

بخش اول تولید داده :

داده های مربوط به گروه ویس :

ضبط ویس برای 96 نسخه پزشکی برای گروه صوت که داروها و یا دوزهای هر نسخه متفاوت بود و متن محاوره ای بود و بعضی از ویس ها هم با نویز بالا گرفته شدند.

داده های مربوط به گروه متن :

- ۱- ساخت ۲۰ نسخه پزشکی برای تست مدل ها انجام شد که این ۲۰ نسخه ها هر کدام ویژگی خاص خودشان را دارند و ساختارشان یکسان نیست. (raw_prescriptions.txt)
- ۲- ساخت 1600 نسخه پزشکی محاوره ای با ایرادات و نواقص مختلف با استفاده از gemini انجام شد . (1600_samples_gemini.docx)
- ۳- ساخت دیتاست پزشکی برای آموزش مدل با کد پایتون انجام شد. (1000_samples_for_fine_tuning_better_than_the_previous_one.xlsx) اما چونکه تنوع استفاده از نام فارسی و دارو و موارد مختلف کم بود و ساختار جملات هم تکراری بود ، برای آموزش و فاین تیون کردن مدل مناسب نیست و نیاز به بهبود کد پایتونش است .

بخش دوم پیش پردازش :

تبدیل فرمت فایل pdf لیست_دارویی به فرمت csv و تمیز کردن فایل csv با کد پایتون و تبدیل آن به فایل drug_lists.xlsx انجام شد که فقط شامل ستونهای نام دارو به فارسی، دوز دارو، نوع دارو است.

در ادامه، مدل های مختلف spell checking بررسی شدند که هیچکدام از آنها مناسب نبودند و نتایج در جدول زیر قابل مشاهده اند:

| Model name | Size(G) | Accuracy |
|---|-----------------------------------|-------------------------------------|
| xlm-roberta-base | 1.13 | 1% |
| Ahmad/parsT5-base | 1.98 | 0% |
| HooshvareLab/bert-fa-base-uncased | 1.31 | 0% |
| HooshvareLab/bert-fa-zwnj-base | 0.95 | 0% |
| https://github.com/eteamin/async_faspell | Python code with your own dataset | 0% |
| meta-llama/Llama-3.1-8B-Instruct | 16 | Need permission from meta(rejected) |

بخش سوم تست مدل های متنی مختلف :

تست مدل های مختلف زبانی سایت هاگینگ فیس برای پیدا کردن مدل های مناسبی که توانایی فرمت بندی نسخه های پزشکی فارسی را داشته باشند. مدل هایی که بررسی شدند، داخل فایل models.xlsx موجود هستند که دقت و زمان و حجم همه آنها بررسی شده است .

| Model name | Cpu/Gpu | Size(G) | Time(S) | Accuracy |
|---|---------|---------|---------|----------|
| TheBloke/openchat_3.5-GPTQ | gpu | 4.16 | 22.38 | 80% |
| PartAI/Dorna2-Llama3.1-8B-Instruct | cpu | 16.07 | 92.16 | 90% |
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 5.01 | 90% |
| mistralai/Mistral-7B-Instruct-v0.2 | cpu | 14.48 | 237.1 | 70% |
| mistralai/Mistral-7B-Instruct-v0.2 | gpu | 14.48 | 10.98 | 70% |
| HuggingFaceH4/zephyr-7b-alpha | cpu | 12.52 | 235.62 | 50% |
| HuggingFaceH4/zephyr-7b-alpha | gpu | 12.52 | 12.02 | 20% |
| NousResearch/Nous-Hermes-2-Mistral-7B-DPO | cpu | 14.48 | 268.19 | 20% |
| NousResearch/Nous-Hermes-2-Mistral-7B-DPO | gpu | 14.48 | 9.67 | 50% |
| teknium/OpenHermes-2.5-Mistral-7B | cpu | 14.48 | 219.15 | 50% |
| teknium/OpenHermes-2.5-Mistral-7B | gpu | 14.48 | 9.24 | 50% |
| MaralGPT/Maral-7B-alpha-1 | cpu | 14.5 | 220.36 | 0% |
| MaralGPT/Maral-7B-alpha-1 | gpu | 14.5 | 12.14 | 0% |
| universitytehran/PersianMind-v1.0 | cpu | 13.63 | 88.02 | 80% |
| universitytehran/PersianMind-v1.0 | gpu | 13.63 | 3.45 | 70% |
| openchat/openchat_3.5 | cpu | 14.48 | 236.25 | 80% |
| openchat/openchat_3.5 | gpu | 14.48 | 12.12 | 70% |
| rishiraj/smol-7b | cpu | 14.48 | 283.87 | 40% |
| rishiraj/smol-7b | gpu | 14.48 | 17.36 | 60% |

| | | | | |
|--|-----|------|-------|------|
| repo_id="asedmammad/PersianMind-v1.0-GGUF", filename="PersianMind-v1.0.q6_K.gguf" | cpu | 5.59 | 17.98 | 40% |
| repo_id="TheBloke/Mistral-7B-Instruct-v0.2-GGUF", filename="mistral-7b-instruct-v0.2.Q6_K.gguf" | cpu | 5.94 | 56.93 | 80% |
| repo_id="bartowski/mistralai_Mistral-Small-3.2-24B-Instruct-2506-GGUF", filename="mistralai_Mistral-Small-3.2-24B-Instruct-2506-Q4_K_M.gguf" | cpu | 14.3 | 43.82 | 85% |
| repo_id="unsloth/Magistral-Small-2506-GGUF", filename="Magistral-Small-2506-Q4_K_M.gguf" | cpu | 14.3 | 47.58 | 100% |
| repo_id="mradermacher/gemma-3-4b-persian-v0-GGUF", filename="gemma-3-4b-persian-v0.Q4_K_M.gguf" | cpu | 2.49 | 4.97 | 20% |
| repo_id="tensorblock/Gemma2_Farsi-GGUF", filename="Gemma2_Farsi-Q4_K_M.gguf" | cpu | 1.63 | 14.69 | 0% |
| repo_id="tensorblock/Maral-7B-alpha-1-GGUF", filename="Maral-7B-alpha-1-Q4_K_M.gguf" | cpu | 4.37 | 32.29 | 0% |
| repo_id="mostafaamiri/persian-llama-7b-GGUF-Q4", filename="persian_llama_7b.Q4_K_M.gguf" | cpu | 4.19 | 12.38 | 0% |
| repo_id="PartAI/Dorna-Llama3-8B-Instruct-GGUF", filename="dorna-llama3-8b-instruct.Q4_K_M.gguf" | cpu | 4.92 | 27.65 | 80% |
| repo_id = "mshojaei77/gemma-3-4b-persian-v0" filename = "gemma-3-4b-persian-v0_q8_0.gguf" | cpu | 4.13 | 8.94 | 60% |
| repo_id="YanaS/llama-2-7b-langchain-chat-GGUF", filename="llama-2-7b-langchain-chat-Q6_K.gguf" | cpu | 5.53 | 54.6 | 70% |

هشت مدل برای تست روی نسخه های متنوع و پرامپت های متفاوت انتخاب شدند که در بخش

فاین تیون ، نتایج بررسی بیشتر این مدل ها را قرار داده ام.

در ادامه ، 8 مدلی که استاد گفتند ، دقیق بررسی شد . در جدول زیر نتایج را مشاهده میکنید ، با 20 نسخه پزشکی و پرامپت های مختلف بررسی شدند :

| Model name | Cpu/Gpu | Size(G) | Time(S) | Accuracy | prompt |
|-----------------------|---------|---------|---------|----------|--|
| HooshvareLab/gpt-2-fa | cpu | 0.49 | 2.8 | 0% | لطفاً اطلاعات زیر را استخراج کن: نام بیمار، تشخیص بیماری، داروها (نام، دوز، نحوه مصرف، توصیه، نوع). پاسخ: |
| HooshvareLab/gpt-2-fa | cpu | 0.49 | 2.8 | 0% | تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص". ####فرمت خروجی: نام بیمار...: تشخیص بیماری...: _____ اسم دارو...: دوز...: نحوه مصرف...: توصیه...: نوع...: (تکرار برای هر دارو) |
| openai-community/gpt2 | cpu | 0.55 | 3.2 | 0% | لطفاً اطلاعات زیر را استخراج کن: نام بیمار، تشخیص بیماری، داروها (نام، دوز، نحوه مصرف، توصیه، نوع). پاسخ: |
| openai-community/gpt2 | cpu | 0.55 | 3.2 | 0% | تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. |

| | | | | | |
|------------------------|-----|------|------|----|---|
| | | | | | <p>فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص".</p> <p>####فرمت خروجی:</p> <p>نام بیمار...:</p> <p>تشخیص بیماری...:</p> <p>اسم دارو...:</p> <p>دوز...:</p> <p>نحوه مصرف...:</p> <p>توصیه...:</p> <p>نوع...:</p> <p>(تکرار برای هر دارو)</p> |
| google/gemma-3n-E2B-it | gpu | 10.9 | 32.8 | 0% | <p>تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص".</p> <p>####فرمت خروجی:</p> <p>نام بیمار...:</p> <p>تشخیص بیماری...:</p> <p>اسم دارو...:</p> <p>دوز...:</p> <p>نحوه مصرف...:</p> <p>توصیه...:</p> <p>نوع...:</p> <p>(تکرار برای هر دارو)</p> |
| google/gemma-3n-E4B-it | gpu | 15.8 | 18.3 | 0% | <p>تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص".</p> <p>####فرمت خروجی:</p> |

| | | | | | |
|------------------------------------|-----|-------|-----|-----|--|
| | | | | | نام بیمار... : تشخیص بیماری... : اسم دارو... : دوز... : نحوه مصرف... : توصیه... : نوع... : (تکرار برای هر دارو) |
| universitytehran/PersianMind-v1.0 | gpu | 13.63 | 3.2 | 5% | "شما یک دستیار هوشمند هستید که باید از نسخه زیر اطلاعات ساختاریافته استخراج کنید": - نام بیمار "\n" - داروها و دوز "\n" مصرفی "\n" - طریقه مصرف "\n" - تشخیص "\n" بیماری "\n" f" نسخه {input_text} "\n" "\n" " پاسخ:" |
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 6.3 | 80% | شما یک دستیار پزشکی هستید. لطفاً فقط اطلاعات زیر را استخراج کن و هیچ توضیح اضافی نده: - نام بیمار - داروها و دوز مصرفی - طریقه مصرف - تشخیص بیماری فقط همین اطلاعات را بده. |
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 7.7 | 85% | تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص". |

| | | | | | |
|------------------------------------|-----|-------|-----|-----|--|
| | | | | | <p>####فرمت خروجی:</p> <p>نام بیمار...:</p> <p>تشخیص بیماری...:</p> <p>اسم دارو...:</p> <p>دوز...:</p> <p>نحوه مصرف...:</p> <p>توصیه...:</p> <p>نوع...:</p> <p>(تکرار برای هر دارو)</p> |
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 8.2 | 70% | <p>از متن نسخه، فقط این اطلاعات را استخراج کن:</p> <p>1. نام بیمار.</p> <p>2. تشخیص بیماری.</p> <p>3. لیست داروها شامل:</p> <p>- نام دارو</p> <p>- دوز (مثلاً: ۵۰۰ میلی گرم، ۱ عدد، نصف قرص)</p> <p>- نحوه مصرف (مثلاً: روزی دو بار بعد غذا)</p> <p>- توصیه پزشک (مثلاً: با آب زیاد، در صورت تهوع قطع شود)</p> <p>- نوع دارو (مثلاً: قرص، شربت، تزریقی)</p> <p>قالب دقیق خروجی:</p> <p>نام بیمار: [مثال: علی رضایی]</p> <p>تشخیص بیماری: [مثال: عفونت گلو]</p> <p>دارو ۱:</p> <p>نام: [مثال: آموکسی سیلین]</p> <p>دوز: [مثال: ۵۰۰ میلی گرم]</p> <p>نحوه مصرف: [مثال: هر ۸ ساعت یک بار بعد غذا]</p> <p>توصیه: [مثال: تا ۷ روز مصرف شود]</p> <p>نوع: [مثال: کپسول]</p> |

| | | | | | |
|------------------------------------|-----|-------|------|-----|---|
| | | | | | (اگر داروهای بیشتری هست، ادامه بده به همین شکل) فقط اطلاعات خواسته شده را بده. توضیح یا تفسیر ننویس. |
| mistralai/Mistral-7B-Instruct-v0.2 | gpu | 14.48 | 4.6 | 20% | شما یک دستیار پزشکی هستید. لطفاً فقط اطلاعات زیر را استخراج کن و هیچ توضیح اضافی نده: -نام بیمار -داروها و دوز مصرفی -طریقه مصرف -تشخیص بیماری فقط همین اطلاعات را بده. |
| mistralai/Mistral-7B-Instruct-v0.2 | gpu | 14.48 | 14.4 | 30% | تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص". ####فرمت خروجی: نام بیمار...: تشخیص بیماری...: _____ اسم دارو...: دوز...: نحوه مصرف...: توصیه...: نوع...: (تکرار برای هر دارو) |
| mistralai/Mistral-7B-Instruct-v0.2 | gpu | 14.48 | 11.5 | 20% | از متن نسخه، فقط این اطلاعات را استخراج کن: 1.نام بیمار 2.تشخیص بیماری 3.لیست داروها شامل: - نام دارو |

| | | | | | |
|------------------------------------|-----|-------|------|-----|---|
| | | | | | <p>- دوز (مثلاً: ۵۰۰ میلی گرم، ۱ عدد، نصف قرص)</p> <p>- نحوه مصرف (مثلاً: روزی دوبار بعد غذا)</p> <p>- توصیه پزشک (مثلاً: با آب زیاد، در صورت تهوع قطع شود)</p> <p>- نوع دارو (مثلاً: قرص، شربت، تزریقی)</p> <p>قالب دقیق خروجی:</p> <p>نام بیمار: [مثال: علی رضایی]</p> <p>تشخیص بیماری: [مثال: عفونت گلو]</p> <p>دارو ۱:</p> <p>نام: [مثال: آموکسی سیلین]</p> <p>دوز: [مثال: ۵۰۰ میلی گرم]</p> <p>نحوه مصرف: [مثال: هر ۸ ساعت یکبار بعد غذا]</p> <p>توصیه: [مثال: تا ۷ روز مصرف شود]</p> <p>نوع: [مثال: کپسول]</p> <p>(اگر داروهای بیشتری هست، ادامه بده به همین شکل)</p> <p>فقط اطلاعات خواسته شده را بده. توضیح یا تفسیر ننویس.</p> |
| mistralai/Mistral-7B-Instruct-v0.2 | gpu | 14.48 | 10.7 | 30% | <p>شما یک دستیار هوشمند هستید که باید از متن نسخه پزشکی، اطلاعات زیر را استخراج کرده و نمایش دهید. جوابی که در خروجی به من میدی به زبان فارسی باشد:</p> <p>- نام بیمار</p> <p>- نوع دارو و دوز مصرفی</p> <p>- طریقه مصرف</p> <p>- تشخیص بیماری توسط پزشک</p> <p>فقط همان اطلاعات را استخراج کن و هیچ توضیح اضافه نده.</p> |

| | | | | | |
|------------------------------------|-----|------|------|-----|--|
| mistralai/Mistral-7B-Instruct-v0.3 | gpu | 14.5 | 13.6 | 50% | شما یک دستیار هوشمند هستید که باید از متن نسخه پزشکی، اطلاعات زیر را استخراج کرده و نمایش دهید. جوابی که در خروجی به من میدی به زبان فارسی باشد: - نام بیمار - نوع دارو و دوز مصرفی - طریقه مصرف - تشخیص بیماری توسط پزشک فقط همان اطلاعات را استخراج کن و هیچ توضیح اضافه نده. |
|------------------------------------|-----|------|------|-----|--|

بخش چهارم ترین مدل :

—

بخش پنجم فاین تیون :

هشت مدل انتخاب شده در بخش تست مدل های زبانی ، با ۲۰ نسخه پزشکی و ۲ پرامپت متفاوت تست و بررسی شدند و داخل combined_prescriptions_output.pdf ذخیره شده اند. نتایج کلی را داخل جدول زیر قرار داده ام :

| Model name | Cpu/Gpu | Size(G) | Time(S) | Accuracy | Prompt(old or new) |
|--|---------|---------|---------|----------|--------------------|
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 5 | 70% | new |
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 5 | 90% | old |
| TheBloke/openchat_3.5-GPTQ | gpu | 4.16 | 54 | 50% | new |
| TheBloke/openchat_3.5-GPTQ | gpu | 4.16 | 54 | 40% | old |
| universitytehran/PersianMind-v1.0 | gpu | 13.63 | 20 | 40% | new |
| universitytehran/PersianMind-v1.0 | gpu | 13.63 | 14 | 0% | old |
| repo_id = "mshojaei77/gemma-3-4b-persian-v0"filename = "gemma-3-4b-persian-v0_q8_0.gguf" | cpu | 4.13 | 15 | 80% | new |
| repo_id = "mshojaei77/gemma-3-4b-persian-v0"filename = "gemma-3-4b-persian-v0_q8_0.gguf" | cpu | 4.13 | 10 | 50% | old |
| repo_id="YanaS/llama-2-7b-langchain-chat- | cpu | 5.53 | 50 | 0% | new |

| | | | | | |
|---|-----|------|----|-----|-----|
| GGUF",filename="llama-2-7b-langchain-chat-Q6_K.gguf | | | | | |
| repo_id="YanaS/llama-2-7b-langchain-chat-GGUF",filename="llama-2-7b-langchain-chat-Q6_K.gguf | cpu | 5.53 | 50 | 0% | old |
| repo_id="PartAI/Doma-Llama3-8B-Instruct-GGUF",filename="doma-llama3-8b-instruct.Q4_K_M.gguf | cpu | 4.92 | 27 | 80% | new |
| repo_id="PartAI/Doma-Llama3-8B-Instruct-GGUF",filename="doma-llama3-8b-instruct.Q4_K_M.gguf | cpu | 4.92 | 22 | 50% | old |
| repo_id="unsloth/Magistral-Small-2506-GGUF",filename="Magistral-Small-2506-Q4_K_M.gguf | cpu | 14.3 | 80 | 70% | new |
| repo_id="unsloth/Magistral-Small-2506-GGUF",filename="Magistral-Small-2506-Q4_K_M.gguf | cpu | 14.3 | 55 | 50% | old |
| repo_id="TheBloke/Mistral-7B-Instruct-v0.2-GGUF",filename="mistral-7b-instruct-v0.2.Q6_K.gguf | cpu | 5.94 | 50 | 50% | new |
| repo_id="TheBloke/Mistral-7B-Instruct-v0.2-GGUF",filename="mistral-7b-instruct-v0.2.Q6_K.gguf | cpu | 5.94 | 80 | 40% | old |

Old prompt:

شما یک دستیار پزشکی هستید. لطفاً فقط اطلاعات زیر را استخراج کن و هیچ توضیح اضافی نده:

-نام بیمار

-داروها و دوز مصرفی

-طریقه مصرف

-تشخیص بیماری

فقط همین اطلاعات را بده.

New prompt:

شما یک دستیار پزشکی هستید. لطفاً فقط اطلاعات زیر را استخراج کن و هیچ توضیح اضافی نده:

نام بیمار:

تشخیص بیماری:

اسم دارو:

دوز:

نحوه مصرف: (چند بار در روز)

توصیه: (نسبت به غذا)

نوع: قرص / آمپول / پماد

فقط همین اطلاعات را بده.

سه مدل زیر به دلیل دقت خوب و قابل قبولشان ، برای ادامه کار انتخاب شدند :

1) PartAI/Dorna2-Llama3.1-8B-Instruct

2) repo_id = "mshojaei77/gemma-3-4b-persian-v0" filename = "gemma-3-4b-persian-v0_q8_0.gguf"

3) repo_id="PartAI/Dorna-Llama3-8B-Instruct-GGUF",filename="dorna-llama3-8b-instruct.Q4_K_M.gguf"

در ادامه ، بررسی تاثیر پارامترهای خروجی روی دقت و زمان مدل انجام شد.

4 مدل انتخاب شدند که وظیفه من ، بررسی بهترین مقدار ۳ پارامتر خروجی تولید جواب بود . همه تست ها روی یک پرامپت مشخص انجام شده اند تا نتایج قابل مقایسه کردن باشند . تست ها روی ۴ مدل تعیین شده انجام شدند . نتایج و توضیحات جامع فعالیتیم را داخل فایل final_results.xlsx قرار داده ام . با بررسی نتایج مشخص شد که پارامترها تاثیر بسیار کمی روی زمان اجرای مدل دارند و بیشتر روی دقت پاسخ تاثیر گذارند .

4 مدلی که تست ها روی آنها انجام شد :

- 1) repo_id = "mshojaei77/gemma-3-4b-persian-v0" filename = "gemma-3-4b-persian-v0_q8_0.gguf"
- 2) repo_id="PartAI/Dorna-Llama3-8B-Instruct-GGUF",filename="dorna-llama3-8b-instruct.Q4_K_M.gguf"
- 3) "repo_id": "unsloth/gemma-3n-E2B-it-GGUF", "filename": "gemma-3n-E2B-it-Q5_K_S.gguf"
- 4) "repo_id": "PartAI/Dorna-Llama3-8B-Instruct-GGUF", "filename": "dorna-llama3-8b-instruct.Q4_K_S.gguf"

نتایج را در جدول زیر قرار داده ام.

| Model name | Cpu/Gpu | Size(G) | token | temperature | top_p | Time(S) | Accuracy |
|--------------|---------|---------|-------|-------------|-------|---------|-----------|
| Dorna Q4_KS | cpu | 4.69 | 80 | 0.35 | 0.85 | 80 | جواب ناقص |
| Dorna Q4_KS | cpu | 4.69 | 90 | 0.35 | 0.85 | 85 | جواب ناقص |
| Dorna Q4_KS | cpu | 4.69 | 95 | 0.35 | 0.85 | 89 | جواب ناقص |
| Dorna Q4_KS | cpu | 4.69 | 100 | 0.35 | 0.85 | 90 | جواب ناقص |
| Dorna Q4_KS | cpu | 4.69 | 105 | 0.35 | 0.85 | 94 | جواب ناقص |
| Dorna Q4_KS | cpu | 4.69 | 110 | 0.35 | 0.85 | 99 | جواب ناقص |
| Dorna Q4_KS | cpu | 4.69 | 115 | 0.35 | 0.85 | 98 | جواب ناقص |
| Dorna Q4_KS | cpu | 4.69 | 120 | 0.35 | 0.85 | 100 | جواب ناقص |
| Dorna Q4_KS | cpu | 4.69 | 125 | 0.35 | 0.85 | 102 | جواب ناقص |
| Dorna Q4_KS | cpu | 4.69 | 130 | 0.35 | 0.85 | 104 | جواب ناقص |
| Dorna Q4_KS | cpu | 4.69 | 135 | 0.35 | 0.85 | 101 | جواب ناقص |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.35 | 0.85 | 81 | 80% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.4 | 0.85 | 88 | 75% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.3 | 0.85 | 101 | 90% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.25 | 0.85 | 101 | 75% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.35 | 0.8 | 103 | 70% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.35 | 0.9 | 96 | 75% |
| Dorna Q4_KS | cpu | 4.69 | 256 | 0.3 | 0.85 | 98 | 90% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.35 | 0.95 | 100 | 75% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.35 | 0.75 | 98 | 70% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.26 | 0.85 | 106 | 70% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.27 | 0.85 | 104 | 70% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.28 | 0.85 | 105 | 71% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.29 | 0.85 | 98 | 69% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.31 | 0.85 | 98 | 85% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.32 | 0.85 | 96 | 85% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.33 | 0.85 | 93 | 85% |
| Dorna Q4_KS | cpu | 4.69 | 140 | 0.34 | 0.85 | 95 | 85% |
| Gemma Q5_K_S | cpu | 3.26 | 140 | 0.35 | 0.85 | 58 | 80% |
| Gemma Q5_K_S | cpu | 3.26 | 140 | 0.4 | 0.85 | 55 | 75% |
| Gemma Q5_K_S | cpu | 3.26 | 140 | 0.3 | 0.85 | 54 | 90% |
| Gemma Q5_K_S | cpu | 3.26 | 140 | 0.25 | 0.85 | 56 | 75% |
| Gemma Q5_K_S | cpu | 3.26 | 140 | 0.3 | 0.8 | 56 | 80% |
| Gemma Q5_K_S | cpu | 3.26 | 140 | 0.3 | 0.95 | 56 | 85% |
| Gemma Q5_K_S | cpu | 3.26 | 160 | 0.28 | 0.85 | 56 | 84% |
| Gemma Q8_0 | cpu | 4.13 | 160 | 0.3 | 0.85 | 85 | 90% |
| Gemma Q8_0 | cpu | 4.13 | 140 | 0.3 | 0.85 | 77 | 90% |
| Gemma Q8_0 | cpu | 4.13 | 140 | 0.35 | 0.85 | 76 | 80% |
| Gemma Q8_0 | cpu | 4.13 | 140 | 0.4 | 0.85 | 74 | 75% |
| Gemma Q8_0 | cpu | 4.13 | 140 | 0.25 | 0.85 | 76 | 75% |

| | | | | | | | |
|-------------|-----|------|-----|------|------|-----|-----|
| Gemma Q8_0 | cpu | 4.13 | 140 | 0.3 | 0.8 | 88 | 80% |
| Gemma Q8_0 | cpu | 4.13 | 140 | 0.3 | 0.9 | 83 | 85% |
| Dorna Q4_KM | cpu | 4.92 | 160 | 0.4 | 0.85 | 101 | 75% |
| Dorna Q4_KM | cpu | 4.92 | 160 | 0.35 | 0.85 | 96 | 80% |
| Dorna Q4_KM | cpu | 4.92 | 160 | 0.3 | 0.85 | 98 | 90% |
| Dorna Q4_KM | cpu | 4.92 | 160 | 0.25 | 0.85 | 95 | 75% |
| Dorna Q4_KM | cpu | 4.92 | 160 | 0.28 | 0.85 | 94 | 84% |
| Dorna Q4_KM | cpu | 4.92 | 160 | 0.3 | 0.8 | 98 | 80% |
| Dorna Q4_KM | cpu | 4.92 | 160 | 0.3 | 0.9 | 96 | 85% |
| Dorna Q4_KM | cpu | 4.92 | 140 | 0.3 | 0.85 | 97 | 90% |

پرامپت استفاده شده برای پیدا کردن بهترین پارامترهای خروجی به شرح زیر است:

""<im_start>system

شما یک دستیار پزشکی دقیق هستید. لطفاً از متن نسخه زیر اطلاعات زیر را با فرمت خواسته شده استخراج کن.

هیچ توضیح اضافی ننویس. اگر اطلاعاتی وجود نداشت، آن را با "نامشخص" پر کن.

خروجی باید دقیقاً به شکل زیر باشد:

نام بیمار... :

تشخیص بیماری... :

اسم دارو... :

دوز... :

نحوه مصرف... :

توصیه... :

نوع... :

اگر بیش از یک دارو در نسخه بود، این بخش را برای هر دارو تکرار کن.

```
<|im_end|>
<|im_start|>user
{input_text}
<|im_end|>
<|im_start|>assistant
.....
```

پارامترهای خروجی با توجه به دقت مدل ها به این شکل انتخاب شدند :

token=256 - temperature=0.3 - top_p=0.85

در ادامه کد فاین تیون مدل درنای اصلی با استفاده از lora ساخته شد و با تعداد داده های کم برای آموزش ، بدون ارور اجرا شد. فایلی که کد فاین تیون ، داخل آن وجود دارد، first_fine_tuning_dorna.ipynb نام دارد.

در ادامه چندین پرامپت مختلف را امتحان کردم تا متوجه شوم که آیا میتوان با مدل درنای اصلی ، نسخه های دارای غلط های املائی و نگارشی را بر طرف کرد که متوجه شدم که مدل درنای اصلی قابلیت برطرف کردن این مشکلات را ندارد. دو مدل mistralv0.2 و persianmind را با نسخه های دارای غلط املائی و نگارشی هم بررسی کردم و بخاطر اینکه این دو مدل دقتشان در نسخه های عادی ، ضعیف بود ، اینجا هم نتایج خوبی نداشتند.

| Model name | Cpu/Gpu | Size(G) | Time(S) | Accuracy | prompt |
|------------------------------------|---------|---------|---------|----------|---|
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 6.1 | 30% | شما یک دستیار پزشکی هستید. لطفاً فقط اطلاعات زیر را استخراج کن و هیچ توضیح اضافی نده: -نام بیمار -داروها و دوز مصرفی -طریقه مصرف -تشخیص بیماری فقط همین اطلاعات را بده. |
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 7.4 | 35% | تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص." ####فرمت خروجی: نام بیمار...: تشخیص بیماری...: |

| | | | | | |
|------------------------------------|-----|-------|-----|-----|---|
| | | | | | <p>اسم دارو... :</p> <p>دوز... :</p> <p>نحوه مصرف :</p> <p>...</p> <p>توصیه... :</p> <p>نوع... :</p> <p>(تکرار برای هر دارو)</p> |
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 7.8 | 25% | <p>از متن نسخه، فقط این اطلاعات را استخراج کن:</p> <p>1. نام بیمار</p> <p>2. تشخیص بیماری</p> <p>3. لیست داروها شامل:</p> <p>- نام دارو</p> <p>- دوز (مثلاً: ۵۰۰ میلی گرم، ۱ عدد، نصف قرص)</p> <p>- نحوه مصرف (مثلاً: روزی دو بار بعد غذا)</p> <p>- توصیه پزشک (مثلاً: با آب زیاد، در صورت تهوع قطع شود)</p> <p>- نوع دارو (مثلاً: قرص، شربت، تزریقی)</p> <p>قالب دقیق خروجی:</p> |

| | | | | | |
|------------------------------------|-----|-------|-----|-----|---|
| | | | | | نام بیمار: [مثال: علی رضایی] تشخیص بیماری: [مثال: عفونت گلو] دارو ۱: نام: [مثال: آموکسی سیلین] دوز: [مثال: ۵۰۰ میلی گرم] نحوه مصرف: [مثال: هر ۸ ساعت یک بار بعد غذا] توصیه: [مثال: تا ۷ روز مصرف شود] نوع: [مثال: کپسول] (اگر داروهای بیشتری هست، ادامه بده به همین شکل) فقط اطلاعات خواسته شده را بده. توضیح یا تفسیر ننویس. |
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 7.3 | 20% | تو یک دستیار پزشکی هوشمند هستی. نسخه هایی که بهت می دیم ممکنه شامل |

| | | | | | |
|--|--|--|--|--|---|
| | | | | | <p>غلط‌های تایپی، نگارشی، گرامری یا اشتباهات دوز و نحوه مصرف باشن. وظیفه تو اینه که با درک دقیق متن نسخه، اطلاعات زیر رو اصلاح شده و استاندارد طبق قالب خواسته شده استخراج کنی. اگر اطلاعات ناقص یا نامشخص بود، فقط بنویس «نامشخص.»</p> <p>فقط و فقط خروجی نهایی رو در قالب زیر نمایش بده. هیچ توضیح یا تفسیر اضافی ننویس.</p> <p>فرمت خروجی: نام بیمار... : تشخیص بیماری... : _____</p> <p>اسم دارو... : دوز... : نحوه مصرف : ... توصیه... :</p> |
|--|--|--|--|--|---|

| نوع ... : (تکرار برای هر دارو) | | | | | |
|--|-----|-------|-----|-----|------------------------------------|
| تو یک دستیار پزشکی هوشمند هستی. نسخه‌هایی که بهت داده می‌شن ممکنه شامل غلط‌های تایپی، نگارشی، گرامری، ترتیب کلمات، دوزهای اشتباه یا حتی اشتباه در نوشتن نام داروها باشن. هدف تو اینه که با درک دقیق معنای نسخه، اطلاعات ساختاریافته زیر رو تصحیح شده و استاندارد استخراج کنی. اگر هر بخش از اطلاعات (مثلاً دوز، توصیه، نوع دارو یا نام بیمار) مشخص نبود یا ناقص بود، فقط بنویس "نامشخص". | gpu | 16.07 | 7.1 | 30% | PartAI/Dorna2-Llama3.1-8B-Instruct |

| | | | | | |
|------------------------------------|-----|-------|-----|-----|--|
| | | | | | <p>فقط و فقط</p> <p>خروجی را در</p> <p>قالب زیر ارائه</p> <p>بده. هیچ</p> <p>توضیح یا</p> <p>تفسیر یا دلیل</p> <p>اضافه ننویس.</p> <p>فرمت</p> <p>خروجی:</p> <p>نام بیمار...:</p> <p>تشخیص</p> <p>بیماری...:</p> <p>-----</p> <p>اسم دارو...:</p> <p>دوز...:</p> <p>نحوه مصرف:</p> <p>...</p> <p>توصیه...:</p> <p>نوع...:</p> <p>(تکرار برای</p> <p>هر دارو)</p> |
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 7.3 | 35% | <p>تو یک دستیار</p> <p>پزشکی</p> <p>هوشمند</p> <p>هستی.</p> <p>نسخه‌هایی که</p> <p>بهت داده</p> <p>می‌شن ممکنه</p> <p>شامل غلط‌های</p> <p>تایپی، نگارشی،</p> <p>گرامری،</p> <p>دوزهای اشتباه</p> <p>یا فواصل</p> <p>مصرف</p> <p>غیرمنطقی</p> <p>باشن.</p> |

| | | | | | |
|--|--|--|--|--|---|
| | | | | | <p>نکات مهمی</p> <p>که باید اصلاح کنی:</p> <p>دوز دارو باید واقع بینانه و رایج باشد. مثلاً «۵۰۰۰ mg» برای یک قرص استامینوفن غیر واقعی هست و باید به مقدار معقول مثل «۵۰۰ mg» اصلاح بشه.</p> <p>نحوه مصرف باید در بازه زمانی معقول باشد. مثلاً اگر گفته شده «هر ۱۰۰ ساعت» اصلاحش کن به چیزی مثل «هر ۶ ساعت» یا بنویس «نامشخص» اگر قابل تشخیص نبود.</p> <p>هدف تو اینه که با درک دقیق متن نسخه، اطلاعات زیر رو اصلاح شده و استاندارد استخراج کنی.</p> |
|--|--|--|--|--|---|

| | | | | | |
|-----------------------------------|-----|-------|-----|-----|--|
| | | | | | <p>اگر هر بخش از اطلاعات (مثلاً دوز، توصیه، نوع دارو یا نام بیمار) مشخص نبود یا غیرقابل تشخیص بود، فقط بنویس "نامشخص".</p> <p>فقط و فقط خروجی را در قالب زیر ارائه بده. هیچ توضیح یا تفسیر یا دلیل اضافه ننویس.</p> <p>فرمت خروجی:</p> <p>نام بیمار...:</p> <p>تشخیص بیماری...:</p> <p>-----</p> <p>اسم دارو...:</p> <p>دوز...:</p> <p>نحوه مصرف:</p> <p>...</p> <p>توصیه...:</p> <p>نوع...:</p> <p>(تکرار برای هر دارو)</p> |
| universitytehran/PersianMind-v1.0 | gpu | 13.63 | 6.8 | 15% | <p>تو یک دستیار پزشکی هوشمند هستی. نسخه‌هایی که بهت می‌دیم</p> |

| | | | | | |
|--|--|--|--|--|--|
| | | | | | <p>ممکنه شامل غلط‌های تایپی، نگارشی، گرامری یا اشتباهات دوز و نحوه مصرف باشن. وظیفه تو اینه که با درک دقیق متن نسخه، اطلاعات زیر رو اصلاح شده و استاندارد طبق قالب خواسته شده استخراج کنی. اگر اطلاعات ناقص یا نامشخص بود، فقط بنویس «نامشخص».</p> <p>فقط و فقط خروجی نهایی رو در قالب زیر نمایش بده. هیچ توضیح یا تفسیر اضافی ننویس.</p> <p>فرمت خروجی: نام بیمار: ... تشخیص بیماری: ... اسم دارو: ... دوز: ...</p> |
|--|--|--|--|--|--|

| | | | | | |
|-----------------------------------|-----|-------|-----|-----|--|
| | | | | | نحوه مصرف: ... توصیه: ... نوع: ... (تکرار برای هر دارو) |
| universitytehran/PersianMind-v1.0 | gpu | 13.63 | 7.1 | 20% | تو یک دستیار پزشکی هوشمند هستی. نسخه‌هایی که بهت داده می‌شن ممکنه شامل غلط‌های تایپی، نگارشی، گرامری، ترتیب کلمات، دوزهای اشتباه یا حتی اشتباه در نوشتن نام داروها باشن. هدف تو اینه که با درک دقیق معنای نسخه، اطلاعات ساختاریافته زیر رو تصحیح شده و استاندارد استخراج کنی. اگر هر بخش از اطلاعات (مثلاً دوز، توصیه، نوع دارو یا نام بیمار) مشخص نبود یا ناقص بود، فقط |

| | | | | | |
|------------------------------------|-----|-------|-----|----|---|
| | | | | | <p>بنویس "نامشخص".</p> <p>فقط و فقط خروجی را در قالب زیر ارائه بده. هیچ توضیح یا تفسیر یا دلیل اضافه ننویس.</p> <p>فرمت خروجی: نام بیمار: ... تشخیص بیماری: ...</p> <hr/> <p>اسم دارو: ... دوز: ... نحوه مصرف: ... توصیه: ... نوع: ... (تکرار برای هر دارو)</p> |
| mistralai/Mistral-7B-Instruct-v0.2 | gpu | 14.48 | 5.2 | 5% | <p>شما یک دستیار پزشکی هستید. لطفاً فقط اطلاعات زیر را استخراج کن و هیچ توضیح اضافی نده: - نام بیمار - داروها و دوز مصرفی - طریقه مصرف - تشخیص بیماری</p> |

| | | | | | |
|------------------------------------|-----|-------|------|----|---|
| | | | | | فقط همین اطلاعات را بده. |
| mistralai/Mistral-7B-Instruct-v0.2 | gpu | 14.48 | 12.4 | 5% | تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشته، بنویس "نامشخص". #### خروجی: نام بیمار: ... تشخیص بیماری: ... اسم دارو: ... دوز: ... نحوه مصرف: ... توصیه: ... نوع: ... (تکرار برای هر دارو) |
| mistralai/Mistral-7B-Instruct-v0.2 | gpu | 14.48 | 12 | 5% | از متن نسخه، فقط این اطلاعات را استخراج کن: ۱. نام بیمار ۲. تشخیص بیماری |

| | | | | | |
|--|--|--|--|--|--|
| | | | | | ۳. لیست داروها شامل: - نام دارو - دوز (مثلاً): ۵۰۰ میلی گرم، ۱ عدد، نصف قرص) - نحوه مصرف (مثلاً): روزی دوبار بعد غذا) - توصیه پزشک (مثلاً: با آب زیاد، در صورت تهوع قطع شود) - نوع دارو (مثلاً: قرص، شربت، تزریقی) قالب دقیق خروجی: نام بیمار: [مثال: علی رضایی] تشخیص بیماری: [مثال: عفونت گلو] _____ دارو ۱: نام: [مثال: آموکسی سیلین] دوز: [مثال: ۵۰۰ میلی گرم] نحوه مصرف: [مثال: هر ۸ ساعت یکبار بعد غذا] |
|--|--|--|--|--|--|

| | | | | | |
|------------------------------------|-----|-------|------|----|--|
| | | | | | <p>توصیه: [مثال]: تا ۷ روز مصرف شود] نوع: [مثال]: کپسول]</p> <p>(اگر داروهای بیشتری هست، ادامه بده به همین شکل)</p> <p>فقط اطلاعات خواسته شده را بده. توضیح یا تفسیر ننویس.</p> |
| mistralai/Mistral-7B-Instruct-v0.2 | gpu | 14.48 | 10.2 | 5% | <p>تو یک دستیار پزشکی هوشمند هستی. نسخه‌هایی که بهت می‌دیم ممکنه شامل غلط‌های تایپی، نگارشی، گرامری یا اشتباهات دوز و نحوه مصرف باشن. وظیفه تو اینه که با درک دقیق متن نسخه، اطلاعات زیر رو اصلاح شده و استاندارد طبق قالب خواسته شده استخراج کنی. اگر اطلاعات ناقص یا نامشخص بود،</p> |

| | | | | | |
|------------------------------------|-----|-------|------|----|--|
| | | | | | فقط بنویس «نامشخص». |
| | | | | | فقط و فقط خروجی نهایی رو در قالب زیر نمایش بده. هیچ توضیح یا تفسیر اضافی ننویس. |
| | | | | | فرمت خروجی: نام بیمار: ... تشخیص بیماری: ... |
| | | | | | اسم دارو: ... دوز: ... نحوه مصرف: ... توصیه: ... نوع: ... (تکرار برای هر دارو) {, "" |
| mistralai/Mistral-7B-Instruct-v0.2 | gpu | 14.48 | 12.1 | 5% | تو یک دستیار پزشکی هوشمند هستی. نسخه‌هایی که بهت داده می‌شن ممکنه شامل غلط‌های تایپی، نگارشی، گرامری، ترتیب کلمات، دوزهای اشتباه یا حتی اشتباه |

| | | | | | |
|--|--|--|--|--|--|
| | | | | | <p>در نوشتن نام داروها باشن.</p> <p>هدف تو اینه که با درک دقیق معنای نسخه، اطلاعات ساختاریافته زیر رو تصحیح شده و استاندارد استخراج کنی.</p> <p>اگر هر بخش از اطلاعات (مثلاً دوز، توصیه، نوع دارو یا نام بیمار) مشخص نبود یا ناقص بود، فقط بنویس "نامشخص".</p> <p>فقط و فقط خروجی را در قالب زیر ارائه بده. هیچ توضیح یا تفسیر یا دلیل اضافه ننویس.</p> <p>فرمت خروجی:</p> <p>نام بیمار: ...</p> <p>تشخیص بیماری: ...</p> <hr/> <p>اسم دارو: ...</p> <p>دوز: ...</p> |
|--|--|--|--|--|--|

| | | | | | |
|------------------------------------|-----|-------|-----|----|---|
| | | | | | <p>نحوه مصرف:</p> <p>...</p> <p>توصیه: ...</p> <p>نوع: ...</p> <p>(تکرار برای هر دارو)</p> |
| mistralai/Mistral-7B-Instruct-v0.2 | gpu | 14.48 | 9.1 | 5% | <p>تو یک دستیار پزشکی هوشمند هستی. نسخه‌هایی که بهت داده می‌شن ممکنه شامل غلط‌های تایپی، نگارشی، گرامری، دوزهای اشتباه یا فواصل مصرف غیرمنطقی باشن.</p> <p>نکات مهمی که باید اصلاح کنی:</p> <p>دوز دارو باید واقع‌بینانه و رایج باشه. مثلاً «mg۵۰۰۰» برای یک قرص استامینوفن غیرواقعی هست و باید به مقدار معقول مثل «mg۵۰۰» اصلاح بشه.</p> |

| | | | | | |
|--|--|--|--|--|---|
| | | | | | <p>نحوه مصرف باید در بازه زمانی معقول باشد. مثلاً اگر گفته شده «هر ۱۰۰ ساعت»، اصلاً حش کن به چیزی مثل «هر ۶ ساعت» یا بنویس «نامشخص» اگر قابل تشخیص نبود.</p> <p>هدف تو اینه که با درک دقیق متن نسخه، اطلاعات زیر رو اصلاح شده و استاندارد استخراج کنی.</p> <p>اگر هر بخش از اطلاعات (مثلاً دوز، توصیه، نوع دارو یا نام بیمار) مشخص نبود یا غیرقابل تشخیص بود، فقط بنویس "نامشخص".</p> <p>فقط و فقط خروجی را در قالب زیر ارائه بده. هیچ توضیح یا</p> |
|--|--|--|--|--|---|

| | | | | | |
|--|--|--|--|--|--|
| | | | | | تفسیر یا دلیل اضافه ننویس. فرمت خروجی: نام بیمار: ... تشخیص بیماری: ... <hr/> اسم دارو: ... دوز: ... نحوه مصرف: ... توصیه: ... نوع: ... (تکرار برای هر دارو) |
|--|--|--|--|--|--|

مدل درنای اصلی را با متن های محاوره ای پزشکی تست کردم و مشخص شد که این مدل میتواند متن محاوره ای را به متن رسمی تبدیل کند .

در ادامه برای بهبود دقت مدل درنای اصلی ، پرامپت های مختلفی را تست کردم که در جدول زیر قرار داده ام :

| Model name | Cpu/Gpu | Size(G) | Time(S) | Accuracy | prompt |
|------------------------------------|---------|---------|---------|----------|--|
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 7 | 97% | تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود |

| | | | | | |
|--|-----|-------|------|-----|--|
| | | | | | <p>نداشت، بنویس "نامشخص".</p> <p>هیچ توضیح یا تفسیر اضافی ننویس.</p> <p>#### فرمت</p> <p>خروجی: نام بیمار: ... تشخیص بیماری: ... _____</p> <p>اسم دارو: ... دوز: ... نحوه مصرف و توصیه: ... نوع: ... (تکرار برای هر دارو)</p> |
| PartAI/Dorna2- Llama3.1-8B- Instruct | gpu | 16.07 | 11.5 | 70% | <p>تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی در قالب زیر ارائه بده و هیچ توضیح یا تفسیر اضافه ننویس. اگر اطلاعات وجود نداشت، بنویس «نامشخص».</p> <p>توصیه‌های کلی پزشک که مستقیماً به نحوه مصرف دارو مربوط نیستند را در بخش</p> |

| | | | | | |
|--|-----|-------|-----|-----|---|
| | | | | | <p>«توصیه‌های کلی پزشک» بنویس.</p> <p>فرمت خروجی: نام بیمار: ... تشخیص بیماری: ... _____</p> <p>اسم دارو: ... دوز: ... نحوه مصرف و توصیه (فقط مربوط به همان دارو): ... نوع: ... (تکرار برای هر دارو) _____</p> <p>توصیه‌های کلی پزشک: ...</p> |
| PartAI/Dorna2- Llama3.1-8B- Instruct | gpu | 16.07 | 8.3 | 60% | <p>تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص".</p> <p>هیچ توضیح یا تفسیر اضافی ننویس.</p> <p>#### فرمت خروجی: نام بیمار: ... تشخیص بیماری: ...</p> |

| توصیه اضافی: (در صورت وجود داشتن داخل نسخه) | اسم دارو: ... | دوز: ... | نحوه مصرف و توصیه: ... | نوع: ... | (تکرار برای هر دارو) |
|---|---------------|----------|------------------------|----------|--|
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 6.8 | 95% | <p>تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص".</p> <p>هیچ توضیح یا تفسیر اضافی ننویس.</p> <p>موردی در فرمت خروجی از خودت اضافه نکن و فقط اطلاعاتی که در نسخه ها موجود است را در مواردی که موجود است، بنویس.</p> <p>#### فرمت خروجی:</p> <p>نام بیمار: ...</p> |

| | | | | | |
|------------------------------------|-----|-------|---|-----|--|
| | | | | | <p>تشخیص بیماری:</p> <p>...</p> <hr/> <p>اسم دارو: ...</p> <p>دوز: ...</p> <p>نحوه مصرف و</p> <p>توصیه: ...</p> <p>نوع: ...</p> <p>(تکرار برای هر دارو)</p> |
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 7 | 80% | <p>تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص".</p> <p>هیچ توضیح یا تفسیر اضافی ننویس.</p> <p>#### فرمت</p> <p>خروجی:</p> <p>نام بیمار: ...</p> <p>تشخیص بیماری:</p> <p>...</p> <hr/> <p>اسم دارو: ...</p> <p>دوز: ...</p> <p>نحوه مصرف و</p> <p>توصیه: ...</p> <p>نوع: ...</p> <p>(تکرار برای هر دارو)</p> |

| | | | | | |
|------------------------------------|-----|-------|-----|-----|---|
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 6.9 | 96% | <p>تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص".</p> <p>هیچ توضیح یا تفسیر اضافی ننویس. موردی در فرمت خروجی از خودت اضافه نکن.</p> <p>#### فرمت خروجی:</p> <p>نام بیمار: ...</p> <p>تشخیص بیماری: ...</p> <hr/> <p>اسم دارو: ...</p> <p>دوز: ...</p> <p>نحوه مصرف و توصیه: ...</p> <p>نوع: ...</p> <p>(تکرار برای هر دارو)</p> |
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 7.1 | 85% | <p>تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن.</p> |

| | | | | | |
|--|-----|-------|-----|-----|--|
| | | | | | <p>فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس «نامشخص». هیچ توضیح یا تفسیر اضافی ننویس. هیچ موردی خارج از ساختار و ترتیب فیلدهای فرمت خروجی اضافه نکن. توصیه‌های کلی پزشک که به داروی خاصی مربوط نیستند را نادیده بگیر و در خروجی بنویس. فرمت خروجی: نام بیمار: ... تشخیص بیماری: ... _____</p> <p>اسم دارو: ... دوز: ... نحوه مصرف و توصیه: ... نوع: ... (تکرار برای هر دارو)</p> |
| PartAI/Dorna2- Llama3.1-8B- Instruct | gpu | 16.07 | 7.1 | 86% | <p>تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی</p> |

| | | | | | |
|------------------------------------|-----|-------|-----|-----|--|
| | | | | | <p>درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص".</p> <p>#### فرمت خروجی:</p> <p>نام بیمار: ...</p> <p>تشخیص بیماری: ...</p> <hr/> <p>اسم دارو: ...</p> <p>دوز: ...</p> <p>نحوه مصرف و توصیه: ...</p> <p>نوع: ...</p> <p>(تکرار برای هر دارو)</p> |
| PartAI/Dorna2-Llama3.1-8B-Instruct | gpu | 16.07 | 7.4 | 75% | <p>تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص".</p> <p>#### فرمت خروجی:</p> <p>نام بیمار: ...</p> <p>تشخیص بیماری: ...</p> <hr/> <p>اسم دارو: ...</p> <p>دوز: ...</p> <p>الگوی مصرف: ...</p> |

| | | | | | |
|--|--|--|--|--|--|
| | | | | | نحوه مصرف: ... نوع: ... (تکرار برای هر دارو) |
|--|--|--|--|--|--|

ذخیره پرامپت مشخص و همچنین مقادیر مشخص پارامترهای خروجی روی فایل مدل اصلی درنا انجام شد و مدل هم داخل گوگل درایو ذخیره شد و دانلود شد. ذخیره مدل اصلی درنا بدون تغییرات در فایل‌هایش انجام شد و مدل هم داخل گوگل درایو ذخیره شد و دانلود شد. همچنین مدل درنای Q4KS هم داخل گوگل درایو ذخیره شد و دانلود شد. مدل درنای Q4KM هم داخل گوگل درایو ذخیره شد و دانلود شد.

فایل زیپ مدل اصلی درنا که پارامترهای خروجی آن و پرامپت آن به شکل زیر داخل مدل تعریف شده اند :

token=256 - temperature=0.3 - top_p=0.85

prompt=

تو یک دستیار پزشکی هستی. اطلاعات نسخه زیر را دقیقاً با فرمت مشخص شده استخراج کن. فقط خروجی درخواست شده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس "نامشخص".

فرمت خروجی:

نام بیمار: ...

تشخیص بیماری: ...

اسم دارو: ...

دوز: ...

نحوه مصرف و توصیه: ...

نوع: ...

(تکرار برای هر دارو)

لینک دانلود مدل درنای اصلی با موارد ذخیره شده از گوگل درایو :

<https://drive.google.com/file/d/1OY76vxLFmry2nSeCqRceLY1vX6mjaXe2/view?usp=sharing>

لینک دانلود مدل درنای اصلی بدون اعمال تغییرات از گوگل درایو :

<https://drive.google.com/file/d/1MB449LK8yDA263eOYppi46GvCcKYE8uM/view?usp=sharing>

لینک دانلود مدل Dorna Q4_K_M بدون اعمال تغییرات از گوگل درایو :

<https://drive.google.com/file/d/19pEYhugHzi5DpW8N72BjPJuffqPCImR3/view?usp=sharing>

لینک دانلود مدل Dorna Q4_K_S بدون اعمال تغییرات از گوگل درایو :

https://drive.google.com/file/d/1pOeaDo5mDCfkkAK3_JTplo4Y2mOI9Lmg/view?usp=sharing