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	0%
	your own dataset	
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%

محمدعلى لطفى كوشالي

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	cpu	0.49	2.8	0%	لطفاً اطلاعات زير را استخراج
2-fa					كن:
					نام بیمار، تشخیص بیماری،
					داروها (نام، دوز، نحوه مصرف،
					توصيه، نوع).
					پاسخ:
HooshvareLab/gpt	cpu	0.49	2.8	0%	تو یک دستیار پزشکی هستی.
2-fa	1				اطلاعات نسخه زير را دقيقاً با
					فرمت مشخص شده استخراج كن.
					فقط خروجی درخواستشده را
					ارائه بده. اگر اطلاعات وجود
					نداشت، بنويس "نامشخص."
					###فرمت خروجي:
					نام بيمار:
					تشخیص بیماری:
					 اسم دارو :
					اسم دارو دوز :
					دور نحوه مصرف :
					توصيه:
					نوع:
					نوع)تکرار برای هر دارو(
)تحورار برای هر دارور
openai-	cpu	0.55	3.2	0%	لطفاً اطلاعات زير را استخراج
community/gpt2	Сри	0.55	3.2	070	کن:
					س. نام بیمار، تشخیص بیماری،
					داروها (نام، دوز، نحوه مصرف،
					توصيه، نوع).
					پاسخ:
					. يسپ
openai- community/gpt2	cpu	0.55	3.2	0%	تو یک دستیار پزشکی هستی.
Tommanity, gpt2					اطلاعات نسخه زير را دقيقاً با
					فرمت مشخص شده استخراج كن.

	ı	1	ı	ı	Γ
					فقط خروجی درخواستشده را
					ارائه بده. اگر اطلاعات وجود
					نداشت، بنویس "نامشخص."
					###فرمت خروجي:
					نام بيمار:
					تشخیص بیماری :
					دوز:
					نحوه مصرف:
					توصيه:
					نوع:
)تکرار برای هر دارو(
google/gemma-3n-	gpu	10.9	32.8	0%	تو یک دستیار پزشکی هستی.
E2B-it	Sp u	10.5	32.0	070	لو یک دستیر پرسامی مسلی. اطلاعات نسخه زیر را دقیقاً با
					فرمت مشخص شده استخراج كن.
					فقط خروجی درخواستشده را
					ارائه بده. اگر اطلاعات وجود
					نداشت، بنویس "نامشخص."
					نداست، بنویس نامستخص،
					###فرمت خروجي:
					نام بيمار :
					تشخیص بیماری:
					اسم دارو:
					دوز:
					نحوه مصرف:
					توصيه:
					نوع:
)تکرار برای هر دارو(
google/gemma-	gpu	15.8	18.3	0%	تو یک دستیار پزشکی هستی.
3n-E4B-it					اطلاعات نسخه زير را دقيقاً با
					فرمت مشخص شده استخراج كن.
					فقط خروجی درخواستشده را
					ارائه بده. اگر اطلاعات وجود
					نداشت، بنويس "نامشخص."
					###فرمت خروجى:
				-	

		1	1		,
					نام بيمار:
					تشخیص بیماری :
					اسم دارو:
					دوز:
					نحوه مصرف:
					توصيه:
					نوع:
)تکوار برای هر دارو(
		12.52		=0.4	
universitytehran	gpu	13.63	3.2	5%	"شما یک دستیار هوشمند هستید
/PersianMind- v1.0					که باید از نسخه زیر اطلاعات
V1.0					ساختاريافته استخراج كنيد"n\:
					-" نام بیمار"n
					-" داروها و دوز
					مصرفی "n'
					-" طريقه مصرف"n
					-" تشخيص
					یان بیماری"n\n
					ا" :input_text:نسخه f''
					}\n\n"
					, عد عد " پاسخ":
PartAI/Dorna2-	gpu	16.07	6.3	80%	شما یک دستیار پزشکی هستید.
Llama3.1-8B-	8r ·				يو پر لطفاً فقط اطلاعات زير را
Instruct					استخراج کن و هیچ توضیح
					اضافی نده:
					-نام بیمار
					-داروها و دوز مصرفی
					-طريقه مصرف
					-تشخیص بیماری
					فقط همین اطلاعات را بده.
PartAI/Dorna2-	gpu	16.07	7.7	85%	تو یک دستیار پزشکی هستی.
Llama3.1-8B- Instruct					اطلاعات نسخه زير را دقيقاً با
					فرمت مشخص شده استخراج كن.
					فقط خروجی درخواستشده را
					ارائه بده. اگر اطلاعات وجود
					نداشت، بنویس "نامشخص."
					<i>J. G</i> -2, J-1, Canton
		I	I	I	

	ı	T		Ι	
					###فرمت خروجي:
					نام بيمار :
					تشخیص بیماری:
					 اسم دارو :
					,
					دوز:
					نحوه مصرف:
					توصيه:
					نوع:
)تکرار برا <i>ی</i> هر دارو(
PartAI/Dorna2- Llama3.1-8B-	gpu	16.07	8.2	70%	از متن نسخه، فقط این اطلاعات
Instruct					را استخراج كن:
					.1نام بیمار
					.2تشخیص بیماری
					.3ليست داروها شامل:
					- نام دارو
					- دوز (مثلاً: ۵۰۰ میلیگرم، 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،
					ف شربت، تزری <i>قی</i>)
					قالب دقيق خروجي:
					نام بیمار: [مثال: علی رضایی]
					تشخیص بیماری: [مثال: عفونت
					گلو]
					 دارو ۱:
					نام: [مثال: آموكسي سيلين]
					دوز: [مثال: ٥٠٠ میلی گرم]
					نحوه مصرف: [مثال: هر ۸ ساعت
					يکبار بعد غذا]
					توصیه: [مثال: تا ۷ روز مصرف
					شود]
					نوع: [مثال: كپسول]

)اگر داروهای بیشتری هست،
					ادامه بده به همین شکل(
					فقط اطلاعات خواسته شده را بده.
					توضيح يا تفسير ننويس.
mistralai/Mistral- 7B-Instruct-v0.2	gpu	14.48	4.6	20%	شما یک دستیار پزشکی هستید.
/B-mstruct-vo.2					لطفاً فقط اطلاعات زير را
					استخراج کن و هیچ توضیح
					اضافی نده:
					-نام بيمار
					-داروها و دوز مصرفی
					-طريقه مصرف
					-تشخیص بیماری
					فقط همين اطلاعات را بده.
					. 3 . 0.
mistralai/Mistral- 7B-Instruct-v0.2	gpu	14.48	14.4	30%	تو یک دستیار پزشکی هستی.
/B-IIIstruct-v0.2					اطلاعات نسخه زير را دقيقاً با
					فرمت مشخص شده استخراج كن.
					فقط خروجی درخواستشده را
					ارائه بده. اگر اطلاعات وجود
					نداشت، بنويس "نامشخص."
					###فرمت خروجي:
					نام بيمار :
					تشخیص بیماری:
					اسم دارو:
					دوز:
					نحوه مصرف:
					توصيه:
					نوع:
					۔)تکرار برای هر دارو(
mistralai/Mistral-	gpu	14.48	11.5	20%	از متن نسخه، فقط این اطلاعات
7B-Instruct-v0.2	Ci				را استخراج كن:
					.1نام بيمار
					.2تشخیص بیماری
					.3لیست داروها شامل:
					۔ - نام دارو
					23

	T	T		1	
					- دوز (مثلاً: ٥٠٠ میلیگرم، 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، تزری <i>قی</i>)
					قالب دقيق خروجي:
					نام بیمار: [مثال: علی رضایی]
					تشخیص بیماری: [مثال: عفونت
					گلو]
					 دارو ۱:
					نام: [مثال: آموكسى سيلين]
					دوز: [مثال: ٥٠٠ میلیگرم]
					نحوه مصرف: [مثال: هر ۸ ساعت
					یکبار بعد غذا]
					توصیه: [مثال: تا ۷ روز مصرف
					شود]
					نوع: [مثال: كپسول]
)اگر داروهای بیشتری هست،
					ادامه بده به همین شکل (
					فقط اطلاعات خواستهشده را بده.
					توضيح يا تفسير ننويس.
mistralai/Mistral-	gpu	14.48	10.7	30%	شما یک دستیار هوشمند هستید که
7B-Instruct-v0.2	gpu	14.40	10.7	3070	باید از متن نسخه پزشکی، اطلاعات
					زیر را استخراج کرده و نمایش دهید.
					جوابی که در خروجی به من میدی به
					زبان فارسى باشد:
					- نام بیمار
					- نوع دارو و دوز مصرفی
					- طريقه مصرف
					- تشخیص بیماری توسط پزشک
					فقط همان اطلاعات را استخراج کن و هیچ توضیح اضافه نده.
					س و هميچ تو صبيح اسات مان

mistralai/Mistral-	gpu	14.5	13.6	50%	شما یک دستیار هوشمند هستید که
7B-Instruct-v0.3					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
					- طريقه مصرف
					- تشخیص بیماری توسط پزشک
					فقط همان اطلاعات را استخراج
					كن و هيچ توضيح اضافه نده.

بخش چهارم ترین مدل:

_

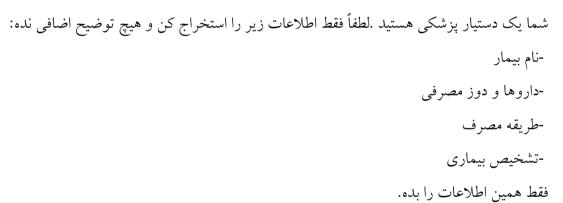
بخش پنجم فاین تیون:

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

COLUMN C1 H11 2 71		Ι	1	1	
GGUF",filename="llama-2-7b-					
"langchain-chat-Q6_K.gguf					
repo_id="YanaS/llama-2-7b-	cpu	5.53	50	0%	old
langchain-chat-	1				
GGUF",filename="llama-2-7b-					
"langchain-chat-Q6_K.gguf					
repo_id="PartAI/Dorna-Llama3-8B-	cpu	4.92	27	80%	new
Instruct-GGUF",filename="dorna-	1	_			
"llama3-8b-instruct.Q4_K_M.gguf					
repo_id="PartAI/Dorna-Llama3-8B-	cpu	4.92	22	50%	old
Instruct-GGUF",filename="dorna-	Pu	,2		20,0	014
"llama3-8b-instruct.Q4_K_M.gguf					
repo_id="unsloth/Magistral-Small-	cpu	14.3	80	70%	new
2506-GGUF",filename="Magistral-	Сри	1 1.5		7070	ne w
"Small-2506-Q4_K_M.gguf					
repo_id="unsloth/Magistral-Small-	cpu	14.3	55	50%	old
2506-GGUF",filename="Magistral-	Сри	1 1.5		3070	old
"Small-2506-Q4_K_M.gguf					
repo id="TheBloke/Mistral-7B-	cnu	5.94	50	50%	new
Instruct-v0.2-	cpu	3.74	30	3070	iie w
GGUF",filename="mistral-7b-					
"instruct-v0.2.Q6_K.gguf					
repo id="TheBloke/Mistral-7B-	cpu	5.94	80	40%	old
Instruct-v0.2-	Cpu	3.74	30	1070	oid
GGUF",filename="mistral-7b-					
"instruct-v0.2.Q6_K.gguf					

Old prompt:



New prompt:

شما یک دستیار پزشکی هستید. لطفاً فقط اطلاعات زیر را استخراج کن و هیچ توضیح اضافی ن	نده:
نام بیمار:	
تشخیص بیماری:	
اسم دارو:	

محمدعلي لطفي كوشالي

دوز:

نحوه مصرف: (چند بار در روز)

توصيه: (نسبت به غذا)

نوع: قرص/آمپول/پماد

فقط همين 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 ک مدل تعیین شده انجام شدند . نتایج و توضیحات جامع فعالیتم 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 تاثیرگذارند .

4 مدلى كه تست ها روى 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%

(متن) محمدعلى لطفى كوشالي

گزارش جامع پروژه پایانی deep learning (متن)

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.

خروجي بايد دقيقاً به شكل زير باشد:

نام بيمار ... :

تشخیص بیماری...:

اسم دارو...:

دوز...:

نحوه مصرف...:

توصيه...:

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|>
<|iim_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 املایی 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 مشکلات را با نسخه های دارای غلط املایی 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-	gpu	16.07	6.1	30%	شما یک
Instruct					دستيار پزشكى
					هستيد. لطفاً
					فقط اطلاعات
					زير را
					استخراج کن و
					هيچ توضيح
					اضافی نده:
					-نام بيمار
					-داروها و دوز
					مصرفي
					-طريقه
					مصرف
					-تشخيص
					بيماري
					فقط همين
					اطلاعات را
					بده.
PartAI/Dorna2-Llama3.1-8B-	gpu	16.07	7.4	35%	تو یک دستیار
Instruct					پزشکی هستی.
					اطلاعات
					نسخه زير را
					دقيقاً با فرمت
					مشخص شده
					استخراج كن.
					فقط خروجي
					درخواستشده
					را ارائه بده.
					اگر اطلاعات
					وجود نداشت،
					بنويس
					"نامشخص."
					###فرمت
					خروجي:
					نام بیمار:
					تشخيص
					بيمارى:

	1				1
					ـــــــــــــــــــــــــــــــــــــ
					دوز:
					نحوه مصرف:
					توصيه:
					نوع:
)تکرار برای
					هر دارو(
PartAI/Dorna2-Llama3.1-8B-	gpu	16.07	7.8	25%	از متن نسخه،
Instruct					فقط این
					اطلاعات را
					استخراج كن:
					(),1
					.1نام بیمار
					2.تشخیص
					بیماری 12
					. 3ليست دا دا دا دا دا
					داروها شامل:
					- نام دارو - دوز (مثلاً:
					۰۰۰ دور رممار. ۵۰۰ میل <i>ی</i> گرم،
					۱ عدد، نصف
					ا عدد، نصف قرص)
					- نحوه
					ر مصرف (مثلاً:
					ر روزی دوبار
					بعد غذا)
					· - توصیه
					پزشک (مثلاً: با
					آب زیاد، در
					صورت تھوع
					قطع شود)
					- نوع دارو
					(مثلاً: قرص،
					شربت،
					تزریقی)
					قالب دقيق
					خروجي:

					نام بیمار:
					[مثال: على
					رضایی]
					تشخيص
					بيمارى: [مثال:
					عفونت گلو]
					 دارو ۱:
					نام: [مثال:
					ا آموكسىسيلين]
					دوز: [مثال:
					٥٠٠ ميليگرم]
					نحوه مصرف:
					[مثال: هر ۸
					ساعت يكبار
					بعد غذا]
					توصيه: [مثال:
					تا ۷ روز
					مصرف شود]
					نوع: [مثال:
					كپسول]
)اگر داروهای
					بیشتری هست،
					ادامه بده به
					همین شکل(
					فقط اطلاعات
					خواستهشده را
					بده. توضیح یا
					خواستهشده را بده. توضیح یا تفسیر ننویس.
PartAI/Dorna2-Llama3.1-8B-	gpu	16.07	7.3	20%	تو یک دستیار
Instruct	<i>C</i> 1				
					پزشکی هوشمند
					هستي.
					هستی. نسخههایی که بهت میدیم ممکنه شامل
					- بهت میدیم
					ممکنه شامل
				<u> </u>	

محمدعلى لطفى كوشالى	گزارش جامع پروژه پایانی deep learning (متن)
	غلطهای تایبی،
	نگارشی،
	گرامری یا
	اشتباهات دوز
	و نحوه مصرف
	باشن. وظيفه
	تو اینه که با
	درک دقیق متن
	نسخه،
	اطلاعات زير
	رو اصلاح شده
	و استاندارد
	طبق قالب
	خواستهشده
	استخراج كني.
	اگر اطلاعات
	ناقص یا
	نامشخص بود،
	فقط بنويس
	«نامشخص.«
	فقط و فقط
	خروجی نهایی
	رو در قالب
	زیر نمایش
	بده. هیچ
	توضيح يا
	تفسير اضافى
	ننویس.
	فرمت
	خروجي:
	نام بيمار :
	تشخیص
	بيمارى :
	اسم دارو:
	دوز:
	نحوه مصرف :
	توصيه:

					نوع:
)تکرار برای
					هر دارو(
D-#41/D2 112 1 9D		1.6.07	7.1	200/	
PartAI/Dorna2-Llama3.1-8B- Instruct	gpu	16.07	7.1	30%	تو یک دستیار
					پزشك <i>ى</i>
					هوشمند
					هستى.
					نسخەھايى كە
					بهت داده
					مىشن ممكنه
					شامل غلطهای
					تايپى، نگارشى،
					گرامری،
					ترتیب کلمات،
					دوزهای اشتباه
					یا حتی اشتباه
					در نوشتن نام
					داروها باشن.
					هدف تو اینه
					که با درک
					دقیق معنای
					نسخه،
					اطلاعات
					ساختاريافته
					زیر رو
					تصحيح شده و
					استاندارد
					استخراج كني.
					اگر هر بخش
					از اطلاعات
					(مثلاً دوز،
					توصيه، نوع
					دارو یا نام
					بيمار) مشخص
					نبود يا ناقص
					بود، فقط
					بنويس
					"نامشخص."

					فقط و فقط
					خطو و قطط خروجی را در
					قالب زير ارائه
					ونب رير ارات بده. هيچ
					بدو. ميچ توضيح يا
					تفسير يا دليل
					اضافه ننویس.
					اصافه ننویس.
					فرمت
					خروجي:
					نام بيمار:
					تشخيص
					بيمارى:
					اسم دارو:
					دوز:
					نحوه مصرف:
					 توصيه:
					نوع :
					ر (تکرار برای
					هر دارو)
PartAI/Dorna2-Llama3.1-8B-	gpu	16.07	7.3	35%	تو یک دستیار
Instruct					يزشكي
					. و هو شمند
					پزشکی هوشمند هستی.
					نسخەھايى كە
					بهت داده
					،، مىشن ممكنه
					شامل غلطهای
					تايپى، نگارشى،
					یپی و یی گرامری،
					ر ری دوزهای اشتباه
					يا فواصل يا فواصل
					مصرف
					غيرمنطقى
					ير باشن.
					<i>J</i> .

محمدعلي لطفي كوشالي	dee (متن)	ep learning	پروژه پایانی ب	گزارش جامع
				نکات مهمی
				که باید اصلاح
				 کنی:
				دوز دارو باید
				واقعبينانه و
				رايج باشه. مثلاً
				mg» o···»
				برای یک
				قرص
				استامينوفن
				غيرواقعي
				هست و باید
				به مقدار معقول
				مثل
				mg» ٥٠٠»
				اصلاح بشه.
				نحوه مصرف
				باید در بازه
				زمانی معقول
				باشه. مثلاً اگر
				گفته شده «هر
				۱۰۰ ساعت»،
				اصلاحش كن
				به چیزی مثل
				«هر ٦ ساعت»
				يا بنويس
				«نامشخص»
				اگر قابل
				تشخيص نبود.
				هدف تو اینه
				هدف نو اینه که با درک
				که با در ت دقیق متن
				دفيق منن نسخه،
				نسخه، اطلاعات زیر
				اطارعات ریر رو اصلاحشده
				و استاندارد
				استخراج كني.

كو شالى	لطف	محمدعلي
ے سے	صعي	ے جہاں جہاں

					اگر هر بخش
					از اطلاعات
					ر مثلاً دوز، (مثلاً دوز،
					رممار دور. توصیه، نوع
					دارو یا نام
					بیمار) مشخص
					نبود یا غیرقابل
					تشخیص بود،
					فقط بنويس
					"نامشخص."
					فقط و فقط
					ر خروج <i>ی</i> را در
					قالب زير ارائه
					. وير و بده. هيچ
					. ت توضیح یا
					تفسیر یا دلیل
					در يــــــــــــــــــــــــــــــــــــ
					0
					فرمت
					خروجي:
					نام بيمار:
					تشخيص
					بیماری:
					اسم دارو:
					دوز:
					نحوه مصرف :
					توصیه :
					نوع:
					(تکرار برای هر دارو)
universitytehran/PersianMind-v1.0	anu	13.63	6.8	15%	
am versity contain i erstamvimu-vi.0	gpu	13.03	0.8	1370	تو یک دستیار
					پزشک <i>ی</i> همشمند
					هوشمند هستی. نسخههایی که بهت میدیم
					ا هستی.
					سحههایی ت
					بهت می دیم

سكند شامل الماده التي الماده التي الماده التي التي الماده التي التي التي التي التي التي التي التي	محمدعلى لطفى كوشالى	گزارش جامع پروژه پایانی deep learning (متن)
شله های تایی، کارشی،		ممكنه شاما
کارشوی یا اشتیاهات دور کراهوی یا اشتیاهات دور و ترجوه مصوف الشیاهات دور و ترجوه مصوف یا بیشت وظیفه و تحویه مصوف یا بیشت وظیفه دوری دونیق متن تو اینه که یا اطلاعات زیر و اصلاح شده طبق قالب و استاندارد و واصلاح شده طبق قالب و استاندارد و تواستخداه مشاهدات استخراء کنی، فقط بنویس نامشخص بود، نافشخص بود، نافشخص بود، نافشخص بود، نافش یا در دو قالب خروجی نهایی و شقط و فقط زیر نمایش و و در قالب بود، هیچ زیر نمایش و در قالب نویس نامشخص نامشخص نویس نامشخص به نامشخص نویس نامشخص به نامشخص داد، هیچ زیر نمایش و در قالب نویس نامشخص نامشخص نامشخص نویس نامشخص نام بیمار نامس دادر نامشخص نام بیمار نامشخص نام بیمار نامشخص نام بیمار نامشخص نام		
گرامری با استداهات دور و نحوه مصرف باشن، وظیفه و الدین که با استخدات دیر اسانداد و استخداج کنی، خواستخداه طبق قالب و استخداج کنی، تخاستخداج کنی، نامیخصی بود، نامیخصی بود، نامیخصی بود، نفوسی، نوبی، نام بیماز، بنام بیماز،		
الشاهات دور و نحوه مصرف بالشن، وظبقه و نحوه مصرف توابد که با السائن، وظبقه السخه السخه السخه السخه السخه الملاعات السخواج کتی، خواسته الملاعات الم		
ر نحوه مصرف توایه که با باشن، وظیه توایه که با درک دقیق مثن درک دقیق مثن را اصلاح شده اطالاعات زیر را استخدای خواسته شده خواسته شده خواسته شده با اگر اطلاعات باشخص بوده نامشخص، فقط بنوس، فقط و فقط مزر بخی نهایی فقط و فقط زیر نمایش رو در قالب بده هیچ زیر نمایش توضیح یا نوس، نتوس، نتوس، نتوس، نتوس، نتوس، نتوس، نتوس، خورجی، خورجی، خورجی، خورجی، نام بیمار: بیماری:		
باشن، وظيفه در دقيق متن در دقيق متن در دقيق متن اطلاحات زير اساتاداره و اساتاداره و اساتاداره و المساتخده طبق قالب الكر اطلاحات الكر اطلاحات الكر اطلاحات الكر اطلاحات المشخص بوده ققط بنوس المشخص، ققط و فقط مامشخص، ققط و فقط مردوجي نهايي و در قالب خروجي نهايي توضيح با يده هيج ترضيح با توسير اضافي توضيح با توسير اضافي خروجي: خورجي: خورجي: خورجي: الم يمادن ييمارين ييمارين		
ورک دقیق متن السخه، اطلاعات زیر و اصلاح شده و اصلاح شده و اصلاح شده طبق قالب و استاندارد و و اصلاح شده طبق قالب خواسته شده المستخراج کنی، خواسته شده یا اگر اطلاعات استخراج کنی، نقش یا اگر اطلاعات نامشخص بود، نقش بنویس نامشخص». فقط و فقط مناسشخص». فقط و فقط در تالب خروجی نهایی توضیع یا یده هیچ زیر نمایش دو در قالب توسی یا یده هیچ نیویس. نشویس. نشویس. خروجی: هومت نام بیمهار: خروجی: شخیص یا ییمهاری: نشخیص یا ییمهاری: ییمهاری:		باشن. وظيفه
نسخه، اطلاعات زیر او اصلاح شده و استانداره و استانداره علی قالب خواستهشده اشخراج کن، اگر اطلاعات اگر اطلاعات اگر اطلاعات اشخص بود، القص یا اقط بنویس الشخص، اقظ و فقط المنافض		تو اینه که با
اطلاعات زبر و اصلاح شده و استانداره و استانداره علی قالب خواسته شده استخراج کن. اگر اطلاعات اگر اطلاعات اگر اطلاعات المشخص بود، القص یا المشخص بود، القط و فقط المشخص المشخص المناب خروجی نهایی ققط و فقط و در قالب خروجی نهایی توضیح یا یده هیچ زبر نمایش توضیح یا یده هیچ نوس. تفسیر اضافی توضیح بام بیمار: سنجیس بام بیمار: سیماری: سیماری: سم دارو: سم مدارو: سم سم دارو: سم مدارو:		درک دقیق متن
رو اصلاح شده و استاندارد و استاندارد طبق قالب خواسته شده خواسته شده اشتخراج کنی، اگر اطلاحات اگر اطلاحات المشخص بوده انامشخص بوده انامشخص به انامشخص به انامشخص به انامشخص به انامشخص به انامشخص به خورجی نهایی توسیح با بده هیچ نوسیح با نوسیح ب		نسخه،
و استانداره طبق قالب خواسته شده خواسته شده استخراج کنی، اگر اطلاعات اگر اطلاعات المشخص، ه ققط بنویس منامشخص، هقط و فقط منامشخص، و در قالب خروجی نهایی توسیح یا یده، هیچ زیر نمایش توسیح یا یده، هیچ نویس، نامسافی توصیح یا یده، هیچ نویس، نام بیمار: خروجی: مومت بیماری: بیماری:		اطلاعات زير
طبق قالب خواسته شده خواسته شده الكر اطلاعات الكر اطلاعات المشخص بود، ناقص با فقط و فقط والمشخص، فقط و فقط والمشخص، وو در قالب خووجي نهايي توضيح يا بده. هيچ توضيح يا نده. هيچ توضيح يا ندوس، تقسير اضافي توضيح يا ندوس، تشخيص نام بيمان خووجي:		رو اصلاحشده
خواسته شده الكراعات المستحص بود، المقط بنويس الماشخص الله المقط و فقط المقط و فقط و فقط و فقط و زر نمايش وو در قالب يده هيچ از بر نمايش توضيح يا يده هيچ الكراعات ال		و استاندارد
المنتخراج كني. الأواطلاعات الأمشخص بود، المشخص بود، المشخص، وققط المنتخص، وققط المنتخص، وققط المنتخص، وقوط المنتخص، وود وقالب المنتخص، وود ود وقالب المنتخص، وود ود وقالب المنتخص، وود ود وقالب المنتخص، وود ود والله المنتخص، وود ود والله المنتخص المنتخص، والمنتخص، وا		
اگر اطلاعات نامشخص بود، نامشخص بود، فقط بنویس فقط و فقط فقط و فقط خروجی نهایی زر دمایش رو در قالب بده. هیچ توضیح یا بده. هیچ توضیح یا ننویس. ننویس. خروجی: فرمت ننام بیمار: تشخیص نام بیمار:		
نامشخص بود، فقط بنویس فقط بنویس فقط و فقط فقط و فقط خروجی نهایی و در قالب زیر نمایش زیر نمایش توضیح یا بلده هیچ توضیح یا ننویس. ننویس. خروجی: فرمت نام بیمار: تشخیص بیماری:		
نامشخص بود، فقط بنویس فقط و فقط فقط و فقط خروجی نهایی زیر نمایش زیر نمایش توضیح یا ننویس. ننویس. خروجی: فرمت نام بیمار: تشخیص بیماری:		
ققط بنویس انامشخصا». افقط و فقط خوجی نهایی رو در قالب رو در قالب بده. هیچ توضیح یا بده. هیچ تنویس. تنویس. خووجی: فرمت نام بیمار: تشخیص بیماری:		
الفقط و فقط فقط خروجى نهايى و در در قالب خروجى نهايى و در در قالب بده. هيچ بيده. هيچ توضيح يا تفسير اضافى توضيح يا ننويس. تفييس فرمت فرمت خروجى: و تشخيص نام بيمار: تشخيص اسم دارو:		
فقط و فقط خروجی نهایی رو در قالب زیر نمایش بده. هیچ توضیح یا تفسیر اضافی تفسیر اضافی ننویس. خروجی: فرمت نام بیمار: تشخیص بیماری: بیماری:		
خروجى نهايى رو در قالب زير نمايش بده. هيچ توضيح يا تنسير اضافى تنويس. فرمت فرمت نام بيمار: تشخيص بيمارى:		«نامشخص».
خروجى نهايى رو در قالب زير نمايش بده. هيچ توضيح يا تنسير اضافى تنويس. فرمت فرمت نام بيمار: تشخيص بيمارى:		فقط و فقط
رو در قالب زیر نمایش بده. هیچ توضیح یا تفسیر اضافی ننویس. فرمت فرمت خروجی: نام بیمار: تشخیص بیماری:		
زير نمايش بلده. هيچ توضيح يا تنويس. ننويس. غروجي: غرمت نام بيمار: تشخيص بيماري:		
بده. هیچ توضیح یا تفسیر اضافی ننویس. فرمت خروجی: نام بیمار: تشخیص بیماری:		
توضيح يا تفسير اضافى ننويس. فرمت فرمت خرو جى: نام بيمار: تشخيص بيمارى:		
تفسير اضافى ننويس. فرمت خروجى: نام بيمار: تشخيص بيمارى: اسم دارو:		
ننویس. فرمت خروجی: نام بیمار: تشخیص بیماری: بیمارو:		
خروجى: نام بيمار: تشخيص بيمارى: 		
خروجى: نام بيمار: تشخيص بيمارى: 		
نام بيمار: تشخيص بيمارى: بيمارى: اسم دارو:		
تشخیص بیماری: ——— اسم دارو:		
بيمارى: اسم دارو:		
اسم دارو:		
		ا بیماری:
		اسم دارو:
دوز:		ر دوز: ا دوز:

					نحوه مصرف:
					توصيه:
					نوع:
					(تکرار برای
					هر دارو)
universitytehran/PersianMind-v1.0	gpu	13.63	7.1	20%	تو یک دستیار
					پزشكى
					هوشمند
					هستى.
					نسخەھايى كە
					بهت داده
					مىشن ممكنه
					شامل غلطهاي
					تايپى، نگارشى،
					گرامري،
					ترتيب كلمات،
					دوزهای اشتباه
					یا حتی اشتباه
					در نوشتن نام
					داروها باشن.
					هدف تو اینه
					که با درک
					دقیق معنای
					نسخه،
					اطلاعات
					ساختاريافته
					زیر رو
					تصحيح شده و
					استاندارد
					استخراج كني.
					اگر هر بخش
					از اطلاعات
					(مثلاً دوز،
					توصيه، نوع
					دارو یا نام
					بيمار) مشخص
					نبود يا ناقص
					بود، فقط

					بنويس
					"نامشخص".
					فقط و فقط
					خروجی را در
					قالب زير ارائه
					بده. هيچ
					توضيح يا
					تفسير يا دليل
					اضافه ننويس.
					فرمت
					خروجي:
					نام بيمار:
					تشخيص
					يمارى:
					اسم دارو:
					دوز:
					نحوه مصرف:
					 توصيه:
					نوع:
					(تکرار برای
1.100.100				-0.4	هر دارو)
mistralai/Mistral-7B-Instruct-v0.2	gpu	14.48	5.2	5%	شما یک
					دستيار پزشكى
					هستيد. لطفاً
					فقط اطلاعات
					زير را
					استخراج کن و
					هیچ توضیح
					اضافی نده:
					- نام بیمار
					هم بیمهر - داروها و
					دوز مصرفی ۱ ت
					– طريقه
					مصرف
					- تشخيص
					بیماری

					. 1 ::
					فقط همین اطلاعات را
					بكركات را
mistralai/Mistral-7B-Instruct-v0.2	gpu	14.48	12.4	5%	تو یک دستیار
	81	_		-	ر ت پزشکی هستی.
					اطلاعات
					نسخه زير را
					دقيقاً با فرمت
					مشخص شده
					استخراج كن.
					فقط خروجى
					درخواستشده
					را ارائه بده.
					اگر اطلاعات
					وجود نداشت،
					بنويس
					"نامشخص".
					### فرمت
					ااااا ورانت خروجي:
					نام بيمار:
					تشخيص
					يىمارى:
					اسم دارو:
					دوز:
					نحوه مصرف:
					توصيه:
					نوع:
					(تکرار برای
1.1.20		1.4.40	10	50 /	هر دارو)
mistralai/Mistral-7B-Instruct-v0.2	gpu	14.48	12	5%	از متن نسخه،
					فقط این
					اطلاعات را
					فقط این اطلاعات را استخراج کن: ۱. نام بیمار ۲. تشخیص
					۱. نام بیمار
					۲. تشخیص
					بيماري

		٣. ليست
		داروها شامل:
		- نام دارو
		- دوز (مثلاً:
		۵۰۰ میلی گرم،
		۱ عدد، نصف
		قرص)
		- نحوه
		مصرف (مثلاً:
		روزی دوبار
		بعد غذا)
		- توصيه
		پزشک (مثلاً: با
		آب زیاد، در
		صورت تهوع
		قطع شود)
		– نوع دارو
		(مثلاً: قرص،
		شربت،
		تزریقی)
		ti:
		قالب دقيق
		خروجي:
		نام بيمار:
		مثال: على
		رضای <i>ی</i>]
		تشخيص
		بيمارى: [مثال:
		عفونت گلو]
		دارو ۱:
		نام: [مثال:
		آموكسىسيلين]
		دوز: [مثال:
		۵۰۰ میلی گرم]
		نحوه مصرف:
		[مثال: هر ۸
		ساعت يكبار
		بعد غذا]

محمدعلى لطفى كوشالي

	ı		l	l	1
					توصيه: [مثال:
					تا ۷ روز
					مصرف شود]
					نوع: [مثال:
					كپسول]
					(اگر داروهای
					بیشتری هست،
					ادامه بده به
					همین شکل)
					فقط اطلاعات
					خواستهشده را
					بده. توضيح يا
					تفسير ننويس.
mistralai/Mistral-7B-Instruct-v0.2	gpu	14.48	10.2	5%	تو یک دستیار
					پزشكى
					هوشمند
					هستى.
					نسخەھايى كە
					بهت میدیم
					ممكنه شامل
					غلطهای تایپی،
					نگارشی،
					گرامری یا
					اشتباهات دوز
					و نحوه مصرف
					باشن. وظيفه
					تو اینه که با
					درک دقیق متن
					نسخه،
					اطلاعات زير
					رو اصلاحشده
					و استاندارد
					طبق قالب
					خواستهشده
					استخراج كني.
					اگر اطلاعات
					ناقص يا
					نامشخص بود،
	1		l	I	

					فقط بنويس
					«نامشخص».
					فقط و فقط
					خروجی نهایی
					رو در قالب
					زیر نمایش
					بده. هیچ
					توضيح يا
					تفسير اضافي
					ننويس.
					فرمت
					خروجى:
					نام بيمار:
					تشخيص
					بیماری:
					اسم دارو:
					، دوز:
					نحوه مصرف:
					توصيه:
					نوع:
					(تکرار برای
					هر دارو)
					,{"""
mistralai/Mistral-7B-Instruct-v0.2	gpu	14.48	12.1	5%	تو یک دستیار
					پزشكى
					هوشمند
					هستى.
					نسخەھاي <i>ى</i> كە
					بهت داده
					مىشن ممكنه
					شامل غلطهای
					تايپى، نگارشى،
					یپی و ی گرامری،
					ترتیب کلمات،
					و یا . دوزهای اشتباه
					یا حتی اشتباه
	l				

محمدعلى لطفى كوشالى	گزارش جامع پروژه پایانی deep learning (متن)
	در نوشتن نام
	داروها باشن.
	هدف تو اینه
	که با درک
	دقیق معنای
	انسخه،
	اطلاعات
	ساختاريافته
	ا ذير رو
	تصحیحشده و
	استاندارد
	استخراج کنی.
	اگر هر بخش
	از اطلاعات
	(مثلاً دوز،
	توصيه، نوع
	دارو یا نام
	بيمار) مشخص
	نبود یا ناقص
	بود، فقط
	بنویس
	"نامشخص".
	فقط و فقط
	خروجی را در
	قالب زير ارائه
	بده. هیچ
	· ع توضیح یا
	تفسير يا دليل
	اضافه ننویس.
	ا فرمت
	خروجي:
	ا نام بیمار:
	تشخیص
	بيمارى:
	اسم دارو:
	ا دوز:

					نحوه مصرف:
					توصيه:
					نوع:
					(تکرار برای
					هر دارو)
mistralai/Mistral-7B-Instruct-v0.2	gpu	14.48	9.1	5%	تو یک دستیار
					پزشكى
					هوشمند
					هستى.
					نسخەھايى كە
					بهت داده
					مىشن ممكنه
					شامل غلطهاي
					تايپى، نگارشى،
					گرامري،
					دوزهای اشتباه
					يا فواصل
					مصرف
					غيرمنطقي
					باشن.
					نکات مهمی
					که باید اصلاح
					كنى:
					the first of
					دوز دارو باید
					واقع بينانه و
					رايج باشه. مثلاً
					«mgo···»
					برای یک
					قرص
					استامينوفن
					غيرواقعي
					هست و باید
					به مقدار معقول
					مثل «mg٥٠٠»
					اصلاح بشه.

محمدعلى لطفى كوشالى	de (متن)	ep learning	پروژه پایانی ;	گزارش جامع
				نحوه مصرف
				باید در بازه
				زمانى معقول
				باشه. مثلاً اگر
				گفته شده «هر
				۱۰۰ ساعت»،
				اصلاحش كن
				به چیزی مثل
				«هر ٦ ساعت»
				يا بنويس
				«نامشخص»
				اگر قابل
				تشخيص نبود.
				هدف تو اینه
				هدف نو اینه که با درک
				دقیق متن نسخه،
				اطلاعات زير
				رو اصلاح شده
				و استاندارد
				استخراج كني.
				اگر هر بخش
				از اطلاعات
				(مثلاً دوز،
				توصيه، نوع
				دارو یا نام
				بيمار) مشخص
				نبود یا غیرقابل
				تشخیص بود،
				فقط بنويس
				"نامشخص".
				فقط و فقط
				خروجی را در
				قالب زير ارائه
				بده. هیچ
				توضيح يا

كو شالى	لطف	محمدعلي
ے سے	صعي	ے جہاں جہاں

سير يا دليل مافه ننويس. رمت بروجي: م بيمار:
رمت تروجى: م بيمار:
م بيمار: م بيمار:
م بيمار: م بيمار:
م بيمار:
شخص
مارى:
سم دارو:
وز:
حوه مصرف:
رصيه:
ے:
ت نکرار برای
ر 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-	gpu	16.07	7	97%	تو یک دستیار
Llama3.1-8B- Instruct					پزشكى هستى.
					اطلاعات نسخه زير
					را دقيقاً با فرمت
					مشخص شده
					استخراج كن. فقط
					خروجي
					درخواستشده را
					ارائه بده. اگر
					اطلاعات وجود

					نداشت، بنويس
					"نامشخص".
					هيچ توضيح يا
					تفسير اضافي
					ننويس.
					### فرمت
					خروجي:
					نام بيمار:
					تشخیص بیماری:
					اسم دارو:
					دوز:
					نحوه مصرف و
					توصيه:
					نوع:
					(تکرار برای هر
					دارو)
PartAI/Dorna2-	gpu	16.07	11.5	70%	تو یک دستیار
Llama3.1-8B- Instruct					پزشكى هستى.
mstruct					اطلاعات نسخه زير
					را دقيقاً با فرمت
					مشخصشده
					استخراج كن.
					رج ں۔ فقط خروجی در
					قالب زير ارائه بده
					. ویر و . و هیچ توضیح یا
					تفسير اضافه
					ننويس.
					اگر اطلاعات وجود
					نداشت، بنویس
					«نامشخص».
					توصیههای کلی
					ر یا کی پزشک که مستقیماً
					به نحوه مصرف به نحوه مصرف
	1	l			
					دارو مربوط نيستند
					دارو مربوط نیستند را در بخش

	T	T			
					«توصیههای کلی
					پزشک» بنویس.
					فرمت خروجي:
					نام بيمار:
					تشخیص بیماری:
					 اسم دارو:
					دوز:
					نحوه مصرف و
					توصيه (فقط مربوط
					به همان دارو):
					نوع:
					تکرار برای هر
					دارو)
					توصیههای کلی
					پزشک:
PartAI/Dorna2-	gpu	16.07	8.3	60%	تو یک دستیار
Llama3.1-8B- Instruct					پزشكى هستى.
mstruct					اطلاعات نسخه زير
					را دقيقاً با فرمت
					مشخص شده
					استخراج كن. فقط
					خروجى
					درخواستشده را
					ارائه بده. اگر
					اطلاعات وجود
					نداشت، بنویس
					"نامشخص".
					نامستحص .
					هیچ توضیح یا
					تفسير اضافي
					ننويس.
					### فرمت
					خروجي:
					نام بيمار:
					تشخیص بیماری:

كوشالي	لطفي	محمدعلي
--------	------	---------

ورورت وجود وجود وجود المنت داخل المنات داخل المنات داخل المنات داخل المنات داخل المنات داخل المنات					
السه داود اسم داود اسم داود و دوزه اسم داود و دوزه الموسية					توصیه اضافی: (در
اسم دارو: اسم دارو: اسم دارو: الاخواء بعدوه مصرف و دون الاخواء براى هر نوع الاخواء براى هر نوع الاخواء براى هر نوع الاخواء بناه في الله بده. الاخواء الله الله الله بده. الاخواء الله بده. الله بده. الاخواء الله بده. الله بده. الله بده. الاخواء الله بده.					
المجادة و المورق و المجادة و المجاد					نسخه)
المجادة و المورق و المجادة و المجاد					 اسم دارو:
المنافلة ال					·
المجادل براى هر المجادل المجا					نحوه مصرف و
المتحال المراك هر المتحال الم					توصيه:
الم المراقب ا					نوع:
PartAl/Dorna2- Llama3.1-8B- Instruct gpu 16.07 6.8 95% 95% برشكي هستي. برشكي هستي. اطلاعات نسخه زير وجوبي المنافق ال					(تکرار برای هر
الملاعات نسخه زیر پزشکی هستی. اطلاعات نسخه زیر پر پزشکی الملاعات نسخه زیر پر پزشکی المشخص شده و را دقیقاً با فرمت استخراج کن دققط مشخص شده و را دوره بی استخراج کن دققط الرائه بده. اگر و برود الرائه بده. اگر الناشت، بنویس اطلاعات و جود الناشت، بنویس تفسیر اضافی هیچ توضیح یا ننویس. تفسیر اضافی موردی در فرمت ننویس. اشافه نکن و فقط خروجی از خودت اطلاعاتی ک در اطلاعاتی ک در اطلاعاتی ک در در ادره موردی در در وردی است را در موردی					دارو)
اطلاعات نسخه زیر اولتها با فرمت را دقیقاً با فرمت را دقیقاً با فرمت استخواج کن. فقط مشخص شده را خروجی استخدام را خروجی درخواست شده را ازائه بده. اگر درخواست شده را اطلاعات وجود ازائه بده. اگر انشت، بنویس اطلاعات وجود انفیس. تفسیر اضافی هیچ توضیع یا تنویس. تفسیر اضافی موردی در فرمت ننویس. اضافی کن و فقط خروجی از خودت اضافه نکن و فقط اطلاعاتی ک در اضافه نکن و مودد	gpu	16.07	6.8	95%	تو یک دستیار
اطلاعات نسخه زیر را دقیقاً با فرمت استخراج کن، فقط عروجی درخواستشده را عروجی ارائه بده. اگر ارائه بده. اگر اطلاعات وجود ارائه بده و اگر انامشخص". انامشخص". تفسیر اضافی هیچ توضیع یا تنویس. تفسیر اضافی موردی در فرمت ننویس. اضافه نکن و فقط اطلاعاتی ک در اطلاعاتی ک در اسخه ها موجود					پزشكى هستى.
مشخص شده استخراج كن. فقط درخواست شده را درخواست شده را اله بده. اگر ارائه بده. اگر اطلاعات وجود اطلاعات وجود الداشت، بنویس المشخص ". المشخص ". تفسیر اضافی هیچ توضیح یا تفسیر اضافی موردی در فرمت ننویس. اضافی خروجی از خودت اضافه نكن و فقط خروجی از خودت اطلاعاتی ک در اطلاعاتی ک در اطلاعاتی ک در است را در مواردی					اطلاعات نسخه زير
استخراج كن. فقط درخواستشده را خروجي ارائه بده. اگر ارائه بده. اگر اطلاعات وجود اطلاعات وجود الناشت، بنویس اظاهات وجود الناشت، بنویس تفسیر اضافی هیچ توضیح یا تفسیر اضافی موردی در فرمت ننویس. عزوجی از خودت موردی در فقط خروجی از خودت اطلاعاتی ک در اضافه نکن و فقط است را در مواردی					را دقيقاً با فرمت
خووجى درخواستشده را درخواستشده را ارائه بده. اگر اطلاعات وجود الشت، بنویس انامشخص". انامشخص". تفسیر اضافی هیچ توضیح یا تنویس. تنویس. موردی در فرمت ننویس. خروجی از خودت ما اضافه نکن و فقط اضافه نکن و فقط اطلاعاتی ک در انسخه ها موجود					مشخص شده
درخواستشده را اگر ارائه بده. اگر اطلاعات وجود الله بده. اگر انداشت، بنویس الله بنویس الله بنویس الله بنویس الله بنویس الله بنویس الله بنویس تفسیر اضافی هیچ توضیح یا ننویس. اننویس. موردی در فرمت خروجی از خودت اطافه نکن و فقط خروجی از موادت اطلاعاتی ک در اطلاعاتی ک در است را در مواردی					استخراج كن. فقط
ارائه بده. اگر اطلاعات وجود "نامشخص". "نامشخص". تفسیر اضافی هیچ توضیح یا تنویس. ننویس. موردی در فرمت خروجی از خودت اضافه نکن و فقط اطلاعاتی ک در اسخه ها موجود است را در مواردی					خروجى
اطلاعات وجود "نامشخص". هیچ توضیح یا تفسیر اضافی ننویس. ننویس. موردی در فرمت خروجی از خودت اضافه نکن و فقط اطلاعاتی ک در اساخه ها موجود است را در مواردی					درخواستشده را
نداشت، بنویس "نامشخص" هیچ توضیح یا تفسیر اضافی ننویس. ننویس. موردی در فرمت خروجی از خودت اضافه نکن و فقط نطلاعاتی ک در اطلاعاتی ک در است دا در مواردی					ارائه بده. اگر
"نامشخص". هیچ توضیح یا تفسیر اضافی ننویس. موردی در فرمت خروجی از خودت اضافه نکن و فقط اطلاعاتی ک در اسخه ها موجود					اطلاعات وجود
هیچ توضیح یا تفسیر اضافی ننویس. موردی در فرمت خروجی از خودت اضافه نکن و فقط اطلاعاتی ک در نسخه ها موجود					
تفسير اضافي ننويس. موردي در فرمت خروجي از خودت اضافه نكن و فقط اطلاعاتي ك در نسخه ها موجود					"نامشخص".
تفسير اضافي ننويس. موردي در فرمت خروجي از خودت اضافه نكن و فقط اطلاعاتي ك در نسخه ها موجود					ھىچ توضىح يا
ننویس. موردی در فرمت خروجی از خودت اضافه نکن و فقط اطلاعاتی ک در نسخه ها موجود					
موردی در فرمت خروجی از خودت اضافه نکن و فقط اطلاعاتی ک در نسخه ها موجود است را در مواردی					
خروجى از خودت اضافه نكن و فقط اطلاعاتى ك در نسخه ها موجود است را در مواردى					
اضافه نکن و فقط اطلاعاتی ک در نسخه ها موجود است را در مواردی					
اطلاعاتی ک در نسخه ها موجود است را در مواردی					
نسخه ها موجود است را در مواردی					_
					نسخه ها موجود
که مه حه د است،					است را در مواردی
					که موجود است،
بنويس.					بنويس.
### فرمت					### فرمت
خروجي:					خروجي:
نام بیمار:					نام بيمار:

محمدعلي لطفي كوشالي

	I				
					تشخیص بیماری:
					اسم دارو:
					دوز:
					نحوه مصرف و
					توصيه:
					نوع:
					(تکرار برای هر
					دارو)
PartAI/Dorna2-	gpu	16.07	7	80%	تو یک دستیار
Llama3.1-8B- Instruct					پزشك <i>ى</i> هستى.
THIS II GOL					اطلاعات نسخه زير
					را با فرمت
					مشخص شده
					استخراج كن. فقط
					خروجي
					روبی درخواستشده را
					ارائه بده. اگر
					اطلاعات وجود
					نداشت، بنویس
					"نامشخص".
					هیچ توضیح یا
					تفسير اضافي
					ننويس.
					عويس.
					### فرمت
					خروجي:
					نام بيمار:
					تشخیص بیماری:
					اسم دارو:
					دوز:
					نحوه مصرف و
					توصيه:
					و نوع:
					تکرار برای هر
					دارو) دارو)
					دارو

PartAI/Dorna2-	gpu	16.07	6.9	96%	تو یک دستیار
Llama3.1-8B-	8r ··			20.2	ر پزشکی هستی.
Instruct					اطلاعات نسخه زير
					را دقيقاً با فرمت
					مشخص شده
					استخراج كن. فقط
					خروجي
					درخواستشده را
					ارائه بده. اگر
					اطلاعات وجود
					نداشت، بنويس
					"نامشخص".
					هیچ توضیح یا
					تفسير اضافي
					ننويس.
					موردی در فرمت
					خروجی از خودت
					اضافه نكن.
					### فرمت
					ر خروجي:
					نام بيمار:
					۰۰۰۰ تشخیص بیماری:
					اسم دارو:
					دوز:
					نحوه مصرف و
					توصيه:
					نوع:
					(تکرار برای هر
					دارو)
PartAI/Dorna2- Llama3.1-8B-	gpu	16.07	7.1	85%	تو یک دستیار
Instruct					پزشكى هستى.
					اطلاعات نسخه زير
					را دقيقاً با فرمت
					مشخص شده
					استخراج كن.

					فقط خروجي
					درخواستشده را
					ارائه بده. اگر
					اطلاعات وجود
					نداشت، بنویس
					«نامشخص».
					هیچ توضیح یا
					ین حرصیں یہ تفسیر اضافی
					ننويس.
					ری ن هیچ موردی خارج
					یچ رو ن ربی از ساختار و ترتیب
					فیلدهای فرمت
					ی ص ر ــــــــــــــــــــــــــــــــــ
					نکن.
					ت توصیههای کلی
					پزشک که به
					پر داروی خاصی
					مربوط نیستند را
					ربو نادیده بگیر و در
					خروجي ننويس.
					<i>C</i> .5 <i>G</i> . 35
					فرمت خروج <i>ی</i> :
					نام بيمار:
					تشخیص بیماری:
					اسم دارو:
					دوز:
					نحوه مصرف و
					توصيه:
					نوع:
					(تکرار برای هر
D (A1/D 2		160=	7 .	0.607	دارو)
PartAI/Dorna2- Llama3.1-8B-	gpu	16.07	7.1	86%	تو یک دستیار
Instruct					پزشكى هستى.
					اطلاعات نسخه زير
					را دقيقاً با فرمت
					مشخص شده
					استخراج كن. فقط
					خروجي

					درخواستشده را ارائه بده. اگر اطلاعات وجود نداشت، بنویس
					"نامشخص".
					### فرمت
					خروجى:
					نام بیمار: تشخیص بیماری:
					اسم دارو:
					دوز:
					نحوه مصرف و
					توصيه:
					نوع: (تکرار برای هر
					رفاکرار برای هر دارو)
PartAI/Dorna2-	gpu	16.07	7.4	75%	تو یک دستیار
Llama3.1-8B-	81, 11	10107	,	, ,	ر . پزشکی هستی.
Instruct					اطلاعات نسخه زير
					را دقيقاً با فرمت
					مشخص شده
					استخراج كن. فقط
					خروجي
					درخواستشده را
					ارائه بده. اگر
					اطلاعات وجود
					نداشت، بنویس "نامشد بنویس
					"نامشخص".
					### فرمت
					خروجي:
					نام بيمار:
					تشخیص بیماری:
					اسم دارو:
					دوز:
					الگوي مصرف:

		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 . مدل درنای مدل درنای وگل 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 خروجي:

نام بيمار: ...

تشخیص بیماری: ...

اسم دارو: ...

دوز: ...

نحوه مصرف و توصیه: ...

نوع: ...

(تکرار برای هر دارو)

محمدعلى لطفى كوشالي

گزارش جامع پروژه پایانی deep learning (متن)

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

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

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

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

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