастье и радость. Модель **nanoLLaVA:fp16** дала общий ответ, не конкретизируя эмоции, а модель **llava-phi-3-mini:q4** включает нерелевантные эмоции, такие как ностальгия. **minicpm-v** показала наиболее детализированный и точный ответ, включая конкретные примеры выражения эмоций через язык тела и мимику.

4.2.2 Вопрос 2: What could be the relationship between the individuals?

Ответ ChatGPT-4o: They could be a couple, friends, or family members, as their body language suggests closeness and affection.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4:** The individuals in the image are likely friends or acquaintances, as they are seen together in a relaxed setting, possibly enjoying a meal or a drink.
- **nanoLLaVA:fp16:** As an AI language model, I do not have access to specific information about the individuals in question. However, based on the given description, it is possible that the individuals could be friends, family members, or colleagues.
- **llava-phi-3-mini:q4:** 1. Family members, such as siblings or parents and children.
2. Friends or acquaintances.
3. Business associates or colleagues.
4. Romantic partners or spouses.
5. Teacher and student.
6. Healthcare provider and patient.
7. Legal representatives, such as lawyer and client.
8. Social worker and client.
9. Mentor and mentee.
10. Other professional relationships.
- **llava-llama-3-8b-v1.1:q4:** 2 people sitting on a couch, one of them is holding a remote. They could be friends, family members, or a couple. The exact relationship cannot be determined from the image alone.
- **llava-v1.6-mistral-7b:q4:** Friends
- **minicpm-v:** The individuals could be a couple, friends or family members.

Вывод: Модели **llava-v1.6-vicuna-7b:q4**, **nanoLLaVA:fp16**, **llava-llama-3-8b-v1.1:q4**, **llava-v1.6-mistral-7b:q4** и **minicpm-v** предоставили релевантные ответы, предлагая различные возможные отношения между людьми, такие как друзья, семья или пара. Модель **llava-phi-3-mini:q4** перечисляет множество вариантов, включая профессиональные отношения, что делает ответ менее сфокусированным, но все же релевантным. **minicpm-v** добавляет конкретные примеры, подтверждая разнообразие возможных отношений.

4.3 Изображение 3: 3.jpeg

4.3.1 Вопрос 1: What might the puppy be feeling based on its expression?

Ответ ChatGPT-4o: The puppy's expression suggests curiosity and innocence, possibly feeling secure and comfortable in its surroundings.



Рис. 3: Молодой, коричневый лабрадор-щенок лежит на мягкой поверхности и смотрит вверх.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4**: The puppy might be feeling curious or playful, as it is looking up at the camera with a smile on its face.
- **nanoLLaVA:fp16**: As an AI language model, I don't have personal experiences or emotions, but based on the given information, it's possible that the puppy is feeling scared or anxious. Its expression might be a mix of fear and excitement, as it's trying to navigate through its new environment.
- **llava-phi-3-mini:q4**: The puppy is looking up.
- **llava-llama-3-8b-v1.1:q4**: 1. The puppy might be feeling happy or content based on its expression.
- **llava-v1.6-mistral-7b:q4**: The puppy might be feeling happy or excited, as it is wagging its tail and looking up.
- **minicpm-v**: The puppy's expression, with its head resting on the surface and eyes looking directly at the camera, suggests a sense of curiosity or mild interest. It may be feeling relaxed but also attentive to something in front of it. The posture could indicate that the puppy is taking a break from play or activity, possibly waiting for someone or something.

Вывод: Модели **llava-v1.6-vicuna-7b:q4**, **llava-llama-3-8b-v1.1:q4**, **llava-v1.6-mistral-7b:q4** и **minicpm-v** предоставили релевантные ответы, описывая позитивные эмоции щенка, такие как любопытство, игривость, счастье и радость. Модель **nanoLLaVA:fp16** дала нерелевантный ответ, предполагая страх и тревогу, что не соответствует исходному описанию. Модель **llava-phi-3-mini:q4** дала минималистичный ответ, не раскрывая эмоционального состояния щенка. **minicpm-v** продемонстрировала наиболее точное понимание, описав одновременно расслабленное и внимательное состояние щенка.

4.3.2 Вопрос 2: How can you tell the puppy is comfortable in its environment?

Ответ ChatGPT-4o: The relaxed body posture and calm gaze indicate that the puppy feels safe and content.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4**: The puppy is sitting calmly on the floor, which indicates that it feels safe and at ease in its surroundings.
- **nanoLLaVA:fp16**:
 1. The puppy's body language: A puppy that is relaxed and comfortable in its environment is likely to be relaxed and relaxed. This can be seen in the puppy's posture, body language, and overall demeanor.
 2. The puppy's exploration: A puppy that is exploring its environment is likely to be comfortable in its environment. This can be seen in the puppy's exploration of different areas, such as the floor, the grass, or the trees.
 3. The puppy's interaction with its environment: A puppy that is interacting with its environment is likely to be comfortable in its environment. This can be seen in the puppy's exploration of different areas, such as the grass, the trees, or the house.
 4. The puppy's overall demeanor: A puppy that is calm and relaxed is likely to be comfortable in its environment. This can be seen in the puppy's overall demeanor, such as its relaxed posture, relaxed body language, and relaxed demeanor.
- **llava-phi-3-mini:q4**:
 1. The puppy is lying down on a soft surface, which is a comfortable position for a dog to rest.
 2. The puppy is in a well-lit room, which suggests that it is safe and secure.
 3. The puppy is in a clean and tidy environment, which indicates that it is well taken care of.
 4. The puppy is in a quiet and peaceful environment, which allows it to rest and relax.
 5. The puppy is in a familiar environment, such as a home or a dog kennel, which makes it feel safe and secure.
 6. The puppy is in a warm and cozy environment, which is essential for its comfort and well-being.
 7. The puppy is in a clean and hygienic environment, which is important for its health and well-being.
 8. The puppy is in a safe and secure environment, which is essential for its emotional well-being.
 9. The puppy is in a comfortable and spacious environment, which allows it to move around and play.
 10. The puppy is in a well-ventilated environment, which is important for its health
- **llava-llama-3-8b-v1.1:q4**:
 1. The puppy is laying down.
 2. The puppy is on a soft surface, such as a bed or a couch.
 3. The puppy is not showing any signs of distress or fear.

These factors suggest that the puppy is comfortable in its environment.
- **llava-v1.6-mistral-7b:q4**: The puppy is sitting down and appears relaxed.
- **minicpm-v**: The puppy appears comfortable in its environment because it is lying down with its head resting on a soft surface, which indicates relaxation. Its body language does not show signs of stress or anxiety; instead, the relaxed posture suggests that it feels safe and at ease in this setting.

Вывод: Все модели, кроме **llava-phi-3-mini:q4**, предоставили релевантные и точные ответы, описывая расслабленную позу и спокойное поведение щенка как признаки его комфорта. Модель **llava-phi-3-mini:q4** дала детализированный, но избыточный ответ, повторяя некоторые пункты. **minicpm-v** предоставила четкий и точный ответ, подчеркивая расслабленное и безопасное состояние щенка.

4.3.3 Вопрос 3: What traits are typical for this breed of dog?

Ответ ChatGPT-4o: Labradors are known for their friendliness, intelligence, and loyalty, making them excellent family pets.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4:** The breed of dog in the image is a Border Collie. Border Collies are known for their intelligence, agility, and high energy levels. They are also known for their distinctive **eye** or **stare**, which is a focused and intense gaze that they use to control and direct livestock. Border Collies are typically black and white in color, and they have a medium to large size. They are a herding breed and are often used on farms and ranches to help manage livestock.
- **nanoLLaVA:fp16:** As an AI language model, I do not have personal preferences or opinions. However, based on the information available, the traits that are typically found in this breed of dog are:
 1. Strong physical strength and endurance
 2. Intelligence and problem-solving abilities
 3. Loyalty and companionship
 4. Adaptability and ability to work with different environments
 5. Good communication skills and ability to understand and respond to commands
 6. Good health and temperament
 7. Ability to work in a team and as part of a larger organization
 8. Good physical fitness and stamina
 9. Good health and temperament
 10. Good social skills and ability to work with people of all ages and abilities.
- **llava-phi-3-mini:q4:** 1. High energy and intelligence. 2. Good with children. 3. High prey drive. 4. Highly trainable. 5. Highly active and requires regular exercise. 6. Can be stubborn. 7. Can be prone to barking. 8. Can be sensitive to changes in routine. 9. Can be strong-willed. 10. Can be sensitive to stress.
- **llava-llama-3-8b-v1.1:q4:** 1. Loyal 2. Protective 3. Independent 4. Energetic 5. Intelligent 6. Trainable 7. Active 8. Social 9. Alert 10. Good with children.
- **llava-v1.6-mistral-7b:q4:** The breed of dog in the image is a golden retriever. They are known for their friendly, intelligent, and loyal nature. They are also known for their beautiful golden coat and their love for water.
- **minicpm-v:** The dog in the picture is a Chocolate Labrador Retriever. This breed of dog typically has several notable traits:
 1. **Friendly and Affectionate:** Labradors are known for their loving nature, making them great family pets.
 2. **High Energy Levels:** They require regular exercise to stay healthy and happy.
 3. **Intelligent:** Labs are often used in various roles such as guide dogs or search-and-rescue due to their intelligence.
 4. **Good Temperament:** Generally

Вывод: Ответы моделей варьируются в зависимости от идентификации породы. Модель **llava-v1.6-vicuna-7b:q4** ошибочно идентифицирует породу как бордер-колли, предоставляя нерелевантную информацию. Модель **llava-v1.6-mistral-7b:q4** неверно определяет породу как золотистый ретривер. Модели **nanoLLaVA:fp16**, **llava-phi-3-mini:q4** и **llava-llama-3-8b-v1.1:q4** перечисляют общие характеристики пород, но без точного указания на лабрадора. Модель **minicpm-v** также ошибочно идентифицировала породу как шоколадного лабрадора, но начала перечислять конкретные черты, хотя ответ не был завершён. Таким образом, все модели не смогли точно идентифицировать породу щенка, но описали общие положительные и активные черты, характерные для лабрадоров.

4.4 Изображение 4: 4.jpeg



Рис. 4: Снежный лесной путь, окруженный высокими деревьями, покрытыми снегом.

4.4.1 Вопрос 1: What time of year could this image represent?

Ответ ChatGPT-4o: This image likely represents winter, as evidenced by the snow-covered trees and pathway.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4:** Winter
- **nanoLLaVA:fp16:** It's hard to definitively say without more context, but the presence of snow on the ground could suggest that it's either winter or late fall.
- **llava-phi-3-mini:q4:** 12:00
- **llava-llama-3-8b-v1.1:q4:** This image could represent the fall season, as the leaves on the tree are changing colors and falling to the ground.
- **llava-v1.6-mistral-7b:q4:** Summer
- **minicpm-v:** This image likely represents winter. The presence of snow on the ground, trees, and branches indicates cold temperatures typical of this season.

Вывод: Модели **llava-v1.6-vicuna-7b:q4** и **minicpm-v** правильно определили сезон как зиму. Модели **llava-phi-3-mini:q4**, **llava-llama-3-8b-v1.1:q4** и **llava-v1.6-mistral-7b:q4** дали неверные или нерелевантные ответы, такие как «12:00», осень и лето соответственно. Таким образом, **minicpm-v** является единственной моделью, которая правильно идентифицировала сезон, соответствующий изображению.

4.4.2 Вопрос 2: How might someone feel walking through this forest path?

Ответ ChatGPT-4o: Someone might feel calm and introspective, or perhaps enchanted by the quiet beauty of the snowy landscape.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4:** Someone walking through this forest path might feel a sense of tranquility and wonder. The snow-covered trees create a picturesque winter scene, which can be both calming and awe-inspiring as one takes in the beauty around them. There may also be feelings of solitude or introspection due to the quietness typically found in such natural settings. Additionally, if someone is an avid nature lover, they might feel a deep appreciation for the pristine environment and its untouched state during winter.
- **nanoLLaVA:fp16:** As an AI language model, I don't have personal experiences or emotions. However, based on the information provided, someone walking through the forest path might feel a sense of tranquility and wonder. The snow-covered trees create a picturesque winter scene, which can be both calming and awe-inspiring as one takes in the beauty around them. There may also be feelings of solitude or introspection due to the quietness typically found in such natural settings. Additionally, if someone is an avid nature lover, they might feel a deep appreciation for the pristine environment and its untouched state during winter.
- **llava-phi-3-mini:q4:**
 1. Peaceful
 2. Calm
 3. Relaxed
 4. Serene
 5. Tranquil
 6. Reflective
 7. Inspired
 8. Curious
 9. Adventurous
 10. Content
- **llava-llama-3-8b-v1.1:q4:** The person walking through the forest path might feel a sense of tranquility and connection with nature. The path is surrounded by trees, providing a canopy of greenery and shade, which can create a peaceful and serene atmosphere. The person might also feel a sense of adventure and exploration as they navigate the winding path through the forest. The natural setting can be a refreshing break from the hustle and bustle of daily life, allowing the person to unwind and enjoy the beauty of the outdoors.
- **llava-v1.6-mistral-7b:q4:** Someone walking through this forest path might feel a sense of tranquility and connection with nature. The path appears to be peaceful and serene, surrounded by trees and greenery. It could be a great place for a quiet walk, meditation, or simply enjoying the beauty of the forest.
- **minicpm-v:** Walking through this forest path might evoke a sense of tranquility and wonder. The snow-covered trees create a picturesque winter scene, which can be both

calming and awe-inspiring as one takes in the beauty around them. There may also be feelings of solitude or introspection due to the quietness typically found in such natural settings. Additionally, if someone is an avid nature lover, they might feel a deep appreciation for the pristine environment and its untouched state during winter.

Вывод: Все модели, кроме **llava-phi-3-mini:q4**, предоставили релевантные и точные ответы, описывая чувства спокойствия, умиротворения и связи с природой. Модель **llava-phi-3-mini:q4** дала краткий и списочный ответ, не раскрывая глубину эмоционального восприятия. **minicpm-v** предоставила детализированный и точный ответ, полностью соответствующий заданному вопросу.

4.4.3 Вопрос 3: What types of animals could live in this environment?

Ответ ChatGPT-4o: This environment could be home to animals like deer, wolves, and various bird species adapted to cold weather.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4:** In this environment, which appears to be a desert or a dry savanna, there are several types of animals that could potentially live. These include:
 1. Desert-adapted mammals such as the antelope, gazelle, and ibex, which are well-adapted to the harsh conditions of the desert.
 2. Reptiles like the desert tortoise, which can survive in the extreme heat and arid conditions.
 3. Birds that have adapted to the desert environment, such as the roadrunner, which is well-known for its ability to run at high speeds in the hot desert sun.
 4. Insects like the locust, which can thrive in the dry and arid conditions of the desert.
 5. Small mammals like the kangaroo rat, which can survive on very little water and food.

It's important to note that the specific types of animals that can be found in this environment will depend on the specific location and the availability of resources.

- **nanoLLaVA:fp16:** Animals that could live in this environment include deer, foxes, squirrels, rabbits, and various bird species.
- **llava-phi-3-mini:q4:**
 1. Fish
 2. Frogs
 3. Turtles
 4. Snakes
 5. Birds
 6. Lizards
 7. Insects
 8. Mammals
 9. Reptiles
 10. Amphibians
 11. Moss
 12. Algae
 13. Plants
 14. Bacteria
 15. Fungi
 16. Protozoa
 17. Viruses
 18. Bacteria

19. Algae
20. Plants
21. Moss
22. Insects
23. Birds
24. Lizards
25. Snakes
26. Turtles
27. Frogs
28. Mammals
29. Reptiles
30. Amphibians
31. Bacteria
32. Fungi
33. Protozoa
34. Viruses
35. Algae
36. Plants
37. Moss
38. Insects
39. Birds
40. Liz

- **llava-llama-3-8b-v1.1:q4**: This environment is a grassy field with a few trees, which suggests that it could be a suitable habitat for various types of animals. Some examples of animals that could live in this environment include herbivores such as deer, antelope, or gazelles, as well as predators like lions, leopards, or hyenas. Additionally, the presence of trees could provide shelter for animals like birds, squirrels, or small mammals. The specific types of animals that could live in this environment would depend on factors such as the region, climate, and availability of resources.
- **llava-v1.6-mistral-7b:q4**: In this environment, which appears to be a desert or arid region, animals that could live there include camels, kangaroos, and various species of reptiles and birds. These animals are well-adapted to survive in harsh, dry conditions.
- **minicpm-v**: Animals that could live in this environment include deer, foxes, squirrels, rabbits, and various bird species.

Вывод: Модель **llava-v1.6-vicuna-7b:q4** ошибочно определяет среду как пустыню, предоставляя нерелевантные примеры животных. Модель **llava-phi-3-mini:q4** дала обширный и повторяющийся список, но без фокусировки на зимнем лесу. Модели **nanoLLaVA:fp16** и **llava-llama-3-8b-v1.1:q4** частично релевантны, упоминая лесных животных, но не полностью соответствуют описанной снежной среде. Модель **llava-v1.6-mistral-7b:q4** также ошибочно определяет среду как пустыню. Модель **minicpm-v** предоставила точные примеры животных, характерных для лесной среды, такие как олени, лисы, белки и кролики, что делает её ответ наиболее релевантным среди всех моделей.

4.5 Изображение 5: 5.jpeg

4.5.1 Вопрос 1: What occasion might this table setting be prepared for?

Ответ ChatGPT-4o: This setting could be for a holiday celebration, such as Christmas or Thanksgiving, given the decorative elements and elaborate food presentation.



Рис. 5: Празднично накрытый рождественский стол с едой, украшениями и огнями.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4**: This table setting appears to be prepared for a formal dinner party or a special event, as it includes a variety of dishes, multiple courses, and a well-arranged table setting. The presence of a wine glass, a water glass, and a bread plate suggests that the guests will be served wine, water, and bread during the meal. The attention to detail in the table setting indicates that the host has put effort into creating an elegant and sophisticated dining experience for their guests.
- **nanoLLaVA:fp16**: It's hard to say for certain without more information, but a table setting could be prepared for a variety of occasions such as a casual meeting, a small gathering, a party, or a formal event.
- **llava-phi-3-mini:q4**:
 1. A formal dinner party
 2. A special occasion like a birthday or anniversary
 3. A romantic dinner for two
 4. A holiday celebration
 5. A dinner party with friends
 6. A family gathering
 7. A dinner party with a theme
 8. A dinner party with a sit-down meal
 9. A dinner party with a buffet-style meal
- **llava-llama-3-8b-v1.1:q4**: 4 people are seated around a table with a white tablecloth, plates, silverware, and wine glasses. This setting might be prepared for a formal dinner party or a special celebration.
- **llava-v1.6-mistral-7b:q4**: Wedding
- **minicpm-v**: This table setting appears to be prepared for a festive occasion, likely Christmas dinner. The presence of multiple dishes including what looks like turkey or roast meat, pies, and various side dishes suggests it's meant for sharing among family members or friends during the holiday season.

Вывод: Модель **llava-v1.6-vicuna-7b:q4** предоставила наиболее детализированный и релевантный ответ, указывая на формальный ужин или особое мероприятие. Модель

llava-llama-3-8b-v1.1:q4 также дала релевантный ответ, связывая сервировку с формальными ужинами или празднованиями. Модель **nanoLLaVA:fp16** дала общий ответ, не конкретизируя. Модель **llava-phi-3-mini:q4** предоставила широкий список возможных событий, но без точности. Модель **llava-v1.6-mistral-7b:q4** ошибочно указала на свадьбу, что не соответствует рождественской тематике. **minicpm-v** корректно определила событие как рождественский ужин, подчеркнув праздничные элементы и предназначение сервировки для семейного или дружеского собрания.

4.5.2 Вопрос 2: What types of food are commonly served at gatherings like this?

Ответ ChatGPT-4o: Traditional holiday meals may include dishes like roast meats, pies, casseroles, and various side dishes, depending on the culture.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4:** At gatherings like this, a variety of finger foods and appetizers are commonly served. Examples include cheese and crackers, vegetable platters, fruit skewers, bruschetta, and chicken wings.
- **nanoLLaVA:fp16:** At gatherings like this, people often enjoy a variety of foods such as salads, sandwiches, pasta, pasta salad, pizza, sandwiches, pasta, pasta salad, and desserts.
- **llava-phi-3-mini:q4:**
 1. Appetizers such as cheese platters, charcuterie boards, and finger foods.
 2. Main courses like grilled meats, pasta dishes, and casseroles.
 3. Side dishes such as salads, roasted vegetables, and bread.
 4. Desserts like cakes, pies, and cookies.
 5. Beverages like wine, beer, and soft drinks.
 6. Snacks like chips, crackers, and nuts.
 7. Condiments like dips, sauces, and spreads.
 8. Beverages like water, juice, and soda.
 9. Fruits like apples, oranges, and grapes.
 10. Vegetables like carrots, celery, and broccoli.
 11. Meats like chicken, beef, and pork.
 12. Seafood like shrimp, lobster, and fish.
 13. Dips like hummus, guacamole, and salsa.
 14. Sandwiches and wraps.
 15. Salads like
- **llava-llama-3-8b-v1.1:q4:**
 1. Finger foods: chips, dips, fruit, cheese, crackers, breadsticks, etc.
 2. Main dishes: sandwiches, salads, pasta, grilled meats, etc.
 3. Side dishes: rice, potatoes, vegetables, etc.
 4. Desserts: cakes, pies, cookies, etc.
 5. Beverages: water, soda, juice, etc.
- **llava-v1.6-mistral-7b:q4:** Appetizers, finger foods, and hors d'oeuvres are commonly served at gatherings like this.
- **minicpm-v:** At gatherings like this, it is common to serve a variety of foods that are typically associated with festive occasions. Some examples include:
 1. Roast meats such as turkey or ham
 2. Vegetables including roasted carrots, Brussels sprouts, and green beans

3. Sides dishes such as stuffing, mashed potatoes, cranberry sauce, and gravy
4. Desserts like pumpkin pie, apple pie, or cake with whipped cream

These foods are often served in abundance to accommodate a large

Вывод: Модели **llava-v1.6-vicuna-7b:q4**, **llava-llama-3-8b-v1.1:q4** и **llava-phi-3-mini:q4** предоставили релевантные ответы, перечисляя различные типы блюд и закусок, характерных для праздничных мероприятий. Модель **nanoLLaVA:fp16** дала повторяющийся и менее структурированный ответ, а модель **llava-v1.6-mistral-7b:q4** дала краткий и неполный ответ. Модель **minicpm-v** корректно перечислила основные блюда, характерные для рождественского ужина, такие как жареное мясо, овощи и десерты, что делает её ответ наиболее релевантным и структурированным среди всех моделей.

4.5.3 Вопрос 3: How could the lighting in the image affect the atmosphere of the gathering?

Ответ ChatGPT-4o: The warm lighting adds a cozy and inviting ambiance, making the gathering feel more intimate and festive.

Ответы моделей:

- **llava-v1.6-vicuna-7b:q4:** The lighting in the image, which appears to be dimly lit, can create a more intimate and cozy atmosphere for the gathering. The warm, soft lighting can encourage conversation and interaction among the guests, while also highlighting the ambiance of the venue. However, it is important to note that the lighting may also make it more difficult for guests to see each other clearly, which could potentially hinder the overall experience.
- **nanoLLaVA:fp16:** The lighting in the image can greatly affect the atmosphere of the gathering. For example, bright and bright lighting can create a lively and energetic atmosphere, while dim lighting can create a more intimate and relaxed atmosphere. Additionally, the lighting can also affect the mood of the attendees, with bright lighting often associated with excitement and enthusiasm, while dim lighting can create a more contemplative or introspective atmosphere.
- **llava-phi-3-mini:q4:**
 1. The lighting in the image creates a warm and inviting atmosphere, which can make the gathering feel more intimate and cozy. The soft glow from the candles and the dim lighting can create a sense of relaxation and comfort, encouraging guests to unwind and enjoy each other's company.
 2. The lighting in the image can create a more intimate and romantic atmosphere, which can make the gathering feel more special and memorable. The candles and dim lighting can create a sense of warmth and intimacy, making the gathering feel more personal and romantic.
 3. The lighting in the image can create a more relaxed and casual atmosphere, which can make the gathering feel more informal and laid-back. The candles and dim lighting can create a sense of warmth and relaxation, encouraging guests to let their guard down and enjoy the moment.
- **llava-llama-3-8b-v1.1:q4:** The lighting in the image, which is dimly lit with a candle and a lamp, creates a warm and intimate atmosphere for the gathering. The soft glow of the candlelight and the gentle illumination from the lamp can make the space feel cozy and inviting, encouraging conversation and connection among the people present. This type of lighting is particularly suitable for small gatherings or intimate settings where a more relaxed and personal ambiance is desired.

- **llava-v1.6-mistral-7b:q4**: The lighting in the image can create a warm and inviting atmosphere for the gathering. The soft, warm light can make the people in the image feel more comfortable and relaxed, encouraging them to engage in conversation and enjoy each other’s company. The lighting can also highlight the details of the room and the people in it, making the image more visually appealing and memorable.
- **minicpm-v**: The lighting in the image, which consists of several lit candles and a warm-toned overhead light, creates an intimate and cozy atmosphere. The soft glow from the candlelight adds warmth to the scene, making it feel inviting and comfortable for guests. This type of lighting is often associated with relaxation and celebration, suggesting that this gathering could be a special occasion such as Christmas dinner or another festive event where people come together to enjoy good food and company in a warm setting. The combination of natural candlelight and

Вывод: Анализ показывает высокую релевантность ответов всех моделей при описании влияния освещения на атмосферу собраний. Лучший результат продемонстрировала llava-phi-3-mini:q4, предоставив наиболее структурированный и детальный анализ. Уникальным оказался подход llava-v1.6-vicuna-7b, отметившей недостаток тусклого освещения, а nanoLLaVA:fp16 выделилась сравнением яркого и приглушенного света. llava-v1.6-mistral-7b сфокусировалась на влиянии освещения на комфорт и общение людей, а minicpm-v детально описала конкретные источники света и их роль в создании праздничной атмосферы. Все модели подчеркнули связь между освещением и социальным взаимодействием, но каждая внесла свой уникальный вклад в анализ.

5 Сводная Таблица Статистики

Модель	Запросов	Релевантных	% Релевантных
llava-v1.6-mistral-7b:q4	15	11	73.3
llava-v1.6-vicuna-7b:q4	15	8	53.3
llava-llama-3-8b-v1.1:q4	15	8	53.3
nanoLLaVA:fp16	15	6	40.0
llava-phi-3-mini:q4	15	5	33.3
minicpm-v	15	12	80.0

Таблица 1: Сравнительный анализ релевантности ответов моделей

Модель	Токенов	Время (с)	Токен/с
llava-v1.6-vicuna-7b:q4	819	22.25	36.83
llava-llama-3-8b-v1.1:q4	905	30.66	29.51
llava-v1.6-mistral-7b:q4	436	12.74	34.19
nanoLLaVA:fp16	986	12.16	81.03
llava-phi-3-mini:q4	1361	33.99	40.08
minicpm-v	615	34.05	18.07

Таблица 2: Анализ производительности моделей

6 Анализ моделей TTS и STT

6.1 Краткое описание раздела

В разделе рассматриваются модели синтеза речи (TTS) и распознавания речи (STT). Базовые эксперименты проводились на платформе Huggingface, так как на NехаAI возникли проблемы с циклическим импортом из-за ошибок в библиотеке.

Исследуются модели:

- *Bark*
- *Bark Mini*
- *Whisper-large-v3-Turbo*

6.2 Методология

Модель *Bark* тестировалась в двух сценариях:

- **Без пресета:** Генерация речи осуществляется без дополнительных настроек.
- **С пресетом `en_speaker_6`:** Использование пресета для улучшения качества генерации.

Для оценки качества транскрипций использовался показатель *Word Error Rate* (WER). Также измерялось использование памяти моделью во время работы и время генерации речи.

Whisper-large-v3-Turbo - SOTA модель, способная корректно оценить качество озвучки моделями *Bark* и *Bark-mini*. Идея была в том, чтобы озвучить заготовленные предложения и расшифровать их через *Whisper-large-v3-Turbo*, а потом сравнить по метрике WER близость получившихся транскрипций к исходным предложениям.

Таблица 3: Подвыборка оригинальных предложений

ID	Предложение
0	I just saw the funniest cat video ever!
1	Another Monday morning...
2	We won the championship! Can you believe it?
3	Excuse me, could I have everyone's attention p...
4	I really need to get more sleep these days...
5	Uh... I have something important to tell you.

6.3 Оценка модели TTS *Bark*

Таблица 4: Сравнение WER и использования памяти для модели *Bark*

Настройка	WER(%)	Использование памяти (ГБ)
Без пресета	317.8	5.84
С пресетом <code>en_speaker_6</code>	34.7	5.84

Таблица 5: Скорость генерации и транскрипции для модели *Bark*

Задача	Время выполнения
Генерация озвучки (без ограничения длины)	9.5 сек
Транскрипция одного предложения	1.8 сек
Суммарное время озвучки (20 предложений)	190 сек
Суммарное время транскрипции (20 предложений)	36 сек

Таблица 6: Транскрипции аудио от *Bark* с *en_speaker_6*

ID	Транскрипция
0	so I just saw the funniest cat video ever.
1	and then another...
2	So, you know, I'm really all ones over top in...
3	Excuse me, could I have everyone's attention,...
4	I really need to get more sleep these days.
5	I have something important to tell you.

Таблица 7: Транскрипции аудио от *Bark* без *en_speaker_6*

ID	Транскрипция
0	so I just saw the funniest cat video ever.
1	in this politics, and my end is a go on sssss...
2	and we won the championship. Can you believe it?
3	Can this, excuse me, kind of everyone's attention...
4	and so I really need to get more sleep these days...
5	I have something important to tell you.

6.3.1 Выводы по модели *Bark*

1. Использование пресета *en_speaker_6* значительно улучшает качество транскрипции - снижение WER с 317.8% до 34.7% при температуре 0.1.
2. Модель требует существенных вычислительных ресурсов (5.84 ГБ памяти) и относительно медленная: 9.5 секунд на генерацию одного предложения.
3. При низкой температуре (0.1) модель без пресета демонстрирует сильные искажения ("in this politics, and my end is a go on sssss..."), с пресетом ошибки в основном затрагивают начало предложений (добавление слов "so "and").
4. При температуре 0.5-0.7 качество генерации улучшается, но результаты становятся нестабильными между запусками.

6.4 Оценка модели TTS *Bark-small*

Таблица 8: Сравнение WER и использования памяти для модели *Bark Mini*

Настройка	WER(%)	Использование памяти (ГБ)
Без пресета	251.9	2.61
С пресетом <i>en_6</i>	47.3	2.61

Таблица 9: Скорость генерации и транскрипции для модели *Bark-small*

Задача	Время выполнения
Генерация озвучки (без ограничения длины)	3.5 сек
Транскрипция одного предложения	1.8 сек
Суммарное время озвучки (20 предложений)	70 сек
Суммарное время транскрипции (20 предложений)	36 сек

Таблица 10: Транскрипции аудио от *Bark-small* с `en_speaker_6`

ID	Транскрипция
0	I just saw the funniest cat video ever.
1	Another Monday morning?
2	We won the championship. Can you believe it?
3	Excuse me. Could I have everyone's attention?
4	I really need to get more sleep these days.
5	Uh, I have something important to tell you.

Таблица 11: Транскрипции аудио от *Bark-small* без `en_speaker_6`

ID	Транскрипция
0	I just saw the funniest cat video ever!
1	Another Monday morning, maybe 2016 from mid-1...
2	We won the championship! Can you believe it?
3	Then they go to appropriate and see...
4	Aaaaaaaaaaaaaaaaaaaaaaaaaa...
5	I have something important to tell you.

6.4.1 Выводы

1. При температуре 0.1 *Bark-small* с пресетом `en_speaker_6` показывает WER 47.3%, без пресета - 251.9%.
2. Модель существенно эффективнее по ресурсам (2.61 ГБ памяти) и быстрее в 2.7 раза (3.5 сек против 9.5 сек на генерацию).
3. При высоких температурах (0.5-0.7) качество сравнимо с полной версией *Bark*, но также наблюдается нестабильность результатов между запусками.
4. Без пресета наблюдаются критические искажения даже при низкой температуре (пример с "Аaaaaaaaaaaaaaaaaaaaaaaaaa...").

6.5 Дополнительная информация

- **batch size:** 4
- **VRAM for *Whisper-large-v3-Turbo*:** 6.40 ГБ
- **GPU:** Nvidia Tesla T4.

6.6 Общий вывод по разделу

1. При температуре 0.1 Bark-small с пресетом `en_speaker_6` показывает WER 47.3%, без пресета - 251.9%.
2. Модель существенно эффективнее по ресурсам (2.61 ГБ памяти) и быстрее в 2.7 раза (3.5 сек против 9.5 сек на генерацию).
3. При высоких температурах (0.5-0.7) качество сравнимо с полной версией Bark, но также наблюдается нестабильность результатов между запусками.
4. Без пресета наблюдаются критические искажения даже при низкой температуре (пример с "Аааааааааааааааааааа...").

6.7 Общий вывод по разделу

1. Результаты исследования показывают противоречивую картину. При температуре 0.1 результаты стабильны, но качество генерации существенно ниже потенциально возможного. При повышении температуры до 0.5-0.7 качество значительно улучшается, но возникает проблема воспроизводимости результатов - между запусками показатели моделей могут существенно различаться, что затрудняет однозначное определение лучшей модели.
2. Оптимальным решением представляется:
 - Использование Bark-small с пресетом `en_speaker_6`
 - Установка температуры на уровне 0.55
 - Принятие некоторой вариативности результатов как компромисса
3. Данная конфигурация обеспечивает:
 - Приемлемое качество генерации
 - Экономии вычислительных ресурсов
 - Высокую скорость работы
 - Разумный баланс между стабильностью и качеством
4. Использование моделей без пресетов нецелесообразно при любых настройках температуры из-за:
 - Критически высокого WER
 - Серьезных искажений смысла
 - Появления артефактов в генерации

7 Заключение

7.1 Платформы и их особенности

NexaAI – фреймворк для оптимизированного edge-computing с широким охватом возможностей (текст, изображения, аудио, SD/Flux), сделавший сознательный выбор в пользу легковесных моделей. Однако во многих категориях нет моделей, а одномоментная загрузка моделей с разными датами релиза указывает на ограниченное количество разработчиков. Отдельно стоит отметить проблемы в работе с TTS и STT моделями: из-за

ошибок в библиотеках и циклического импорта не удалось провести тестирование на платформе, а представленные две модели TTS показывают низкое качество, в то время, как на Huggingface есть более современные решения, например `amphion/MaskGCT`.

Ollama – open-source платформа с активным комьюнити, специализирующаяся на текстовых моделях, мультимодальном анализе и эмбедингах, демонстрирующая высокое качество обработки в рамках специализации.

7.2 Сравнительный анализ

В базовых LLM NехаAI показывает более высокую скорость (52.93 против 64.25 токен/сек), что соответствует их фокусу на легковесных моделях. NехаAI осознанно предпочли развиваться в сторону легковесных моделей, поэтому там не найти многие экземпляры выше 10B параметров, которые уже давно есть на ollama (примером служит лидер в этом сравнении - модель `minicpm-v`). Обе платформы предоставляют API, но имеют разный фокус: NехаAI стремится к широкому охвату типов данных при сохранении легковесности, тогда как Ollama концентрируется на LLM, мультимодальных LLM и моделях эмбедингов. При этом попытка охватить все направления в NехаAI иногда приводит к проблемам: например, раздел TTS/STT моделей существенно уступает специализированным платформам как по количеству доступных моделей, так и по их качеству и стабильности работы.

7.3 Перспективы

Небольшое комьюнити NехаAI может затруднить следование трендам ИИ, в то время как активное сообщество Ollama обеспечивает стабильное развитие платформы в выбранном направлении. Выбор платформы зависит от приоритетов проекта: NехаAI подходит для edge-computing и разнообразных медиа-задач с умеренными требованиями к качеству (за исключением TTS/STT направления, где рекомендуется использовать специализированные решения), Ollama – для задач, где качество обработки важнее дешевизны решения.