как мислиш валиден ли е или да променя gradio

huggingface\_hub

Copilot said: Във файла cyber-chat/requirements..txt в момента са посочени

Във файла cyber-chat/requirements..txt в момента са посочени следните пакети:

Code

gradio

requests

Ако питаш дали трябва да смениш gradio с huggingface\_hub, зависи какво използва проектът ти:

Ако използваш функционалности на Gradio (за уеб интерфейс на ML модели), остави gradio.

Ако използваш huggingface\_hub (за теглене/качване на модели от Hugging Face Hub), добави huggingface\_hub, но не е нужно да махаш gradio, освен ако вече не го ползваш.

gradio и huggingface\_hub са различни библиотеки и не са взаимозаменяеми. Можеш да имаш и двете в requirements файла, ако ти трябват.

Пример:

Code

gradio

huggingface\_hub

requests

Ако не ти трябва Gradio, можеш да го махнеш. Ако не си сигурен(на), кажи каква е целта на проекта/коя библиотека реално използваш в кода и ще ти помогна по-конкретно!

You said: дай да направя комит промених файла както

дай да направя комит промених файла както каза

Copilot said: Страхотно! За да направиш commit след промяна

Страхотно! За да направиш commit след промяна на файла requirements..txt, изпълни следните стъпки във вашата локална Git среда:

Добави файла към stage:

bash

git add cyber-chat/requirements..txt

Направи commit с подходящо съобщение:

bash

git commit -m "Обновен requirements..txt: добавен huggingface\_hub"

Изпрати промените към GitHub:

bash

git push

Ако имаш нужда от помощ с нещо по-конкретно (например работа с GitHub Web интерфейс или друга команда), кажи!

You said: gradio huggingface\_hub requests ===== Application Startup at 2025-06-25

gradio

huggingface\_hub

requests ===== Application Startup at 2025-06-25 10:45:54 =====

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Trying to transpile functions from Python -> JS for performance

\* (1/1) <lambda>: ✅

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Running on local URL: http://0.0.0.0:7860, with SSR ⚡ (experimental, to disable set `ssr\_mode=False` in `launch()`)

\* To create a public link, set `share=True` in `launch()`.

Traceback (most recent call last):

File "/usr/local/lib/python3.10/site-packages/huggingface\_hub/utils/\_http.py", line 409, in hf\_raise\_for\_status

response.raise\_for\_status()

File "/usr/local/lib/python3.10/site-packages/requests/models.py", line 1026, in raise\_for\_status

raise HTTPError(http\_error\_msg, response=self)

requests.exceptions.HTTPError: 402 Client Error: Payment Required for url: https://router.huggingface.co/hf-inference/models/HuggingFaceH4/zephyr-7b-beta/v1/chat/completions

The above exception was the direct cause of the following exception:

Traceback (most recent call last):

File "/usr/local/lib/python3.10/site-packages/gradio/queueing.py", line 625, in process\_events

response = await route\_utils.call\_process\_api(

File "/usr/local/lib/python3.10/site-packages/gradio/route\_utils.py", line 322, in call\_process\_api

output = await app.get\_blocks().process\_api(

File "/usr/local/lib/python3.10/site-packages/gradio/blocks.py", line 2218, in process\_api

result = await self.call\_function(

File "/usr/local/lib/python3.10/site-packages/gradio/blocks.py", line 1729, in call\_function

prediction = await anyio.to\_thread.run\_sync( # type: ignore

File "/usr/local/lib/python3.10/site-packages/anyio/to\_thread.py", line 56, in run\_sync

return await get\_async\_backend().run\_sync\_in\_worker\_thread(

File "/usr/local/lib/python3.10/site-packages/anyio/\_backends/\_asyncio.py", line 2470, in run\_sync\_in\_worker\_thread

return await future

File "/usr/local/lib/python3.10/site-packages/anyio/\_backends/\_asyncio.py", line 967, in run

result = context.run(func, \*args)

File "/usr/local/lib/python3.10/site-packages/gradio/utils.py", line 894, in wrapper

response = f(\*args, \*\*kwargs)

File "/home/user/app/app.py", line 11, in chat\_fn

completion = client.chat.completions.create(

File "/usr/local/lib/python3.10/site-packages/huggingface\_hub/inference/\_client.py", line 924, in chat\_completion

data = self.\_inner\_post(request\_parameters, stream=stream)

File "/usr/local/lib/python3.10/site-packages/huggingface\_hub/inference/\_client.py", line 280, in \_inner\_post

hf\_raise\_for\_status(response)

File "/usr/local/lib/python3.10/site-packages/huggingface\_hub/utils/\_http.py", line 482, in hf\_raise\_for\_status

raise \_format(HfHubHTTPError, str(e), response) from e

huggingface\_hub.errors.HfHubHTTPError: 402 Client Error: Payment Required for url: https://router.huggingface.co/hf-inference/models/HuggingFaceH4/zephyr-7b-beta/v1/chat/completions (Request ID: Root=1-685bd392-6f216f6366f7f90c484eef61;cd79dabf-9213-48db-8d1d-2e7bec59d60d)

You have exceeded your monthly included credits for Inference Providers. Subscribe to PRO to get 20x more monthly included credits. не раобти

Copilot said: Грешката, която получаваш: ``` huggingface\_hub.errors.HfHubHTTPError: 402 Client Error: Payment

Грешката, която получаваш:

Code

huggingface\_hub.errors.HfHubHTTPError: 402 Client Error: Payment Required for url: https://router.huggingface.co/hf-inference/models/HuggingFaceH4/zephyr-7b-beta/v1/chat/completions

You have exceeded your monthly included credits for Inference Providers. Subscribe to PRO to get 20x more monthly included credits.

означава, че си надхвърлил безплатния лимит за inference на Hugging Face Hub за този месец.

Какво да направиш:

Имаш два избора:

Изчакай до следващия месец, когато безплатните кредити ще се презаредят.

Абонирай се за Hugging Face PRO, за да получиш повече inference кредити (платено).

Алтернативи:

Можеш да използваш друг модел, който работи локално и не изисква inference през Hugging Face API.

Ако имаш собствен GPU/сървър, можеш да хостваш модела локално и да извикаш inference без нужда от Hugging Face Hub API.

Кодът и requirements.txt са ти наред – проблемът не е в тях, а в лимита на Hugging Face API.

Решение:

Ако искаш да продължиш с Hugging Face Hub този месец, трябва да минеш на платен план.

Ако не искаш да плащаш, потърси модел, който може да се стартира локално (без API заявки към huggingface.co).

Ако искаш конкретна помощ за локално стартиране на модел или смяна на модела, кажи!

You said: Можеш да използваш друг модел, който работи

Можеш да използваш друг модел, който работи локално и не изисква извод чрез API на Hugging Face. как

Copilot said: Много добър въпрос! Ето как може да

Много добър въпрос! Ето как може да използваш локален модел без нужда от Hugging Face API (без онлайн inference и кредитни лимити):

1. Свали модела локално

Отиди на страницата на желания модел в Hugging Face и го изтегли на машината си. Например, използвай командата:

bash

huggingface-cli download <owner>/<model-name>

или просто ръчно свали цялата папка на модела.

2. Зареди модела с Transformers (пример за текстов модел)

Python

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_name\_or\_path = "./път/към/модела" # например: "./zephyr-7b-beta"

tokenizer = AutoTokenizer.from\_pretrained(model\_name\_or\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_name\_or\_path)

Важно:

Някои големи модели изискват GPU.

Нужно е да си инсталирал transformers и torch (pip install transformers torch).

3. Използвай модела локално

Python

inputs = tokenizer("Твоят въпрос тук", return\_tensors="pt")

outputs = model.generate(\*\*inputs)

result = tokenizer.decode(outputs[0])

print(result)

4. Ако използваш Gradio

Промени функцията, така че да използва локалния модел:

Python

def my\_chat\_function(message):

inputs = tokenizer(message, return\_tensors="pt")

outputs = model.generate(\*\*inputs, max\_new\_tokens=100)

return tokenizer.decode(outputs[0])

5. Пример за Gradio интерфейс

Python

import gradio as gr

iface = gr.Interface(fn=my\_chat\_function, inputs="text", outputs="text")

iface.launch()

Накратко:

Свали модела локално

Зареди го с Transformers (или друга библиотека)

Използвай го в кода си, вместо да викаш API на Hugging Face

Ако кажеш кой точно модел искаш да ползваш, мога да ти дам конкретен пример!

Uploaded image: image.png

You said: Describe this image

Describe this image

Copilot said: На изображението се вижда терминален прозорец (Git

На изображението се вижда терминален прозорец (Git Bash), в който потребителят изпълнява няколко команди:

Успешен push към GitHub репозитория stellababy2004/stella-projects-final. Git съобщава, че промяната (добавен huggingface\_hub в requirements..txt) е качена успешно.

Опит за изпълнение на команда:

Code

huggingface-cli download <owner>/<model-name>

Тази команда връща грешка:

Code

bash: syntax error near unexpected token 'newline'

Причината е, че <owner>/<model-name> е примерен плейсхолдър, а не реални стойности. Bash го интерпретира некоректно.

В обобщение: Git push е успешен, но командата за изтегляне на модел от Hugging Face не е изпълнена правилно заради използване на примерни плейсхолдъри вместо истински имена.

Uploaded image: image.png

You said: така ства ли

така ства ли

Copilot said: Да, така е напълно коректно! Това

Да, така е напълно коректно!

Това е правилният начин да клонираш модел от Hugging Face локално, особено големи модели, използвайки git lfs.

Стъпките от изображението:

Инсталирай git lfs (ако не си):

bash

git lfs install

Клонирай модела (ще изтегли и големите файлове):

bash

git clone https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3

При нужда ще поиска Hugging Face access token (можеш да го създадеш от линка на изображението).

Алтернатива: Ако не искаш да теглиш големите файлове веднага, а само "пойнтърите":

bash

GIT\_LFS\_SKIP\_SMUDGE=1 git clone https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3

После можеш да изтеглиш конкретно само нужните файлове.

В обобщение:

Това е правилният и препоръчителен начин за локално теглене на модели от Hugging Face, които са големи и използват git lfs. След като клонираш модела, можеш да го заредиш с transformers или друга библиотека ЛОКАЛНО, без да използваш API кредити!

Ако имаш нужда от примерен код за зареждане на такъв модел, пиши!

You said: Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final (main) $ git lfs

Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final (main)

$ git lfs install

Updated Git hooks.

Git LFS initialized.

Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final (main)

$ git clone https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3

Cloning into 'Mistral-7B-Instruct-v0.3'...

remote: Password authentication in git is no longer supported. You must use a user access token or an SSH key instead. See https://huggingface.co/blog/password-git-deprecation

fatal: Authentication failed for 'https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3/'

Copilot said: Грешката, която виждаш: ``` remote: Password authentication in git

Грешката, която виждаш:

Code

remote: Password authentication in git is no longer supported. You must use a user access token or an SSH key instead. See https://huggingface.co/blog/password-git-deprecation

fatal: Authentication failed for 'https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3/'

означава, че вече не можеш да клонираш с парола, а трябва да използваш Hugging Face user access token или SSH ключ.

Решение: Клониране с Hugging Face access token

Генерирай Hugging Face access token:

Отиди на: https://huggingface.co/settings/tokens

Натисни “New token”, избери role (Read) и копирай токена.

Клонирай модела:

Изпълни командата:

bash

git clone https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3

Когато поиска потребителско име, въведи:

Code

<твоето Hugging Face потребителско име>

Когато поиска парола, постави генерирания access token.

Ако искаш да не въвеждаш токена всеки път:

Можеш да го сложиш директно в URL-то (НЕ се препоръчва от гледна точка на сигурността):

bash

git clone https://<USERNAME>:<TOKEN>@huggingface.co/mistralai/Mistral-7B-Instruct-v0.3

Пример:

Code

git clone https://stellababy2004:hf\_xxxxxxxx@huggingface.co/mistralai/Mistral-7B-Instruct-v0.3

Алтернатива: Използвай SSH ключ

Ако предпочиташ SSH, добави публичния си ключ в Hugging Face профила (Settings > SSH Keys).

Клонирай с SSH адреса (ако е наличен).

В обобщение:

Трябва да използваш Hugging Face access token като парола при клониране чрез HTTPS. Това ще реши проблема с authentication failed.

Ако имаш нужда от помощ при създаването на токена или въвеждането му, пиши!

You said: Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final (main) $ git clone

Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final (main)

$ git clone https://stellababy2004:hf\_niEGaaaVRYpJovyNFvGSWfbxESRYpWusna@huggingface.co/mistralai/Mistral-7B-Instruct-v0.3

Cloning into 'Mistral-7B-Instruct-v0.3'...

remote: Access to model mistralai/Mistral-7B-Instruct-v0.3 is restricted and you are not in the authorized list. Visit https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3 to ask for access.

fatal: unable to access 'https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3/': The requested URL returned error: 403

Copilot said: Виждаш грешка 403 с това съобщение: > Access

Виждаш грешка 403 с това съобщение:

Access to model mistralai/Mistral-7B-Instruct-v0.3 is restricted and you are not in the authorized list.

Това означава, че този модел в Hugging Face е ограничен (restricted access) и не всеки може да го изтегли.

Какво да направиш:

Посети страницата на модела:

https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3

Натисни бутона “Request Access” (или подобен)

Ще трябва да поискаш достъп чрез Hugging Face профила си.

Изчакай одобрение по имейл или в профила си.

Докато не получиш одобрение, няма да можеш да клонираш/теглиш този модел.

Алтернатива:

Използвай модел, който е със свободен достъп (public). Например:

google/gemma-2b

tiiuae/falcon-7b-instruct

mistralai/Mistral-7B-v0.1 (ако е с публичен достъп)

Пример за клониране на публичен модел:

bash

git clone https://huggingface.co/google/gemma-2b

Накратко:

Трябва да поискаш достъп или да избереш модел с отворен достъп. Ако искаш препоръка за добър публичен модел, кажи какъв ти трябва и ще ти дам линк!

You said: Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final (main) $ git clone

Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final (main)

$ git clone https://huggingface.co/google/gemma-2b

Cloning into 'gemma-2b'...

remote: Password authentication in git is no longer supported. You must use a user access token or an SSH key instead. See https://huggingface.co/blog/password-git-deprecation

fatal: Authentication failed for 'https://huggingface.co/google/gemma-2b/'

Copilot said: Тази грешка се появява, защото Hugging Face

Тази грешка се появява, защото Hugging Face вече не позволява клониране чрез парола. Трябва да използваш Hugging Face access token (както и при другия модел).

Как да клонираш модел с Hugging Face access token

Генерирай си Hugging Face access token:

Отиди на: https://huggingface.co/settings/tokens

Натисни “New token”, избери “Read” и копирай токена.

Клонирай модела:

В терминала изпълни:

bash

git clone https://<твоето-потребителско-име>:<твоят-токен>@huggingface.co/google/gemma-2b

Пример:

bash

git clone https://stellababy2004:hf\_niEGaaaVRYpJovyNFvGSWfbxESRYpWusna@huggingface.co/google/gemma-2b

Алтернатива:

Ако не искаш токенът ти да се вижда в историята, просто напиши:

bash

git clone https://huggingface.co/google/gemma-2b

и когато поиска потребителско име и парола, въведи:

Потребителско име: твоето Hugging Face потребителско име

Парола: access token-а, който си генерирал

Важно:

Без access token няма да можеш да клонираш модели от Hugging Face, дори и да са публични.

Токенът трябва да има поне Read права.

Кажи, ако имаш затруднения с някоя от стъпките!

You said: jdud :r git clone https://stellababy2004:hf\_ttNqZRXMoJLvtObBOThziJazwBOvytdiKI @huggingface.co/google/gemma-2b

jdud :r git clone https://stellababy2004:hf\_ttNqZRXMoJLvtObBOThziJazwBOvytdiKI @huggingface.co/google/gemma-2b

Copilot said: В твоята команда има малка грешка —

В твоята команда има малка грешка — има интервал преди „@huggingface.co“ и това ще доведе до грешка.

Трябва да е без интервал!

Правилен синтаксис:

bash

git clone https://stellababy2004:hf\_ttNqZRXMoJLvtObBOThziJazwBOvytdiKI@huggingface.co/google/gemma-2b

Важно:

Не споделяй публично своя Hugging Face Token (можеш да го ревокираш от settings след това).

Ако командата не работи, увери се, че токенът ти има поне “Read” права.

Ако имаш още въпроси — насреща съм!

You said: face.co/google/gemma-2b Cloning into 'gemma-2b'... remote: Access to model google/gemma-2b

face.co/google/gemma-2b

Cloning into 'gemma-2b'...

remote: Access to model google/gemma-2b is restricted and you are not in the authorized list. Visit https://huggingface.co/google/gemma-2b to ask for access.

fatal: unable to access 'https://huggingface.co/google/gemma-2b/': The requested URL returned error: 403

Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final (main)

hfjflf

Copilot said: Моделът \*\*google/gemma-2b\*\* също е с ограничен достъп

Моделът google/gemma-2b също е с ограничен достъп (restricted) и изисква предварително одобрение от Hugging Face, преди да можеш да го клонираш или изтеглиш. Затова виждаш грешка 403:

Code

Access to model google/gemma-2b is restricted and you are not in the authorized list.

Какво можеш да направиш:

Посети страницата на модела:

https://huggingface.co/google/gemma-2b

Натисни бутона "Request Access" (или подобен, като "Access repository" или "Request repository access").

Следвай инструкциите и изчакай одобрение.

Използвай модел с отворен достъп (public):

Ако не искаш да чакаш, можеш да избереш друг модел с публичен достъп. Някои предложения:

tiiuae/falcon-7b-instruct

mistralai/Mistral-7B-v0.1

TheBloke/Llama-2-7B-Chat-GGML (има и други варианти с отворен достъп)

Провери описанието на модела:

В Hugging Face винаги гледай дали пише "This repository is public" или "This repository is private/restricted".

Ако искаш, мога да ти извадя линкове към модели, които са достъпни без заявка за достъп. Кажи ако искаш такъв списък или препоръка!

Uploaded image: image.png

You said: hfjflfucoe

hfjflfucoe

Copilot said: На този екран виждаш страницата на Hugging

На този екран виждаш страницата на Hugging Face за модела google/gemma-2b.

Важно за достъп:

Моделът е публично достъпен, но трябва да приемеш условията за ползване на Google, за да получиш достъп до файловете и съдържанието му.

За да го направиш, трябва да си логнат и да натиснеш бутона „Acknowledge license“ (Потвърди лиценза).

След това ще можеш да клонираш или теглиш модела.

Стъпки:

Влез в профила си в Hugging Face.

Натисни бутона „Acknowledge license“ (вижда се в центъра на екрана).

След потвърждение опитай отново командата за клониране с твоя Hugging Face access token.

Резюме:

Трябва да приемеш условията (лиценза) на модела, преди да получиш достъп до файловете му. След това всичко ще работи!

Uploaded image: image.png

You said: къде д натисна

къде д натисна

Copilot said: Трябва да натиснеш големия сив бутон по

Трябва да натиснеш големия сив бутон по средата на екрана, където пише:

"Acknowledge license"

Този бутон се намира под текста:

“By agreeing you accept to share your contact information (email and username) with the repository authors.”

След като натиснеш този бутон и потвърдиш, ще получиш достъп до тегленето на модела. После ще можеш да клонираш без проблем!

Uploaded image: image.png

You said: къде да натисна

къде да натисна

Copilot said: Вече не е нужно да натискаш нищо

Вече не е нужно да натискаш нищо повече – всичко е готово!

На изображението виждаш съобщение “Gated model You have been granted access to this model”, което означава, че достъпът ти е одобрен.

Сега можеш спокойно да клонираш модела с командата:

bash

git clone https://huggingface.co/google/gemma-2b

(Използвай Hugging Face access token, когато бъдеш попитан за парола.)

Резюме:

Вече имаш достъп. Продължи с тегленето!

You said: коя п $ git clone https://huggingface.co/google/gemma-2b Cloning into 'gemma-2b'... remote:

коя п

$ git clone https://huggingface.co/google/gemma-2b

Cloning into 'gemma-2b'...

remote: Password authentication in git is no longer supported. You must use a user access token or an SSH key instead. See https://huggingface.co/blog/password-git-deprecation

fatal: Authentication failed for 'https://huggingface.co/google/gemma-2b/'

Copilot said: Тази грешка означава, че Hugging Face вече

Тази грешка означава, че Hugging Face вече не поддържа клониране чрез парола. Вместо това трябва да използваш Hugging Face access token или SSH ключ.

Коя парола да въведеш?

Не трябва да въвеждаш обикновената си парола!

Трябва да направиш следното:

Генерирай Hugging Face access token:

Отиди на: https://huggingface.co/settings/tokens

Натисни “New token”, избери “Read” и копирай токена.

Когато изпълниш:

Code

git clone https://huggingface.co/google/gemma-2b

Когато командата поиска:

Username: въведи Hugging Face потребителското си име (stellababy2004)

Password: постави копирания access token (не твоята парола!)

Може и директно така:

bash

git clone https://stellababy2004:ТОКЕНЪТ\_ТИ@huggingface.co/google/gemma-2b

Замени "ТОКЕНЪТ\_ТИ" със своя Hugging Face access token.

Резюме:

Паролата, която иска, е Hugging Face access token, а не обикновената ти парола!

You said: след като клонирах модела какво следва

след като клонирах модела какво следва

Copilot said: След като успешно клонира модела от Hugging

След като успешно клонира модела от Hugging Face, следващите стъпки зависят от това за какво искаш да го използваш. Ето най-честите варианти:

1. Зареждане и използване на модела в Python

Обикновено се използва библиотеката transformers или друга подходяща библиотека. Пример с transformers:

Python

from transformers import AutoModelForCausalLM, AutoTokenizer

# Път до локалната директория на модела (например "gemma-2b")

model\_path = "./gemma-2b"

# Зареждане на токенизатор и модел

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

# Примерно генериране на текст

inputs = tokenizer("Hello, how are you?", return\_tensors="pt")

outputs = model.generate(\*\*inputs)

print(tokenizer.decode(outputs[0], skip\_special\_tokens=True))

2. Използване на модела в друга среда

Ако ползваш друг фреймуърк (например llama.cpp, vLLM, Ollama и др.), виж в документацията как се зарежда модел от локална директория.

Ако искаш да го дообучаваш (файнтьонинг), използвай съответните скриптове на Hugging Face или друг инструмент.

3. Проверка на файловете

Увери се, че имаш нужните файлове (например: config.json, pytorch\_model.bin, tokenizer.model, tokenizer.json и т.н.)

Ако липсват големи файлове, може да се наложи да изпълниш:

bash

git lfs pull

в директорията на модела, за да изтеглиш всички LFS файлове.

4. Пример за стартиране на чат с модела (transformers + TextGenerationPipeline):

Python

from transformers import pipeline

model\_path = "./gemma-2b"

pipe = pipeline("text-generation", model=model\_path)

print(pipe("What is the capital of France?", max\_new\_tokens=20))

Ако имаш конкретна цел (пример: чатбот, анализ, fine-tuning), кажи я, за да ти дам по-точни инструкции!

You said: Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final (main) $ git clone

Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final (main)

$ git clone https://huggingface.co/google/gemma-2b

Cloning into 'gemma-2b'...

remote: Enumerating objects: 120, done.

remote: Total 120 (delta 0), reused 0 (delta 0), pack-reused 120 (from 1)

Receiving objects: 100% (120/120), 48.19 KiB | 9.64 MiB/s, done.

Resolving deltas: 100% (49/49), done.

Filtering content: 100% (5/5), 2.03 GiB | 3.80 MiB/s, done.

готово давай нататък

Copilot said: Страхотно! Моделът е успешно клониран на твоя

Страхотно! Моделът е успешно клониран на твоя компютър.

Следващи стъпки

1. Увери се, че всички файлове са изтеглени (особено ако виждаш файлове с размер 0 байта)

Ако има файлове с разширение като .bin или .safetensors, които са много малки, изпълни в директорията на модела:

bash

git lfs pull

Така ще се изтеглят всички големи (LFS) файлове.

2. Използване на модела в Python (пример с transformers)

Инсталирай transformers и нужните библиотеки:

bash

pip install transformers torch

Примерен код (в директорията на проекта ти):

Python

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "./gemma-2b"

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

inputs = tokenizer("Hello, how are you?", return\_tensors="pt")

outputs = model.generate(\*\*inputs, max\_new\_tokens=50)

print(tokenizer.decode(outputs[0], skip\_special\_tokens=True))

3. Ако искаш нещо специфично:

Файнтьонинг: Можеш да използваш Hugging Face Trainer или друг инструмент.

Интеграция с друга система: Кажи каква е, ще дам насоки.

Демо чатбот: Мога да ти дам код за бърз чат интерфейс.

Кажи за какво ще ползваш модела – чат, анализ, обучение, друг проект? Така ще ти дам най-точните инструкции!

You said: Примерен код (в директорията на проекта ти): Пайтън from

Примерен код (в директорията на проекта ти):

Пайтън

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "./gemma-2b"

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

inputs = tokenizer("Hello, how are you?", return\_tensors="pt")

outputs = model.generate(\*\*inputs, max\_new\_tokens=50)

print(tokenizer.decode(outputs[0], skip\_special\_tokens=True))

това къде го слагам във хтмл

Copilot said: Този примерен код е на Python и

Този примерен код е на Python и не може да се използва директно в HTML или в уеб страница, защото HTML е за уеб съдържание, а Python код (особено с transformers) се изпълнява на сървър или локално на твоя компютър.

Как да го използваш:

Кодът се слага във .py файл

Например, създай файл с име gemma\_test.py и постави кода там:

gemma\_test.py

v1

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "./gemma-2b"

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

Стартирай го през терминала:

bash

python gemma\_test.py

Ако искаш да използваш модела през HTML (уеб страница):

Трябва да направиш бекенд (напр. с Flask, FastAPI или Django на Python), който да приема заявки от HTML/JS и да връща отговори от модела.

HTML не може да зарежда такъв модел директно, защото това изисква Python и много RAM/CPU.

Примерна схема:

HTML формата изпраща заявка (prompt) към Python бекенд.

Python бекенд (Flask/FastAPI) обработва заявката с модела и връща отговор към HTML.

Ако искаш, мога да ти дам пример за такава уеб интеграция! Кажи ако искаш пример Flask + HTML форма.

Uploaded image: image.png

You said: така ли

така ли

Copilot said: Да, точно така! Това е правилният

Да, точно така!

Това е правилният начин – създаваш файл като gemma\_test.py в директорията на проекта (както си показала на снимката), поставяш Python кода вътре и после го стартираш през терминал:

bash

python gemma\_test.py

Ако имаш инсталирани transformers и torch – всичко трябва да работи.

Ако нямаш нужните библиотеки:

Инсталирай ги с:

bash

pip install transformers torch

Резюме:

Създаваш .py файл с кода (както си направила).

Стартираш го с Python от тази папка.

Ако има грешки – копирай ми ги тук, ще помогна.

Всичко си направила правилно!

You said: Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final/gemma-2b (main) $ python gemma\_test.py C:\Python313\python.exe:

Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final/gemma-2b (main)

$ python gemma\_test.py

C:\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\MyProjects\\stella-projects-final\\gemma-2b\\gemma\_test.py': [Errno 2] No such file or directory

Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final/gemma-2b (main)

$

Copilot said: Грешката означава, че файлът `gemma\_test.py` не се

Грешката означава, че файлът gemma\_test.py не се намира в папката, в която си в момента (gemma-2b).

Какво да направиш:

Провери къде е файлът:

От скрийншота по-горе, файлът gemma\_test.py е в папката stella-projects-final, а не вътре в gemma-2b.

Решения:

Вариант 1: Отиди една папка нагоре и стартирай:

bash

cd ..

python gemma\_test.py

Вариант 2: Стартирай файла с абсолютен път:

bash

python ../gemma\_test.py

(ако си вътре в gemma-2b)

Вариант 3: Премести файла gemma\_test.py вътре в папката gemma-2b и стартирай оттам.

Най-лесно:

Отиди в главната папка, където е файлът, с тази команда:

bash

cd ~/MyProjects/stella-projects-final

python gemma\_test.py

Резюме:

Трябва да стартираш скрипта от папката, където се намира самият файл. Ако има въпроси — питай!

You said: Python 3.13.5 (tags/v3.13.5:6cb20a2, Jun 11 2025, 16:15:46)

Python 3.13.5 (tags/v3.13.5:6cb20a2, Jun 11 2025, 16:15:46) [MSC v.1943 64 bit (AMD64)] on win32

Type "help", "copyright", "credits" or "license" for more information.

>>> cd ..

... python gemma\_test.py

... python ../gemma\_test.py

...

Copilot said: В Python интерактивната конзола (REPL), командите `cd

В Python интерактивната конзола (REPL), командите cd .., python gemma\_test.py и python ../gemma\_test.py няма да работят така, защото те са команден ред (shell) команди, а не Python код.

Как да стартираш Python скрипта правилно:

Излез от Python REPL

Ако виждаш >>> в началото на реда, значи си вътре в интерактивната Python среда.

Натисни Ctrl+Z и после Enter (на Windows), за да излезеш.

Използвай терминала/командния ред, а не Python REPL

В терминала (PowerShell, Command Prompt или Git Bash), изпълни:

bash

cd ~/MyProjects/stella-projects-final

python gemma\_test.py

или ако файлът е на друго място, отиди в папката, където се намира.

Важно:

Не въвеждай тези команди в Python REPL (там, където пише >>>), а ги въвеждай в самия терминал (Git Bash, PowerShell или Command Prompt).

Ако имаш нужда от още помощ – кажи!

You said: PS C:\WINDOWS\system32> cd ~/MyProjects/stella-projects-final PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python

PS C:\WINDOWS\system32> cd ~/MyProjects/stella-projects-final

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma\_test.py

C:\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\MyProjects\\stella-projects-final\\gemma\_test.py': [Errno 2] No such file or directory

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Copilot said: Грешката означава, че файлът `gemma\_test.py` не се

Грешката означава, че файлът gemma\_test.py не се намира в папката C:\Users\Stella Barbarella\MyProjects\stella-projects-final.

Какво да направиш:

Провери точното име на файла и къде се намира.

По твоята снимка, файлът се казва gemma\_test.py и се вижда в главната папка на проекта, така че командата трябва да работи, ако си в правилната папка.

Провери дали не си в грешна папка!

Ако случайно си отворила терминал в друга папка, командата ще даде тази грешка.

Провери с тази команда:

PowerShell

dir

Трябва да видиш сред файловете gemma\_test.py. Ако го няма, значи си в грешната папка.

Ако файлът е в поддиректория (например gemma-2b):

Влез в съответната папка:

PowerShell

cd gemma-2b

dir

След като си вътре и виждаш файла, стартирай:

PowerShell

python gemma\_test.py

Резюме:

Увери се, че си в папката, където е файлът.

Провери с dir.

Ако файлът е в друга папка, премини там с cd.

После стартирай скрипта с python gemma\_test.py.

Ако пак има проблем, копирай резултата от dir тук, за да видим какво има в папката!

You said: Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final/gemma-2b (main) $ ls README.md

Stella Barbarella@NucBox\_M3Plus MINGW64 ~/MyProjects/stella-projects-final/gemma-2b (main)

$ ls

README.md generation\_config.json special\_tokens\_map.json

config.json model-00001-of-00002.safetensors tokenizer.json

gemma-2b.gguf model-00002-of-00002.safetensors tokenizer.model

gemma\_test.py.txt model.safetensors.index.json tokenizer\_config.json

Copilot said: Твоят файл се казва \*\*gemma\_test.py.txt\*\* (има разширение .txt). Python

Твоят файл се казва gemma\_test.py.txt (има разширение .txt).

Python няма да го разпознае като изпълним скрипт, ако стартираш python gemma\_test.py.

Какво да направиш:

Преименувай файла на gemma\_test.py (махни .txt от края).

В Git Bash или PowerShell напиши:

bash

mv gemma\_test.py.txt gemma\_test.py

или с графичния интерфейс (десен бутон → Rename → изтрий “.txt”).

После стартирай:

bash

python gemma\_test.py

Резюме:

Файлът трябва да се казва gemma\_test.py, не gemma\_test.py.txt.

Преименувай го и пак стартирай командата!

You said: PS C:\WINDOWS\system32> cd ~/MyProjects/stella-projects-final PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python

PS C:\WINDOWS\system32> cd ~/MyProjects/stella-projects-final

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma\_test.py

C:\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\MyProjects\\stella-projects-final\\gemma\_test.py': [Errno 2] No such file or directory

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> mv gemma\_test.py.txt gemma\_test.py

mv : Impossible de trouver le chemin d'accès «C:\Users\Stella

Barbarella\MyProjects\stella-projects-final\gemma\_test.py.txt», car il n'existe pas.

Au caractère Ligne:1 : 1

+ mv gemma\_test.py.txt gemma\_test.py

+ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

+ CategoryInfo : ObjectNotFound: (C:\Users\Stella...mma\_test.py.txt:String) [Move-Item], ItemNotFoundExce

ption

+ FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.MoveItemCommand

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Copilot said: Файлът `gemma\_test.py.txt` не се намира в папката `stella-projects-final`, затова командата

Файлът gemma\_test.py.txt не се намира в папката stella-projects-final, затова командата не работи. От твоята по-ранна снимка, файлът се намира в поддиректорията gemma-2b!

Какво да направиш:

Отиди в папката, където е файлът:

PowerShell

cd gemma-2b

Виж дали файлът е там:

PowerShell

dir

Трябва да видиш gemma\_test.py.txt в списъка.

Преименувай файла:

PowerShell

ren gemma\_test.py.txt gemma\_test.py

(В PowerShell ren е правилната команда за преименуване.)

Стартирай скрипта:

PowerShell

python gemma\_test.py

Резюме:

Влез в папката gemma-2b

Преименувай файла с ren gemma\_test.py.txt gemma\_test.py

Стартирай с python gemma\_test.py

Ако имаш нужда от още помощ — пиши!

You said: Mode

Mode LastWriteTime Length Name

---- ------------- ------ ----

-a---- 25/06/2025 13:27 1657 .gitattributes

-a---- 25/06/2025 13:27 654 config.json

-a---- 25/06/2025 13:36 10031780672 gemma-2b.gguf

-a---- 25/06/2025 13:41 388 gemma\_test.py.txt

-a---- 25/06/2025 13:27 144 generation\_config.json

-a---- 25/06/2025 13:28 4945242264 model-00001-of-00002.safetensors

-a---- 25/06/2025 13:27 67121608 model-00002-of-00002.safetensors

-a---- 25/06/2025 13:27 13660 model.safetensors.index.json

-a---- 25/06/2025 13:27 21959 README.md

-a---- 25/06/2025 13:27 670 special\_tokens\_map.json

-a---- 25/06/2025 13:27 17518497 tokenizer.json

-a---- 25/06/2025 13:27 4241003 tokenizer.model

-a---- 25/06/2025 13:27 35072 tokenizer\_config.json

Copilot said: Виждам, че файлът \*\*gemma\_test.py.txt\*\* все още е

Виждам, че файлът gemma\_test.py.txt все още е с разширение .txt.

За Python трябва да е .py (без .txt накрая), за да можеш да го стартираш с:

bash

python gemma\_test.py

Как да го преименуваш правилно в PowerShell:

Убеди се, че си в папката, където е файлът (gemma-2b).

Изпълни тази команда в PowerShell:

PowerShell

ren gemma\_test.py.txt gemma\_test.py

(Можеш да копираш и поставиш тази команда!)

Провери, че имаш вече файл gemma\_test.py с:

PowerShell

dir

Стартирай скрипта:

PowerShell

python gemma\_test.py

Ако има грешка/съобщение – копирай го тук и ще ти помогна веднага!

You said: Répertoire : C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b Mode

Répertoire : C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b

Mode LastWriteTime Length Name

---- ------------- ------ ----

-a---- 25/06/2025 13:27 1657 .gitattributes

-a---- 25/06/2025 13:27 654 config.json

-a---- 25/06/2025 13:36 10031780672 gemma-2b.gguf

-a---- 25/06/2025 13:41 388 gemma\_test.py

-a---- 25/06/2025 13:27 144 generation\_config.json

-a---- 25/06/2025 13:28 4945242264 model-00001-of-00002.safetensors

-a---- 25/06/2025 13:27 67121608 model-00002-of-00002.safetensors

-a---- 25/06/2025 13:27 13660 model.safetensors.index.json

-a---- 25/06/2025 13:27 21959 README.md

-a---- 25/06/2025 13:27 670 special\_tokens\_map.json

-a---- 25/06/2025 13:27 17518497 tokenizer.json

-a---- 25/06/2025 13:27 4241003 tokenizer.model

-a---- 25/06/2025 13:27 35072 tokenizer\_config.json

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> python gemma\_test.py

Traceback (most recent call last):

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 470, in cached\_files

hf\_hub\_download(

~~~~~~~~~~~~~~~^

path\_or\_repo\_id,

^^^^^^^^^^^^^^^^

...<10 lines>...

local\_files\_only=local\_files\_only,

^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

)

^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 106, in \_inner\_fn

validate\_repo\_id(arg\_value)

~~~~~~~~~~~~~~~~^^^^^^^^^^^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 160, in validate\_repo\_id

raise HFValidationError(

...<3 lines>...

)

huggingface\_hub.errors.HFValidationError: Repo id must use alphanumeric chars or '-', '\_', '.', '--' and '..' are forbidden, '-' and '.' cannot start or end the name, max length is 96: './gemma-2b'.

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b\gemma\_test.py", line 5, in <module>

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\models\auto\tokenization\_auto.py", line 950, in from\_pretrained

tokenizer\_config = get\_tokenizer\_config(pretrained\_model\_name\_or\_path, \*\*kwargs)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\models\auto\tokenization\_auto.py", line 782, in get\_tokenizer\_config

resolved\_config\_file = cached\_file(

pretrained\_model\_name\_or\_path,

...<12 lines>...

\_commit\_hash=commit\_hash,

)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 312, in cached\_file

file = cached\_files(path\_or\_repo\_id=path\_or\_repo\_id, filenames=[filename], \*\*kwargs)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 523, in cached\_files

\_get\_cache\_file\_to\_return(path\_or\_repo\_id, filename, cache\_dir, revision) for filename in full\_filenames

~~~~~~~~~~~~~~~~~~~~~~~~~^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 140, in \_get\_cache\_file\_to\_return

resolved\_file = try\_to\_load\_from\_cache(path\_or\_repo\_id, full\_filename, cache\_dir=cache\_dir, revision=revision)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 106, in \_inner\_fn

validate\_repo\_id(arg\_value)

~~~~~~~~~~~~~~~~^^^^^^^^^^^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 160, in validate\_repo\_id

raise HFValidationError(

...<3 lines>...

)

huggingface\_hub.errors.HFValidationError: Repo id must use alphanumeric chars or '-', '\_', '.', '--' and '..' are forbidden, '-' and '.' cannot start or end the name, max length is 96: './gemma-2b'.

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b>

Copilot said: Грешката възниква, защото HuggingFace Transformers \*\*не приема

Грешката възниква, защото HuggingFace Transformers не приема пътя с ./gemma-2b като валиден локален път, когато вече се намираш вътре в тази папка. Той очаква или:

само име на папката (ако си една папка по-нагоре)

или просто "." за текущата директория.

Как да го оправиш:

Ако си в папката gemma-2b (както сега):

Промени тези два реда:

Python

model\_path = "."

или дори:

Python

model\_path = ""

и използвай model\_path по-надолу, както досега.

Примерен работещ код:

gemma\_test.py

v2

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "."

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

Алтернативно:

Ако искаш да ползваш model\_path = "./gemma-2b", трябва да стартираш скрипта от една папка по-нагоре (от stella-projects-final).

Резюме:

Ако стартираш скрипта в папката gemma-2b, ползвай model\_path = ".".

Ако си една папка по-нагоре, ползвай model\_path = "./gemma-2b".

Пробвай тази промяна! Ако има нова грешка – копирай я тук!

You said: from transformers import AutoModelForCausalLM, AutoTokenizer model\_path = "./gemma-2b" tokenizer

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "./gemma-2b"

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

inputs = tokenizer("Hello, how are you?", return\_tensors="pt")

outputs = model.generate(\*\*inputs, max\_new\_tokens=50)

print(tokenizer.decode(outputs[0], skip\_special\_tokens=True))

Copilot said: Този код ще работи, ако се стартира

Този код ще работи, ако се стартира от директорията, която е една папка по-нагоре от gemma-2b (т.е. от stella-projects-final).

В момента обаче си вътре в папката gemma-2b.

Как да го използваш правилно:

Вариант 1: Стартирай от родителската папка

Отиди една папка нагоре:

bash

cd ..

Стартирай:

bash

python gemma-2b/gemma\_test.py

или ако е в главната:

bash

python gemma\_test.py

Вариант 2: Промени пътя в кода, ако искаш да стартираш от gemma-2b

Промени само тази линия:

Python

model\_path = "."

или

Python

model\_path = ""

Така ще работи, когато скриптът е в същата папка като моделните файлове.

Примерен коригиран код за работа от папка gemma-2b:

gemma\_test.py

v3

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "."

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

Избери един от двата подхода според това от коя папка ще стартираш!

Ако има друга грешка или въпрос — пиши!

You said: huggingface\_hub.errors.HFValidationError: Repo id must use alphanumeric chars

huggingface\_hub.errors.HFValidationError: Repo id must use alphanumeric chars or '-', '\_', '.', '--' and '..' are forbidden, '-' and '.' cannot start or end the name, max length is 96: './gemma-2b'.

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> cd ..

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma-2b/gemma\_test.py

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma\_test.py

C:\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\MyProjects\\stella-projects-final\\gemma\_test.py': [Errno 2] No such file or directory

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Copilot said: Чудесно! След като си една папка по-нагоре

Чудесно! След като си една папка по-нагоре (в stella-projects-final), командата:

PowerShell

python gemma-2b/gemma\_test.py

е правилният начин да стартираш скрипта, когато пътят в кода е:

Python

model\_path = "./gemma-2b"

Забележка:

Грешката

Code

C:\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\MyProjects\\stella-projects-final\\gemma\_test.py': [Errno 2] No such file or directory

се появява, защото файлът gemma\_test.py не е в текущата папка – той е вътре в поддиректорията gemma-2b. Затова трябва да стартираш с:

PowerShell

python gemma-2b/gemma\_test.py

или в Git Bash:

bash

python gemma-2b/gemma\_test.py

Резюме:

Ако си в папка stella-projects-final, стартирай с:

PowerShell

python gemma-2b/gemma\_test.py

Ако си в gemma-2b, стартирай с:

PowerShell

python gemma\_test.py

и смени в кода model\_path = "."

Ако скриптът не дава нищо:

Ако просто се върнеш на нов ред без изход, вероятно има проблем с модела или зависимостите.

Ако не виждаш грешка, пробвай да добавиш print("Starting inference…") най-отгоре, за да видиш дали изобщо изпълнява скрипта.

Ако има нова грешка или нещо не работи — копирай съобщението тук!

Така ще те водя стъпка по стъпка!

You said: face\_hub.errors.HFValidationError: Repo id must use alphanumeric chars

face\_hub.errors.HFValidationError: Repo id must use alphanumeric chars or '-', '\_', '.', '--' and '..' are forbidden, '-' and '.' cannot start or end the name, max length is 96: './gemma-2b'.

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> cd ..

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma-2b/gemma\_test.py

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma\_test.py

C:\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\MyProjects\\stella-projects-final\\gemma\_test.py': [Errno 2] No such file or directory

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> model\_path = "./gemma-2b"

model\_path : Le terme «model\_path» n'est pas reconnu comme nom d'applet de commande, fonction, fichier de script ou programme exécutable. Vérifiez l'orthographe du nom, ou si un chemin d'accès existe, vérifiez que le chemin d'accès est

correct et réessayez.

Au caractère Ligne:1 : 1

+ model\_path = "./gemma-2b"

+ ~~~~~~~~~~

+ CategoryInfo : ObjectNotFound: (model\_path:String) [], CommandNotFoundException

+ FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Copilot said: Ето какво става и как да го

Ето какво става и как да го оправиш стъпка по стъпка:

1. PowerShell не разбира Python синтаксис

Това:

Code

model\_path = "./gemma-2b"

е Python код, а ти го въвеждаш в PowerShell (затова дава грешка CommandNotFoundException).

Не въвеждай Python код в терминала!

Промени model\_path директно в gemma\_test.py с текстов редактор (Notepad, VS Code и т.н.).

2. Използвай правилния път, съобразно папката, от която стартираш

Вариант А: Стартираш от stella-projects-final

Увери се, че в gemma\_test.py имаш:

Python

model\_path = "./gemma-2b"

Стартирай с:

PowerShell

python gemma-2b/gemma\_test.py

Вариант Б: Стартираш от gemma-2b

В gemma\_test.py сложи:

Python

model\_path = "."

Стартирай с:

PowerShell

python gemma\_test.py

3. Не очаквай изход, ако стартираш с грешния път или от грешна папка!

Ако стартираш с грешен път, ще получаваш HFValidationError.

Ако стартираш командата от папка, където файлът не съществува, ще получиш [Errno 2] No such file or directory.

Пример за правилна последователност (от stella-projects-final):

Провери, че в gemma\_test.py пише:

Python

model\_path = "./gemma-2b"

Стартирай:

PowerShell

python gemma-2b/gemma\_test.py

Пример за правилна последователност (от gemma-2b):

Провери, че в gemma\_test.py пише:

Python

model\_path = "."

Стартирай:

PowerShell

python gemma\_test.py

Не въвеждай Python променливи в терминала! Променяй ги само в самия файл.

Ако следваш тези стъпки, всичко ще тръгне. Ако има нова грешка – копирай цялото съобщение и ще ти помогна!

You said: Type "help", "copyright", "credits" or "license" for

Type "help", "copyright", "credits" or "license" for more information.

>>> model\_path = "./gemma-2b"

>>> model\_path = "./gemma-2b"

>>> model\_path = "./gemma-2b"

>>> python gemma-2b/gemma\_test.py

File "<stdin>", line 1

python gemma-2b/gemma\_test.py

^

SyntaxError: invalid decimal literal

>>>

Copilot said: Виждам, че опитваш да изпълниш командата `python gemma-2b/gemma\_test.py` в

Виждам, че опитваш да изпълниш командата python gemma-2b/gemma\_test.py в Python интерактивната среда (REPL), където пише >>>. Това няма да работи!

Как да стартираш правилно Python скрипта:

Излез от Python REPL!

Ако виждаш >>>, натисни Ctrl+Z и после Enter (на Windows) или напиши exit() и натисни Enter.

Върни се в терминала (PowerShell, Command Prompt или Git Bash).

Там ще имаш стандартния промпт (например: PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>).

Стартирай скрипта от терминала, не от Python REPL!

Командата трябва да се напише там така:

PowerShell

python gemma-2b/gemma\_test.py

Резюме:

Не въвеждай команди като python ... в интерактивния Python прозорец (>>>).

Изпълнявай ги само в терминала/PowerShell/Bash.

Ако направиш това и имаш грешка при стартиране — копирай грешката тук, за да помогна!

You said: ~~~~ + CategoryInfo

~~~~

+ CategoryInfo : ObjectNotFound: (model\_path:String) [], CommandNotFoundException

+ FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma-2b/gemma\_test.py

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma-2b/gemma\_test.py

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Copilot said: Виждам, че вече изпълняваш правилната команда в

Виждам, че вече изпълняваш правилната команда в PowerShell:

PowerShell

python gemma-2b/gemma\_test.py

Ако терминалът просто се връща на нов ред без никакъв изход или грешка, това означава едно от следните:

Скриптът приключва без да отпечата нищо (може би няма print или има празен try/except).

Python процесът се е сринал много рано (но обикновено показва грешка).

Скриптът е приключил успешно, но няма видим резултат.

Какво да направиш:

1. Провери дали скриптът изобщо работи

Добави най-отгоре на gemma\_test.py:

Python

print("Script started!")

Ако видиш това съобщение, значи скриптът се изпълнява.

2. Увери се, че имаш print() за резултата

В твоя скрипт трябва да има в края примерно:

Python

print(tokenizer.decode(outputs[0], skip\_special\_tokens=True))

3. Ако нещо увисва/замръзва

Ако моделът е голям, може просто да чакаш дълго: изчакай няколко минути!

Ако няма грешка, но няма и изход - добави още print() между стъпките, за да видиш докъде стига скриптът.

4. Ако искаш – покажи съдържанието на gemma\_test.py тук (копирай целия файл), за да го проверя!

В обобщение:

Провери дали имаш поне един print() в началото на скрипта.

Стартирай отново.

Ако няма съобщение, копирай съдържанието на скрипта тук или опиши какво прави.

Така ще намерим къде е проблемът!

You said: PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma-2b/gemma\_test.py PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma-2b/gemma\_test.py

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> print("Script started!")

Impossible d’initialiser le périphérique PRN

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Copilot said: Това поведение показва: 1. \*\*Скриптът се изпълнява, но

Това поведение показва:

Скриптът се изпълнява, но не извежда нищо и не дава грешка.

Когато пишеш print("Script started!") директно в PowerShell, получаваш грешка Impossible d’initialiser le périphérique PRN — това е нормално, защото print е Python команда, а ти я пишеш в PowerShell!

Какво да правиш:

1. Добави print в самия Python файл

Отвори gemma\_test.py с текстов редактор и добави на първия ред:

Python

print("Script started!")

Запази файла.

2. Стартирай отново с:

PowerShell

python gemma-2b/gemma\_test.py

Ако не се появи „Script started!“:

Скриптът може да е празен или да не се изпълнява.

Провери, че редактираш файла gemma\_test.py в папка gemma-2b.

Ако излезе „Script started!“, но после нищо не става:

Добави още print() след всеки основен ред в кода (например след зареждането на tokenizer, след зареждането на model и т.н.), за да разбереш докъде стига изпълнението.

Пример:

gemma-2b/gemma\_test.py

v1

print("Script started!")

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "./gemma-2b"

print("Loading tokenizer...")

Това ще ти покаже на кой ред евентуално „увисва“ или спира скрипта!

Ако все още няма изход, копирай съдържанието на целия скрипт тук, за да ти помогна!

You said: >>> print("Script started!") Script started! >>>

>>> print("Script started!")

Script started!

>>>

Copilot said: Перфектно! Това означава, че когато въвеждаш `print("Script started!")` в

Перфектно! Това означава, че когато въвеждаш print("Script started!") в Python REPL (>>>), работи, но това НЕ е начинът да стартираш твоя скрипт. Трябва да го добавиш в самия файл gemma\_test.py.

Как да продължиш (стъпка по стъпка):

Отвори gemma\_test.py с текстов редактор (Notepad, VS Code, др.).

В началото на файла добави този ред:

Python

print("Script started!")

Запази файла.

Върни се в PowerShell (или друг терминал, НЕ в Python REPL!).

Стартирай:

PowerShell

python gemma-2b/gemma\_test.py

Ако виждаш “Script started!” — скриптът се изпълнява!

Ако не виждаш нищо, копирай съдържанието на gemma\_test.py тук.

ВАЖНО:

Не въвеждай Python код (като print(...)) в PowerShell или терминала, освен ако не си в Python REPL (>>>).

Стартирай скрипта само с: python gemma-2b/gemma\_test.py

Ако искаш да тестваме дали скриптът стига до някъде — добавяй print(...) след всяка важна стъпка.

Ако нещо не работи, копирай съдържанието на gemma\_test.py тук и ще ти помогна!

You said: PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> print("Script started!") Impossible d’initialiser le

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> print("Script started!")

Impossible d’initialiser le périphérique PRN

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma-2b/gemma\_test.py

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma-2b/gemma\_test.py

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Copilot said: Виждам, че скриптът не дава никакъв изход

Виждам, че скриптът не дава никакъв изход и не показва грешка. Това може да се случи ако:

В самия Python файл няма нито един print()

Скриптът "увисва" при зареждане на модел (което може да отнеме дълго време)

Има скрита грешка, която се "поглъща" (например в try/except без print на грешките)

Скриптът приключва успешно, но не прави нищо видимо

Как да разберем какво става

Провери съдържанието на файла!

Моля, копирай и постави целия код на gemma\_test.py тук.

Добави print-и за диагностика!

Добави следното в началото и между всяка стъпка:

gemma-2b/gemma\_test.py

v2

print("Script started!")

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "./gemma-2b"

print("Loading tokenizer...")

Запази файла и стартирай пак:

PowerShell

python gemma-2b/gemma\_test.py

Ако пак няма нищо — провери:

Вярно ли редактираш правилния файл?

В тази папка ли си?

Имаш ли достатъчно RAM? (Моделът е огромен и може да "убие" процеса без да даде грешка!)

Ако отново няма изход

Копирай и постави съдържанието на gemma\_test.py тук, за да го проверя!

Кажи дали компютърът ти има достатъчно RAM (над 12-16GB?).

Ако PowerShell просто се връща на нов ред, моделът може да е твърде голям за машината ти и Python процесът се убива от Windows без съобщение.

Изпълни горните стъпки и пиши какво се случва или дай кода на скрипта!

Uploaded image: image.png

You said: Describe this image

Describe this image

Copilot said: The image shows a Windows PowerShell terminal

The image shows a Windows PowerShell terminal where you have successfully run the command:

Code

python gemma-2b/gemma\_test.py

The script prints progress messages:

"Script started!"

"Loading tokenizer..."

"Tokenizer loaded!"

"Loading model..."

A progress bar saying "Loading checkpoint shards: 100%"

"Model loaded!"

"Inputs tokenized."

"Output generated."

"Hello, how are you?"

Finally, the script prints a generated response:

Code

I'm a 20 year old student from Brazil. I'm a big fan of the game and I'm looking for a team to play with. I'm a good player and I'm looking for a team to

This indicates that your local language model (likely Gemma 2B) loaded successfully and generated a text completion in response to the prompt "Hello, how are you?". The output is a plausible English sentence introducing the speaker as a student from Brazil looking for a team, showing the model is functioning as expected.

Uploaded image: image.png

You said: Describe this image

Describe this image

Copilot said: This image shows a terminal session where

This image shows a terminal session where Python's interactive shell (REPL) is running, indicated by the >>> prompt. The user attempts to execute the command:

Code

python gemma-2b/gemma\_test.py

This results in the error:

Code

SyntaxError: invalid decimal literal

This error occurs because python gemma-2b/gemma\_test.py is a shell command, not valid Python syntax. It cannot be run inside the Python REPL.

The user also runs:

Code

print("Script started!")

which works, since this is valid Python code.

In summary:

Shell commands (like running a .py script) must be entered in the terminal/PowerShell, not in the Python REPL (>>>).

Python code (like print(...)) should be run inside the Python REPL or in a .py file.

You said:

^

SyntaxError: invalid decimal literal

>>> print("Script started!")

Script started!

>>> python gemma-2b/gemma\_test.py

File "<stdin>", line 1

python gemma-2b/gemma\_test.py

^

SyntaxError: invalid decimal literal

>>> print("Script started!")

Script started!

>>>

Copilot said: You are seeing this error because you

You are seeing this error because you are trying to run a shell command (python gemma-2b/gemma\_test.py) inside the Python interactive prompt (REPL), which starts with >>>.

In the Python REPL, only valid Python code is accepted. Shell commands like python ... will always cause a SyntaxError.

What to do:

Exit the Python REPL:

Type exit() or press Ctrl+Z then Enter on Windows.

Run your script in the terminal/powershell, NOT in the Python REPL.

You should see a prompt like:

Code

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Then, type:

Code

python gemma-2b/gemma\_test.py

and press Enter.

Summary:

Use the Python REPL (>>>) only for Python statements like print("hello").

Run scripts (like python gemma-2b/gemma\_test.py) in the regular terminal, not in the REPL.

If you follow these steps, your script will run without that SyntaxError.

You said: checkpoint shards: 100%|██████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████| 2/2 [00:05<00:00, 2.58s/it] Model

checkpoint shards: 100%|██████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████| 2/2 [00:05<00:00, 2.58s/it]

Model loaded!

Inputs tokenized.

Output generated.

Hello, how are you?

I'm a 20 year old student from Brazil. I'm a big fan of the game and I'm looking for a team to play with. I'm a good player and I'm looking for a team to

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma-2b/gemma\_test.py

Script started!

Loading tokenizer...

Tokenizer loaded!

Loading model...

Loading checkpoint shards: 100%|██████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████| 2/2 [00:01<00:00, 1.01it/s]

Model loaded!

Inputs tokenized.

Output generated.

Hello, how are you?

I'm a 20 year old student from Brazil. I'm a big fan of the game and I'm looking for a team to play with. I'm a good player and I'm looking for a team to

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Copilot said: Great work! Your output shows that

Great work!

Your output shows that your script gemma\_test.py runs successfully, loads the Gemma-2B model, processes the input, and generates a text response. Here’s what’s happening:

What’s working:

Model and tokenizer are loading fine:

You see progress bars and the "Model loaded!" message.

The script processes the input:

"Hello, how are you?"

You get a generated output:

Code

I'm a 20 year old student from Brazil. I'm a big fan of the game and I'm looking for a team to play with. I'm a good player and I'm looking for a team to

(the output seems to be cut off, which is typical for short max tokens or model output truncation)

What this means:

Your Python environment, Hugging Face Transformers library, and model files are all set up correctly.

The script is running as intended.

The output of the model is being printed to your terminal.

What’s next?

If you want to:

1. Change the prompt

Edit your script and change the line:

Python

inputs = tokenizer("Hello, how are you?", return\_tensors="pt")

to something else, e.g.:

Python

inputs = tokenizer("Tell me a joke.", return\_tensors="pt")

2. Generate more text (longer output)

Increase the max\_new\_tokens parameter:

Python

outputs = model.generate(\*\*inputs, max\_new\_tokens=100)

(Default in your script is likely 50.)

3. Save or use the output programmatically

Assign the result to a variable, save to file, or process further in Python.

4. Experiment with model parameters

You can adjust sampling, temperature, top\_k, etc. in the generate() function.

5. Make it interactive

Accept user input with input() and feed that to the model.

If you have a new question or want to try something else, just ask!

If you want help with making the script interactive, saving results, changing the prompt, or understanding the output—let me know!

You said: искам да се зарежда нормално със поздрав

искам да се зарежда нормално със поздрав и да го включа в сайта ми

Copilot said: Разбира се! Ето как можеш да направиш

Разбира се! Ето как можеш да направиш две неща:

1. Моделът да отговаря с поздрав

За да започва винаги с поздрав (например „Здравей, как мога да помогна?“), можеш да смениш входния prompt в твоя скрипт:

Python

inputs = tokenizer("Здравей! Как мога да помогна?", return\_tensors="pt")

или друг поздрав по твой избор.

2. Интеграция в сайт

Вариант А: Ако сайтът е на Python (например Flask, Django)

Можеш да направиш малък HTTP сървър, който при заявка зарежда модела и връща отговор. Пример с Flask:

app.py

from flask import Flask, request, jsonify

from transformers import AutoModelForCausalLM, AutoTokenizer

app = Flask(\_\_name\_\_)

model\_path = "./gemma-2b"

Тогава от сайта си правиш POST заявка към /generate с JSON:

JSON

{"prompt": "Здравей! Как се казваш?"}

и ще получиш автоматичен отговор.

Вариант Б: Ако сайтът е на JavaScript/HTML

Трябва бекендът (като горния Flask пример) да е пуснат и сайтът да прави заявка към него с JavaScript (fetch/AJAX).

JavaScript

fetch('http://localhost:5000/generate', {

method: 'POST',

headers: { 'Content-Type': 'application/json' },

body: JSON.stringify({ prompt: "Здравей!" })

})

.then(response => response.json())

.then(data => {

document.getElementById('answer').innerText = data.response;

});

Ако сайтът ти е на WordPress, Wix, или друг език — кажи какъв е, за да ти дам конкретен пример!

Обобщение:

Смени prompt-а в Python за поздрав.

Добави Flask сървър за връзка със сайта.

В сайта пращай заявка към този сървър и показвай отговора.

Кажи ми какво точно искаш — поздрав само при първо стартиране, или всеки път? И на какъв език/платформа е сайтът ти?

You said: Python 3.13.5 (tags/v3.13.5:6cb20a2, Jun 11 2025, 16:15:46)

Python 3.13.5 (tags/v3.13.5:6cb20a2, Jun 11 2025, 16:15:46) [MSC v.1943 64 bit (AMD64)] on win32

Type "help", "copyright", "credits" or "license" for more information.

>>> cd ..

... python gemma\_test.py

... python ../gemma\_test.py

... ^Z

... inputs = tokenizer("Здравей! Как мога да помогна?", return\_tensors="pt") тук така ли

Copilot said: Почти си уцелил! Позволи ми да обясня: -

Почти си уцелил! Позволи ми да обясня:

В интерактивния Python REPL (>>>), не можеш да използваш shell команди като cd .. или python gemma\_test.py.

Редът inputs = tokenizer("Здравей! Как мога да помогна?", return\_tensors="pt") е Python код, който трябва да бъде в твоя Python файл (gemma\_test.py), а не да го пишеш директно в терминала/REPL, ако пускаш скрипта като програма.

Как да го направиш стъпка по стъпка:

1. Редактирай файла gemma\_test.py

Отвори файла с текстов редактор и намери реда:

Python

inputs = tokenizer("Hello, how are you?", return\_tensors="pt")

Смени го на:

Python

inputs = tokenizer("Здравей! Как мога да помогна?", return\_tensors="pt")

2. Запази файла.

3. Стартирай скрипта от PowerShell/Terminal (НЕ от Python REPL!)

Върни се в терминала, така че да виждаш нещо като:

Code

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Там напиши:

Code

python gemma-2b/gemma\_test.py

или ако файлът е в текущата директория:

Code

python gemma\_test.py

Резюме

Редът който написа е правилен, но трябва да е във файла.

Не ползвай shell команди в REPL.

Пускай скрипта от терминала, не от Python REPL.

Кажи ако искаш примерен цял файл или имаш въпроси за интеграция със сайт!

You said: Script started! Loading tokenizer... Tokenizer loaded! Loading model... Loading checkpoint shards:

Script started!

Loading tokenizer...

Tokenizer loaded!

Loading model...

Loading checkpoint shards: 100%|██████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████| 2/2 [00:02<00:00, 1.47s/it]

Model loaded!

Inputs tokenized.

Output generated.

Hello, how are you?

I'm a 20 year old student from Brazil. I'm a big fan of the game and I'm looking for a team to play with. I'm a good player and I'm looking for a team to

Copilot said: Всичко работи успешно! Моделът зарежда, обработва

Всичко работи успешно!

Моделът зарежда, обработва входа и генерира отговор.

За да го направиш с поздрав на български:

Редактирай скрипта – смени реда, в който се подава входният текст на модела на:

Python

inputs = tokenizer("Здравей! Как мога да помогна?", return\_tensors="pt")

Така моделът ще отговаря на поздрав на български.

Ако искаш да го интегрираш в сайт:

Направи Flask API (ако сайтът е на Python) – мога да ти дам готов пример.

Свържи сайта чрез JavaScript fetch/AJAX към това API.

Примерен Flask сървър:

chat\_api.py

v1

from flask import Flask, request, jsonify

from transformers import AutoModelForCausalLM, AutoTokenizer

app = Flask(\_\_name\_\_)

model\_path = "./gemma-2b"

Как да го ползваш от сайт (JavaScript):

JavaScript

fetch('http://localhost:5000/generate', {

method: 'POST',

headers: { 'Content-Type': 'application/json' },

body: JSON.stringify({ prompt: "Здравей! Как мога да помогна?" })

})

.then(response => response.json())

.then(data => {

document.getElementById('answer').innerText = data.response;

});

Ако имаш специфичен сайт (WordPress, PHP, React и т.н.) — кажи, за да дам подходящ пример!

Ако искаш примерен цял файл за Flask или друго — ще ти го напиша!

You said: зашо не отговаряш

зашо не отговаряш

Copilot said: Отговарям! :) Понякога има малко забавяне,

Отговарям! :)

Понякога има малко забавяне, но вече получаваш отговорите ми.

Ако имаш конкретен въпрос или искаш помощ с кода, сайта или нещо друго – пиши направо! Ако имаш предвид нещо конкретно по последната инструкция (за поздрав или интеграция в сайт), кажи какво не се получава или какво искаш да стане. Тук съм, за да помогна!

You said: сайта е на хтмл

сайта е на хтмл

Copilot said: Страхотно! Ако сайтът ти е на HTML

Страхотно! Ако сайтът ти е на HTML (и JavaScript), най-лесният начин да включиш Python модел като Gemma е да го стартираш като локален уеб сървър (например с Flask) и да пращаш заявки към него с JavaScript (fetch/AJAX).

Примерно решение

1. Стартирай Python Flask сървър

Създай файл chat\_api.py:

chat\_api.py

v2

from flask import Flask, request, jsonify

from transformers import AutoModelForCausalLM, AutoTokenizer

app = Flask(\_\_name\_\_)

model\_path = "./gemma-2b"

Стартирай сървъра с:

Code

python chat\_api.py

Той ще слуша на http://localhost:5000/generate

2. HTML + JavaScript за сайта

Примерен HTML файл с форма за чат:

index.html

<!DOCTYPE html>

<html lang="bg">

<head>

<meta charset="UTF-8">

<title>Бот с Gemma</title>

<style>

Как работи:

Пускаш Flask сървъра:

python chat\_api.py

Отваряш твоя HTML файл (index.html) в браузър.

Пишеш въпрос, натискаш "Питай" — изпраща заявка до Python сървъра, получава отговор и го показва на екрана.

Важно:

Този пример работи само локално (на твоя компютър), освен ако не направиш сървъра достъпен в интернет.

Ако искаш да го качваш на хостинг, трябва бекендът (Flask + Gemma) да е наличен и там, а не само HTML-а.

Ако искаш повече персонализация или имаш специфичен въпрос — питай!

You said: NameError: name 'tokenizer' is not defined >>> python

NameError: name 'tokenizer' is not defined

>>> python chat\_api.py

File "<stdin>", line 1

python chat\_api.py

^^^^^^^^

SyntaxError: invalid syntax

Copilot said: Виждам две различни грешки: --- ## 1. \*\*NameError: name

Виждам две различни грешки:

1. NameError: name 'tokenizer' is not defined

Причина:

Пишеш Python код (например inputs = tokenizer(...)) в интерактивната Python среда (REPL), но не си дефинирал променливата tokenizer преди това.

Решение:

Този код трябва да се изпълнява в скрипта ти (например chat\_api.py), където вече имаш редове като:

Python

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

Не въвеждай отделни редове от скрипта директно в REPL. Стартирай целия скрипт през терминала.

2. SyntaxError: invalid syntax

Виждаш тази грешка, защото се опитваш да изпълниш shell команда (python chat\_api.py) вътре в Python REPL (>>>).

Правилно:

Shell командите (стартиране на скрипт) се пишат само в командния ред/PowerShell, например:

Code

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python chat\_api.py

НЕ в Python REPL (който започва с >>>).

Как да го направиш правилно:

Затвори Python REPL

Ако виждаш >>>, напиши exit() и натисни Enter.

Върни се в терминал/PowerShell

Ще видиш нещо като:

Code

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Стартирай там скрипта:

Code

python chat\_api.py

Резюме:

Всички shell команди се пишат в терминала, не в Python REPL.

Не изпълнявай отделни парчета от Python скрипта си в REPL, а целия файл чрез python chat\_api.py.

Ако нещо пак не тръгва – копирай цялото съобщение за грешка тук!

You said: SyntaxError: invalid syntax >>> tokenizer = AutoTokenizer.from\_pretrained(model\_path) Traceback (most

SyntaxError: invalid syntax

>>> tokenizer = AutoTokenizer.from\_pretrained(model\_path)

Traceback (most recent call last):

File "<stdin>", line 1, in <module>

NameError: name 'AutoTokenizer' is not defined

>>>

Copilot said: Тази грешка означава, че се опитваш да

Тази грешка означава, че се опитваш да използваш AutoTokenizer в интерактивната Python среда (REPL), но не си го импортирал преди това. Освен това, както обсъдихме по-горе, не трябва да пишеш тези редове в REPL. Скриптът трябва да се стартира директно от терминала, а не ред по ред в Python REPL.

Какво да направиш:

1. Не използвай Python REPL (>>>) за стартиране на скрипта!

Излез от REPL като напишеш exit() и натиснеш Enter.

2. Импортирай правилно модулите, ако работиш в скрипт

В твоя Python файл (например chat\_api.py или gemma\_test.py) трябва да имаш най-горе:

Python

from transformers import AutoModelForCausalLM, AutoTokenizer

3. Стартирай целия скрипт от терминала, а не от REPL

В терминала (PowerShell, CMD):

Code

python chat\_api.py

или

Code

python gemma\_test.py

Резюме:

Не пишеш tokenizer = ... директно в REPL!

Всичко трябва да е във файл, който после стартираш от терминала.

Увери се, че си импортирал нужните библиотеки в началото на файла.

Примерно начало на твой Python файл:

gemma\_test.py

v4

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "./gemma-2b"

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

Стартирай го така:

Code

python gemma\_test.py

Ако искаш да ти напиша цял примерен файл, кажи!

You said: >>> from transformers import AutoModelForCausalLM, AutoTokenizer Traceback (most

>>> from transformers import AutoModelForCausalLM, AutoTokenizer

Traceback (most recent call last):

File "<stdin>", line 1, in <module>

ModuleNotFoundError: No module named 'transformers'

>>>

Copilot said: Грешката `ModuleNotFoundError: No module named 'transformers'` означава,

Грешката ModuleNotFoundError: No module named 'transformers' означава, че библиотеката transformers (от Hugging Face) не е инсталирана в твоя Python.

Решение: Инсталирай transformers

Излез от Python REPL:

Ако виждаш >>>, напиши:

Code

exit()

и натисни Enter.

Върни се в PowerShell или терминал

Трябва да виждаш нещо като:

Code

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Инсталирай transformers с pip:

Напиши:

Code

pip install transformers

Ако ползваш специфична версия на Python (примерно python3.11), може да пробваш:

Code

python -m pip install transformers

или:

Code

python3 -m pip install transformers

Изчакай да завърши инсталацията.

Пробвай отново кода или стартирай твоя скрипт.

Ако видиш грешка „pip не е разпознат“:

Може би трябва да инсталираш pip или да добавиш Python към PATH.

Ако ползваш Anaconda, пробвай:

Code

conda install -c conda-forge transformers

След като успешно инсталираш transformers, можеш да стартираш твоя Python скрипт!

Ако срещнеш нова грешка, копирай я тук!

You said: S C:\Users\Stella Barbarella> cd C:\Users\Stella Barbarella\MyProjects\stella-projects-final> Set-Location :

S C:\Users\Stella Barbarella> cd C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Set-Location : Impossible de trouver un paramètre positionnel acceptant l'argument «

Barbarella\MyProjects\stella-projects-final>».

Au caractère Ligne:1 : 1

+ cd C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

+ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

+ CategoryInfo : InvalidArgument : (:) [Set-Location], ParameterBindingException

+ FullyQualifiedErrorId : PositionalParameterNotFound,Microsoft.PowerShell.Commands.SetLocationCommand

PS C:\Users\Stella Barbarella>

Copilot said: Грешката е, защото си сложил излишен символ

Грешката е, защото си сложил излишен символ > накрая на пътя при командата за смяна на директория.

В PowerShell командата трябва да изглежда така:

PowerShell

cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final"

Без символа > накрая!

Ако има интервали в пътя (като „Stella Barbarella“), винаги слагай кавички около целия път.

Как да го направиш правилно:

Копирай тази команда и я постави в PowerShell:

Code

cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final"

Натисни Enter.

След това можеш спокойно да стартираш Python скрипта си с:

Code

python chat\_api.py

или

Code

python gemma\_test.py

Ако имаш други въпроси или грешки – пиши!

You said: :\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\MyProjects\\stella-projects-final\\chat\_api.py': [Errno

:\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\MyProjects\\stella-projects-final\\chat\_api.py': [Errno 2] No such file or directory

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final> python gemma\_test.py

C:\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\MyProjects\\stella-projects-final\\gemma\_test.py': [Errno 2] No such file or directory

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final>

Copilot said: Тези съобщения означават, че във \*\*C:\Users\Stella Barbarella\MyProjects\stella-projects-final\*\* няма файлове `chat\_api.py` или `gemma\_test.py`.

Тези съобщения означават, че във C:\Users\Stella Barbarella\MyProjects\stella-projects-final няма файлове chat\_api.py или gemma\_test.py.

Възможни причини:

Файловете не са създадени.

Файловете се намират в друга папка.

Името на файла е изписано грешно (например Gemma\_test.py, GemmaTest.py и т.н.).

Какво да направиш:

1. Провери дали файлът съществува

В PowerShell напиши:

Code

dir

Ще видиш списък с файловете в текущата папка.

Ако няма chat\_api.py и/или gemma\_test.py — трябва да ги създадеш.

Ако са с друго име — използвай точния файл, който виждаш.

2. Ако НЯМА такъв файл — създай го

Например, отвори Notepad и копирай този код за тест:

gemma\_test.py

v5

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "./gemma-2b"

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

Запази файла като gemma\_test.py в папката C:\Users\Stella Barbarella\MyProjects\stella-projects-final

3. Ако файлът Е там, но не е в текущата папка

Отиди в правилната папка с:

Code

cd "C:\Users\Stella Barbarella\MyProjects\стара-папка"

Или премести файла в желаната папка.

4. След като файлът е там

Стартирай:

Code

python gemma\_test.py

Ако не си сигурна, просто напиши dir и ми копирай списъка с файловете — ще ти кажа как да продължиш!

You said: PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> dir

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> dir

Répertoire : C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b

Mode LastWriteTime Length Name

---- ------------- ------ ----

-a---- 25/06/2025 13:27 1657 .gitattributes

-a---- 25/06/2025 14:18 711 chat\_appi.py.txt

-a---- 25/06/2025 13:27 654 config.json

-a---- 25/06/2025 13:36 10031780672 gemma-2b.gguf

-a---- 25/06/2025 14:03 584 gemma\_test.py

-a---- 25/06/2025 13:27 144 generation\_config.json

-a---- 25/06/2025 13:28 4945242264 model-00001-of-00002.safetensors

-a---- 25/06/2025 13:27 67121608 model-00002-of-00002.safetensors

-a---- 25/06/2025 13:27 13660 model.safetensors.index.json

-a---- 25/06/2025 13:27 21959 README.md

-a---- 25/06/2025 13:27 670 special\_tokens\_map.json

-a---- 25/06/2025 13:27 17518497 tokenizer.json

-a---- 25/06/2025 13:27 4241003 tokenizer.model

-a---- 25/06/2025 13:27 35072 tokenizer\_config.json

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b>

Copilot said: Виждам, че в директорията имаш два файла,

Виждам, че в директорията имаш два файла, които звучат като твои Python скриптове:

gemma\_test.py

chat\_appi.py.txt (забележи, че тук името е chat\_appi.py.txt – вероятно трябва да бъде chat\_api.py)

Какво да направиш:

1. Стартирай gemma\_test.py

В твоята директория напиши:

Code

python gemma\_test.py

Това ще стартира тестовия скрипт.

2. Ако искаш да стартираш Flask сървъра

Преименувай файла chat\_appi.py.txt на chat\_api.py:

Можеш да го направиш с:

Code

ren chat\_appi.py.txt chat\_api.py

Или с десен бутон → Rename в Windows Explorer.

След това стартирай:

Code

python chat\_api.py

Важно:

Работи само с файлове, които завършват на .py, не на .py.txt.

Ако редактираш файлове, запази ги с коректно разширение.

Ако получиш грешка или имаш въпроси – копирай съобщението тук и ще ти помогна!

You said: Type "help", "copyright", "credits" or "license" for

Type "help", "copyright", "credits" or "license" for more information.

>>> python chat\_api.py

File "<python-input-0>", line 1

python chat\_api.py

^^^^^^^^

SyntaxError: invalid syntax

>>>

Copilot said: Тази грешка се появява, защото опитваш да

Тази грешка се появява, защото опитваш да стартираш Python скрипт („python chat\_api.py“) вътре в самата Python интерактивна среда (REPL), която започва с >>>.

Командата python chat\_api.py трябва да се изпълнява само в терминал/PowerShell, НЕ в Python REPL.

Какво да направиш:

Излез от Python REPL:

Ако виждаш >>>, напиши:

Code

exit()

и натисни Enter.

Върни се в PowerShell/Command Prompt:

Трябва да виждаш нещо като:

Code

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b>

Там напиши:

Code

python chat\_api.py

или

Code

python gemma\_test.py

(в зависимост кой файл искаш да стартираш)

Никога не използвай командата python ... вътре в REPL (там, където е >>>). В REPL се пише само Python код, не shell команди.

Ако имаш други съобщения за грешки или нещо не се получава — копирай ги тук и ще ти помогна!

You said: Mode

Mode LastWriteTime Length Name

---- ------------- ------ ----

-a---- 25/06/2025 13:27 1657 .gitattributes

-a---- 25/06/2025 14:18 711 chat\_appi.py.txt

-a---- 25/06/2025 13:27 654 config.json

-a---- 25/06/2025 13:36 10031780672 gemma-2b.gguf

-a---- 25/06/2025 14:03 584 gemma\_test.py

-a---- 25/06/2025 13:27 144 generation\_config.json

-a---- 25/06/2025 13:28 4945242264 model-00001-of-00002.safetensors

-a---- 25/06/2025 13:27 67121608 model-00002-of-00002.safetensors

-a---- 25/06/2025 13:27 13660 model.safetensors.index.json

-a---- 25/06/2025 13:27 21959 README.md

-a---- 25/06/2025 13:27 670 special\_tokens\_map.json

-a---- 25/06/2025 13:27 17518497 tokenizer.json

-a---- 25/06/2025 13:27 4241003 tokenizer.model

-a---- 25/06/2025 13:27 35072 tokenizer\_config.json

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> python gemma\_test.py

Script started!

Loading tokenizer...

Traceback (most recent call last):

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 470, in cached\_files

hf\_hub\_download(

~~~~~~~~~~~~~~~^

path\_or\_repo\_id,

^^^^^^^^^^^^^^^^

...<10 lines>...

local\_files\_only=local\_files\_only,

^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

)

^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 106, in \_inner\_fn

validate\_repo\_id(arg\_value)

~~~~~~~~~~~~~~~~^^^^^^^^^^^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 160, in validate\_repo\_id

raise HFValidationError(

...<3 lines>...

)

huggingface\_hub.errors.HFValidationError: Repo id must use alphanumeric chars or '-', '\_', '.', '--' and '..' are forbidden, '-' and '.' cannot start or end the name, max length is 96: './gemma-2b'.

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b\gemma\_test.py", line 7, in <module>

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\models\auto\tokenization\_auto.py", line 950, in from\_pretrained

tokenizer\_config = get\_tokenizer\_config(pretrained\_model\_name\_or\_path, \*\*kwargs)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\models\auto\tokenization\_auto.py", line 782, in get\_tokenizer\_config

resolved\_config\_file = cached\_file(

pretrained\_model\_name\_or\_path,

...<12 lines>...

\_commit\_hash=commit\_hash,

)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 312, in cached\_file

file = cached\_files(path\_or\_repo\_id=path\_or\_repo\_id, filenames=[filename], \*\*kwargs)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 523, in cached\_files

\_get\_cache\_file\_to\_return(path\_or\_repo\_id, filename, cache\_dir, revision) for filename in full\_filenames

~~~~~~~~~~~~~~~~~~~~~~~~~^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 140, in \_get\_cache\_file\_to\_return

resolved\_file = try\_to\_load\_from\_cache(path\_or\_repo\_id, full\_filename, cache\_dir=cache\_dir, revision=revision)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 106, in \_inner\_fn

validate\_repo\_id(arg\_value)

~~~~~~~~~~~~~~~~^^^^^^^^^^^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 160, in validate\_repo\_id

raise HFValidationError(

...<3 lines>...

)

huggingface\_hub.errors.HFValidationError: Repo id must use alphanumeric chars or '-', '\_', '.', '--' and '..' are forbidden, '-' and '.' cannot start or end the name, max length is 96: './gemma-2b'.

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> ren chat\_appi.py.txt chat\_api.py

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> python chat\_api.py

Traceback (most recent call last):

File "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b\chat\_api.py", line 1, in <module>

from flask import Flask, request, jsonify

ModuleNotFoundError: No module named 'flask'

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b>

Copilot said: Имаш две различни грешки, които са много

Имаш две различни грешки, които са много чести при работа с HuggingFace модели и Python библиотеки. Ето какво означават и как да ги оправиш:

1. HFValidationError: Repo id must use alphanumeric chars ... './gemma-2b'

Причина:

Библиотеката transformers очаква като път към модела или име на модел в HuggingFace Hub, а не локална директория с точка и наклонена черта (./gemma-2b).

Решение:

Когато зареждаш модел и токенайзер от локална папка, използвай абсолютен път или просто името на папката, ако вече си в нея.

Ти си в папката:

C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b

Промени в твоя Python код:

Python

model\_path = "." # ако скриптът е в същата папка като модела

или

Python

model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

Примерно начало на скрипта:

gemma\_test.py

v6

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "." # текущата директория, където е моделът

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

2. ModuleNotFoundError: No module named 'flask'

Причина:

Не си инсталирала библиотеката Flask, която е нужна за уеб сървъра.

Решение:

Затвори Python REPL, ако си в него (exit()).

В терминала/PowerShell напиши:

Code

pip install flask

Изчакай да завърши инсталацията.

После пак стартирай:

Code

python chat\_api.py

Резюме

За HuggingFace моделите:

Използвай абсолютен път или . ако си в папката на модела!

За Flask:

Инсталирай Flask с pip install flask.

Ако има нови грешки, копирай съобщението тук! Ще ти помогна пак!

You said: SyntaxError: invalid syntax >>> model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b" >>>

SyntaxError: invalid syntax

>>> model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

>>>

Copilot said: Отново си въвела Python код в интерактивната

Отново си въвела Python код в интерактивната Python среда (REPL), където виждаш >>>.

Това не е грешка за този ред! Редът

Python

model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

е напълно валиден в Python.

Причината за "SyntaxError: invalid syntax" по-горе е, че си опитала да изпълниш shell команда като python chat\_api.py в REPL, което е невъзможно.

Как да работиш правилно

1. Никога не пиши shell команди като python ... в REPL (>>>)!

Всичко, което започва с python ..., се пише в терминала/PowerShell, не в REPL.

2. Редактирай Python файл, не въвеждай ред по ред

Пиши своя код в Notepad, VSCode или друг редактор и го запази като .py файл (например: gemma\_test.py).

3. Стартирай цял Python файл от терминала

Влез в съответната папка, например:

Code

cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

Стартирай файла:

Code

python gemma\_test.py

Накратко

Python код (като model\_path = ...) — в .py файл!

Shell команди (python gemma\_test.py) — в терминал/PowerShell!

Не смесвай двете среди.

Ако имаш нужда от примерен файл, пиши и ще ти дам готов файл за копиране!

Ако има нови съобщения за грешка — копирай ги тук.

You said: PS C:\Users\Stella Barbarella> model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b" model\_path

PS C:\Users\Stella Barbarella> model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

model\_path : Le terme «model\_path» n'est pas reconnu comme nom d'applet de commande, fonction, fichier de script ou

programme exécutable. Vérifiez l'orthographe du nom, ou si un chemin d'accès existe, vérifiez que le chemin d'accès

est correct et réessayez.

Au caractère Ligne:1 : 1

+ model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-f ...

+ ~~~~~~~~~~

+ CategoryInfo : ObjectNotFound: (model\_path:String) [], CommandNotFoundException

+ FullyQualifiedErrorId : CommandNotFoundException

Copilot said: Това съобщение показва, че си въвела Python

Това съобщение показва, че си въвела Python код (model\_path = ...) в PowerShell, където трябва да се пишат само команди за операционната система (като cd, dir, python ...).

Python код се пише само в .py файл или в Python REPL (>>>).

Как да работиш правилно:

1. Редактирай Python кода си в .py файл

Например, отвори Notepad, VSCode или друг редактор и напиши там:

Python

model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

(и останалия код)

2. Стартирай Python файла от PowerShell/Терминал

В PowerShell напиши:

Code

python gemma\_test.py

или

Code

python chat\_api.py

(Името трябва да съответства на твоя .py файл.)

Обобщение:

Python код → само във файл .py или в REPL (>>>)

Команди като python ... → само в терминал/PowerShell

Ако искаш примерен файл, ето:

gemma\_test.py

v7

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

Запази го като gemma\_test.py и го стартирай така:

Code

python gemma\_test.py

Ако имаш нова грешка или нещо не се получава – пиши тук!

You said: PS C:\Users\Stella Barbarella> python gemma\_test.py C:\Python313\python.exe: can't open

PS C:\Users\Stella Barbarella> python gemma\_test.py

C:\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\gemma\_test.py': [Errno 2] No such file or directory

PS C:\Users\Stella Barbarella>

Copilot said: Грешката означава, че файлът `gemma\_test.py` не се намира в

Грешката означава, че файлът gemma\_test.py не се намира в текущата директория (C:\Users\Stella Barbarella).

Твоят файл gemma\_test.py е в:

Code

C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b

Какво да направиш:

Отиди в правилната папка с тази команда:

PowerShell

cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

Увери се, че си в правилната папка!

Можеш да провериш със:

Code

dir

Трябва да виждаш gemma\_test.py сред файловете.

Стартирай скрипта с:

Code

python gemma\_test.py

Ако имаш други файлове, които искаш да стартираш, винаги първо отиди в папката, където се намират!

Ако получиш нова грешка, копирай съобщението тук!

You said: PS C:\Users\Stella Barbarella> cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b" PS C:\Users\Stella

PS C:\Users\Stella Barbarella> cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> dir

Répertoire : C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b

Mode LastWriteTime Length Name

---- ------------- ------ ----

-a---- 25/06/2025 13:27 1657 .gitattributes

-a---- 25/06/2025 14:18 711 chat\_api.py

-a---- 25/06/2025 13:27 654 config.json

-a---- 25/06/2025 13:36 10031780672 gemma-2b.gguf

-a---- 25/06/2025 14:03 584 gemma\_test.py

-a---- 25/06/2025 13:27 144 generation\_config.json

-a---- 25/06/2025 13:28 4945242264 model-00001-of-00002.safetensors

-a---- 25/06/2025 13:27 67121608 model-00002-of-00002.safetensors

-a---- 25/06/2025 13:27 13660 model.safetensors.index.json

-a---- 25/06/2025 13:27 21959 README.md

-a---- 25/06/2025 13:27 670 special\_tokens\_map.json

-a---- 25/06/2025 13:27 17518497 tokenizer.json

-a---- 25/06/2025 13:27 4241003 tokenizer.model

-a---- 25/06/2025 13:27 35072 tokenizer\_config.json

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> python gemma\_test.py

Script started!

Loading tokenizer...

Traceback (most recent call last):

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 470, in cached\_files

hf\_hub\_download(

~~~~~~~~~~~~~~~^

path\_or\_repo\_id,

^^^^^^^^^^^^^^^^

...<10 lines>...

local\_files\_only=local\_files\_only,

^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

)

^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 106, in \_inner\_fn

validate\_repo\_id(arg\_value)

~~~~~~~~~~~~~~~~^^^^^^^^^^^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 160, in validate\_repo\_id

raise HFValidationError(

...<3 lines>...

)

huggingface\_hub.errors.HFValidationError: Repo id must use alphanumeric chars or '-', '\_', '.', '--' and '..' are forbidden, '-' and '.' cannot start or end the name, max length is 96: './gemma-2b'.

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b\gemma\_test.py", line 7, in <module>

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\models\auto\tokenization\_auto.py", line 950, in from\_pretrained

tokenizer\_config = get\_tokenizer\_config(pretrained\_model\_name\_or\_path, \*\*kwargs)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\models\auto\tokenization\_auto.py", line 782, in get\_tokenizer\_config

resolved\_config\_file = cached\_file(

pretrained\_model\_name\_or\_path,

...<12 lines>...

\_commit\_hash=commit\_hash,

)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 312, in cached\_file

file = cached\_files(path\_or\_repo\_id=path\_or\_repo\_id, filenames=[filename], \*\*kwargs)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 523, in cached\_files

\_get\_cache\_file\_to\_return(path\_or\_repo\_id, filename, cache\_dir, revision) for filename in full\_filenames

~~~~~~~~~~~~~~~~~~~~~~~~~^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 140, in \_get\_cache\_file\_to\_return

resolved\_file = try\_to\_load\_from\_cache(path\_or\_repo\_id, full\_filename, cache\_dir=cache\_dir, revision=revision)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 106, in \_inner\_fn

validate\_repo\_id(arg\_value)

~~~~~~~~~~~~~~~~^^^^^^^^^^^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 160, in validate\_repo\_id

raise HFValidationError(

...<3 lines>...

)

huggingface\_hub.errors.HFValidationError: Repo id must use alphanumeric chars or '-', '\_', '.', '--' and '..' are forbidden, '-' and '.' cannot start or end the name, max length is 96: './gemma-2b'.

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b>

Copilot said: Тази грешка се появява, защото HuggingFace Transformers

Тази грешка се появява, защото HuggingFace Transformers очаква пътят към локалния модел да не започва с точка или наклонена черта, когато се намираш вече в самата папка на модела.

В момента ти се намираш в:

Code

C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b

Точно там са и файловете за модела.

Как да оправиш грешката

Вариант 1: Използвай "." като път

В твоя файл gemma\_test.py промени:

Python

model\_path = "./gemma-2b"

на:

Python

model\_path = "."

или още по-добре:

Python

model\_path = os.getcwd()

(но тогава трябва да добавиш най-горе import os)

Вариант 2: Абсолютен път

Може и с абсолютен път:

Python

model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

Примерен работещ файл:

gemma\_test.py

v8

from transformers import AutoModelForCausalLM, AutoTokenizer

import os

model\_path = "." # или os.getcwd() за абсолютния път

print("Script started!")

Резюме:

Ако си в папката на модела — използвай model\_path = "."

Не използвай ./gemma-2b, когато вече си в тази папка!

Запази, стартирай пак с:

Code

python gemma\_test.py

Ако има нова грешка, копирай тук!

You said: PS C:\Users\Stella Barbarella> model\_path = "./gemma-2b" model\_path :

PS C:\Users\Stella Barbarella> model\_path = "./gemma-2b"

model\_path : Le terme «model\_path» n'est pas reconnu comme nom d'applet de commande, fonction, fichier de script ou

programme exécutable. Vérifiez l'orthographe du nom, ou si un chemin d'accès existe, vérifiez que le chemin d'accès

est correct et réessayez.

Au caractère Ligne:1 : 1

+ model\_path = "./gemma-2b"

+ ~~~~~~~~~~

+ CategoryInfo : ObjectNotFound: (model\_path:String) [], CommandNotFoundException

+ FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\Stella Barbarella>

Copilot said: Това съобщение означава, че отново се опитваш

Това съобщение означава, че отново се опитваш да изпълниш Python код (model\_path = "./gemma-2b") директно в PowerShell, а не в Python файл или в Python интерактивна среда (>>>).

Какво трябва да направиш?

Редактирай или създай Python файл (например gemma\_test.py) с твоя код вътре:

Python

model\_path = "./gemma-2b"

# останалият ти Python код...

Запази файла.

Стартирай Python файла от PowerShell така:

Code

python gemma\_test.py

Важно:

Команди от вида python ... се пишат в PowerShell.

Редове като model\_path = ... се пишат само във .py файл или в Python REPL (>>>), не в PowerShell.

Ако искаш примерен файл, кажи – ще ти подготвя!

Ако има нова грешка, копирай я тук!

You said: PS C:\Users\Stella Barbarella> = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b" =

PS C:\Users\Stella Barbarella> = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

= : Le terme «=» n'est pas reconnu comme nom d'applet de commande, fonction, fichier de script ou programme

exécutable. Vérifiez l'orthographe du nom, ou si un chemin d'accès existe, vérifiez que le chemin d'accès est correct

et réessayez.

Au caractère Ligne:1 : 2

+ = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma ...

+ ~

+ CategoryInfo : ObjectNotFound: (=:String) [], CommandNotFoundException

+ FullyQualifiedErrorId : CommandNotFoundException

PS C:\Users\Stella Barbarella> model\_path : Le terme «model\_path» n'est pas reconnu comme nom d'applet de comman

Copilot said: Това съобщение означава, че опитваш да въвеждаш

Това съобщение означава, че опитваш да въвеждаш Python код (например model\_path = ...) директно в PowerShell, което не е правилно.

В PowerShell се пишат само команди за операционната система и за стартиране на Python скриптове!

Какво трябва да правиш?

1. Създай/Редактирай Python файл (например gemma\_test.py):

gemma\_test.py

v9

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

2. Запази файла.

3. Стартирай файла от PowerShell така:

PowerShell

cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

python gemma\_test.py

Никога не пиши Python код (редове с =) в PowerShell!

В PowerShell се пишат само команди за стартиране на скриптове или навигиране по папки.

Python код се пише само във файл .py или в интерактивната среда на Python (която започва с >>>).

Ако има нова грешка – копирай я тук!

You said: S C:\Users\Stella Barbarella> cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b" PS C:\Users\Stella

S C:\Users\Stella Barbarella> cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> python gemma\_test.py

Loading checkpoint shards: 100%|█████████████████████████████████████████████████████████| 2/2 [00:02<00:00, 1.21s/it]

Здравей! Как мога да помогна?

$\displaystyle \frac{1}{2}x^{2} + \frac{1}{3}x + \frac{1}{6} = 0$

$\displaystyle x^{2} + 3x + 3 = 0$

$\displaystyle D = 9 - 12 = -3$

$\displaystyle x\_{1} = \frac{-3 + \sqrt{-3}}{2} = \frac{-3 + \sqrt{3}i

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b>

Copilot said: Чудесно! 🎉 Твоят скрипт \*\*gemma\_test.py\*\* работи правилно и

Чудесно! 🎉

Твоят скрипт gemma\_test.py работи правилно и зарежда модела — виждаш генерирани отговори.

Това означава:

Пътят към модела е коректен.

Всички необходими файлове са налични.

Моделът успешно генерира текст (отговорът ти е на български и съдържа формули).

Какво следва?

Ако искаш да тестваш други подсказки (prompts), промени текста в:

Python

inputs = tokenizer("ТВОЯТ ТЕКСТ ТУК", return\_tensors="pt")

Ако искаш да направиш уеб API (например с Flask), работи по chat\_api.py (инсталирай Flask, ако още не си: pip install flask).

Ако имаш други въпроси или искаш пример за Flask API, кажи!

Можеш спокойно да продължиш с твоя проект. Ако има нови грешки — пиши тук! 😊

You said:

^^^^^^^^

SyntaxError: invalid syntax

>>> model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

>>> inputs = tokenizer("ТВОЯТ ТЕКСТ ТУК", return\_tensors="pt")

Traceback (most recent call last):

File "<python-input-2>", line 1, in <module>

inputs = tokenizer("ТВОЯТ ТЕКСТ ТУК", return\_tensors="pt")

^^^^^^^^^

NameError: name 'tokenizer' is not defined

>>>

Copilot said: Тази грешка означава, че се опитваш да

Тази грешка означава, че се опитваш да ползваш променлива tokenizer, която не е дефинирана в текущата Python REPL сесия (>>>). Това е, защото не си изпълнила кода за създаване на токенайзера (например tokenizer = AutoTokenizer.from\_pretrained(...)).

Как да го оправиш:

Вариант 1: Работи винаги през .py файл!

Това е най-сигурният и правилен начин.

Направи пълен Python файл, например gemma\_test.py:

gemma\_test.py

v10

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

Запази файла и го стартирай от PowerShell така:

Code

python gemma\_test.py

Вариант 2: Ако все пак искаш в REPL (>>>), трябва винаги първо да дефинираш всичко:

Python

from transformers import AutoModelForCausalLM, AutoTokenizer

model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

inputs = tokenizer("ТВОЯТ ТЕКСТ ТУК", return\_tensors="pt")

outputs = model.generate(\*\*inputs, max\_new\_tokens=100)

response = tokenizer.decode(outputs[0], skip\_special\_tokens=True)

print(response)

Забележка: Ако случайно изпуснеш някой ред, ще получаваш грешки като тази.

Препоръка:

Винаги работи с .py файлове и ги стартирай от терминала!

Така всичко ще работи по-надеждно и ще е по-лесно за дебъгване.

Ако имаш нужда от примерен файл или друга помощ — кажи!

You said: :\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\gemma\_test.py': [Errno

:\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\gemma\_test.py': [Errno 2] No such file or directory

PS C:\Users\Stella Barbarella> cd 'C:\\Users\\Stella Barbarella\\gemma\_test.py'

cd : Impossible de trouver le chemin d'accès « C:\Users\Stella Barbarella\gemma\_test.py », car il n'existe pas.

Au caractère Ligne:1 : 1

+ cd 'C:\\Users\\Stella Barbarella\\gemma\_test.py'

+ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

+ CategoryInfo : ObjectNotFound: (C:\Users\Stella Barbarella\gemma\_test.py:String) [Set-Location], ItemNo

tFoundException

+ FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.SetLocationCommand

PS C:\Users\Stella Barbarella>

Copilot said: Виждам, че се опитваш да „влезеш“ във

Виждам, че се опитваш да „влезеш“ във файл с cd 'C:\\Users\\Stella Barbarella\\gemma\_test.py', но cd работи само с папки (директории), не с файлове!

Грешката ти казва, че не може да намери такъв път, защото:

Или файлът не съществува на това място

Или се опитваш да ползваш командата cd за файл, а не за папка

Какво да направиш

1. Открий къде е твоят файл

Твоят файл gemma\_test.py е в:

Code

C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b

(виж предишните ти съобщения с командата dir)

2. Отиди в папката, където е файлът (НЕ във файла!):

PowerShell

cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

3. Стартирай Python файла:

PowerShell

python gemma\_test.py

Важно:

cd се използва само за папки (директории), никога за файлове.

За да стартираш Python файл, трябва да си в папката, където се намира този файл, и да изпълниш python име\_на\_файла.py.

Ако имаш други файлове за стартиране или нова грешка – копирай тук!

You said: ory PS C:\Users\Stella Barbarella> cd 'C:\\Users\\Stella Barbarella\\gemma\_test.py' cd :

ory

PS C:\Users\Stella Barbarella> cd 'C:\\Users\\Stella Barbarella\\gemma\_test.py'

cd : Impossible de trouver le chemin d'accès « C:\Users\Stella Barbarella\gemma\_test.py », car il n'existe pas.

Au caractère Ligne:1 : 1

+ cd 'C:\\Users\\Stella Barbarella\\gemma\_test.py'

+ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

+ CategoryInfo : ObjectNotFound: (C:\Users\Stella Barbarella\gemma\_test.py:String) [Set-Location], ItemNo

tFoundException

+ FullyQualifiedErrorId : PathNotFound,Microsoft.PowerShell.Commands.SetLocationCommand

PS C:\Users\Stella Barbarella> C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b

C:\Users\Stella : Le terme «C:\Users\Stella» n'est pas reconnu comme nom d'applet de commande, fonction, fichier de

script ou programme exécutable. Vérifiez l'orthographe du nom, ou si un chemin d'accès existe, vérifiez que le chemin

d'accès est correct et réessayez.

Au caractère Ligne:1 : 1

+ C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b

+ ~~~~~~~~~~~~~~~

+ CategoryInfo : ObjectNotFound: (C:\Users\Stella:String) [], CommandNotFoundException

+ FullyQualifiedErrorId : CommandNotFoundException

Copilot said: Виждам, че се опитваш да използваш командата

Виждам, че се опитваш да използваш командата cd с път към файл, а също така си опитала да изпълниш директория като команда. Ето как да го направиш правилно:

Как да се придвижиш до папка и да стартираш Python скрипт

1. Отиди в папката с твоя проект

В терминала (PowerShell) напиши:

PowerShell

cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

Важно: Пътят трябва да е в кавички "..." ако има интервали!

2. Виж дали си в правилната папка

Напиши:

PowerShell

dir

Трябва да виждаш файловете като gemma\_test.py, chat\_api.py и други.

3. Стартирай Python скрипта

Напиши:

PowerShell

python gemma\_test.py

или

PowerShell

python chat\_api.py

(в зависимост от това кой файл искаш да стартираш)

Често срещани грешки:

cd към файл:

НЕ трябва да правиш cd 'C:\\Users\\Stella Barbarella\\gemma\_test.py' — това е файл, а не папка.

Изпълнение на директория:

НЕ трябва да пишеш просто път до папка като команда (например C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b). Това ще даде грешка.

Ако има нова грешка или въпрос — копирай тук!

Мога да ти дам и конкретни примери или файлове, ако имаш нужда.

You said: PS C:\Users\Stella Barbarella> python gemma\_test.py C:\Python313\python.exe: can't open

PS C:\Users\Stella Barbarella> python gemma\_test.py

C:\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\gemma\_test.py': [Errno 2] No such file or directory

PS C:\Users\Stella Barbarella> python chat\_api.py

C:\Python313\python.exe: can't open file 'C:\\Users\\Stella Barbarella\\chat\_api.py': [Errno 2] No such file or directory

PS C:\Users\Stella Barbarella>

Copilot said: Виждаш тази грешка, защото се опитваш да

Виждаш тази грешка, защото се опитваш да стартираш файлове (gemma\_test.py или chat\_api.py), които не се намират в текущата ти папка (C:\Users\Stella Barbarella). Затова Python не ги намира.

Как да го оправиш стъпка по стъпка

Отиди в папката, където са файловете ти!

В твоя случай файловете са в:

Code

C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b

В PowerShell напиши:

PowerShell

cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

Провери дали си в правилната папка и дали файловете са там

Изпълни:

PowerShell

dir

Трябва да видиш gemma\_test.py и/или chat\_api.py в списъка.

Стартирай скрипта

PowerShell

python gemma\_test.py

или

PowerShell

python chat\_api.py

Важно да запомниш

cd се използва за навигиране до папката, където се намира файлът.

python ... се използва само когато си в папката с файла.

Ако си в грешна папка, Python не може да намери файла и да го стартира.

Ако имаш нова грешка, просто я копирай тук! Ще ти помогна веднага.

You said: 25/06/2025

25/06/2025 13:27 1657 .gitattributes

-a---- 25/06/2025 14:18 711 chat\_api.py

-a---- 25/06/2025 13:27 654 config.json

-a---- 25/06/2025 13:36 10031780672 gemma-2b.gguf

-a---- 25/06/2025 14:33 501 gemma\_test.py

-a---- 25/06/2025 13:27 144 generation\_config.json

-a---- 25/06/2025 13:28 4945242264 model-00001-of-00002.safetensors

-a---- 25/06/2025 13:27 67121608 model-00002-of-00002.safetensors

-a---- 25/06/2025 13:27 13660 model.safetensors.index.json

-a---- 25/06/2025 13:27 21959 README.md

-a---- 25/06/2025 13:27 670 special\_tokens\_map.json

-a---- 25/06/2025 13:27 17518497 tokenizer.json

-a---- 25/06/2025 13:27 4241003 tokenizer.model

-a---- 25/06/2025 13:27 35072 tokenizer\_config.json

Copilot said: Чудесно! Вече си в правилната папка и

Чудесно! Вече си в правилната папка и виждаш файловете gemma\_test.py и chat\_api.py.

За да стартираш някой от тези скриптове, изпълни:

PowerShell

python gemma\_test.py

или

PowerShell

python chat\_api.py

Важно: Увери се, че пишеш тези команди точно така, в PowerShell, докато си в тази папка.

Ако има нова грешка:

Копирай съобщението тук, за да ти помогна веднага.

Ако искаш да промениш текста, който подаваш към модела:

Редактирай реда в gemma\_test.py:

Python

inputs = tokenizer("ТВОЯТ ТЕКСТ ТУК", return\_tensors="pt")

и пробвай пак.

Ако имаш въпроси за Flask, уеб API или нещо друго — питай смело!

You said: PS C:\Users\Stella Barbarella> cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b" PS C:\Users\Stella

PS C:\Users\Stella Barbarella> cd "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> python gemma\_test.py

Loading checkpoint shards: 100%|█████████████████████████████████████████████████████████| 2/2 [00:02<00:00, 1.21s/it]

Здравей! Как мога да помогна?

$\displaystyle \frac{1}{2}x^{2} + \frac{1}{3}x + \frac{1}{6} = 0$

$\displaystyle x^{2} + 3x + 3 = 0$

$\displaystyle D = 9 - 12 = -3$

$\displaystyle x\_{1} = \frac{-3 + \sqrt{-3}}{2} = \frac{-3 + \sqrt{3}i

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> python gemma\_test.py

Loading checkpoint shards: 100%|█████████████████████████████████████████████████████████| 2/2 [00:02<00:00, 1.22s/it]

Здравей! Как мога да помогна?

$\displaystyle \frac{1}{2}x^{2} + \frac{1}{3}x + \frac{1}{6} = 0$

$\displaystyle x^{2} + 3x + 3 = 0$

$\displaystyle D = 9 - 12 = -3$

$\displaystyle x\_{1} = \frac{-3 + \sqrt{-3}}{2} = \frac{-3 + \sqrt{3}i

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b>

Copilot said: Това означава, че всичко работи успешно! 🎉 Твоят

Това означава, че всичко работи успешно! 🎉

Твоят модел се зарежда и генерира отговори. Примерният отговор е:

Code

Здравей! Как мога да помогна?

$\displaystyle \frac{1}{2}x^{2} + \frac{1}{3}x + \frac{1}{6} = 0$

$\displaystyle x^{2} + 3x + 3 = 0$

$\displaystyle D = 9 - 12 = -3$

$\displaystyle x\_{1} = \frac{-3 + \sqrt{-3}}{2} = \frac{-3 + \sqrt{3}i

Какво означава това?

Моделът зарежда успешно чекпойнта и токенайзера.

Генерира отговор (в случая — текст с формули).

Скриптът работи всеки път, когато го стартираш.

Какво можеш да правиш нататък?

Промени входния текст (prompt):

Отвори gemma\_test.py и смени реда:

Python

inputs = tokenizer("Здравей! Как мога да помогна?", return\_tensors="pt")

с твой въпрос или задача.

Експериментирай с параметрите:

Промени max\_new\_tokens=100 на по-голяма или по-малка стойност.

Създай API (по избор):

Ако искаш да направиш уеб интерфейс (например с Flask), можеш да стартираш или модифицираш chat\_api.py.

Ако искаш примерен Flask API, кажи и ще ти дам работещ шаблон.

Ако имаш друг въпрос, нова цел или проблем — пиши!

Всичко по зареждане на модела вече работи перфектно! 👏

You said: искам да го кача в сайта ми

искам да го кача в сайта ми

Copilot said: Разбира се! За да „качиш“ (интегрираш) твоя

Разбира се! За да „качиш“ (интегрираш) твоя модел или функционалност в уеб сайт, най-често се прави така:

1. Създаваш уеб API (пример с Flask)

Можеш да използваш Python и Flask, за да направиш лесен уеб сървър. Така твоят сайт (или фронтенд) ще прави заявки към този сървър, който ще изпраща отговори от модела.

Примерен Flask API:

chat\_api.py

v3

from flask import Flask, request, jsonify

from transformers import AutoModelForCausalLM, AutoTokenizer

import torch

app = Flask(\_\_name\_\_)

2. Стартираш сървъра

В терминала (в папката с този файл) изпълни:

PowerShell

python chat\_api.py

3. Свързваш сайта с твоя API

На сайта ти (фронтенд) пращаш POST заявка към http://<ip-адрес>:5000/chat с JSON:

JSON

{ "prompt": "Твоят въпрос тук" }

и ще получиш отговор:

JSON

{ "response": "Генерираният текст" }

Пример с JavaScript (fetch):

JavaScript

fetch("http://localhost:5000/chat", {

method: "POST",

headers: { "Content-Type": "application/json" },

body: JSON.stringify({ prompt: "Здравей, модел!" })

})

.then(res => res.json())

.then(data => console.log(data.response));

4. Ако сайтът е публичен:

Ако сайтът ти е онлайн, трябва да настроиш сървъра да е достъпен и отвън, и евентуално да използваш reverse proxy (напр. nginx).

Ако сайтът е на друг компютър – използвай публичния IP на машината със сървъра.

Искаш ли примерен HTML/JS код за формуляр за чат? Или помощ с public hosting?

Кажи какъв точно сайт имаш – статичен HTML, WordPress, React...?

Така ще помогна най-точно!

You said: и да работи от тав(PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b>

и да работи от тав(PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> python chat\_api.py

Traceback (most recent call last):

File "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b\chat\_api.py", line 1, in <module>

from flask import Flask, request, jsonify

ModuleNotFoundError: No module named 'flask'

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b>

Copilot said: Грешката ти означава, че \*\*Flask не е

Грешката ти означава, че Flask не е инсталиран в твоя Python. Това е нужно, за да работи твоят уеб API.

Как да оправиш проблема:

1. Инсталирай Flask

В терминала (PowerShell), напиши:

PowerShell

pip install flask

или (ако имаш няколко версии на Python):

PowerShell

python -m pip install flask

Ако искаш да ползваш и RESTful API разширения:

PowerShell

pip install flask flask-cors

2. Стартирай пак твоя скрипт

PowerShell

python chat\_api.py

3. Ако видиш "Running on http://127.0.0.1:5000/..."

Значи всичко работи! Можеш да достъпваш API-то си от твоя сайт или с Postman, или с браузър (GET заявка до /), или с JavaScript fetch.

Ако има нова грешка – копирай я тук!

Ако искаш примерен HTML код за чат с това API – кажи!

You said: in cached\_files hf\_hub\_download(

in cached\_files

hf\_hub\_download(

~~~~~~~~~~~~~~~^

path\_or\_repo\_id,

^^^^^^^^^^^^^^^^

...<10 lines>...

local\_files\_only=local\_files\_only,

^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

)

^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 106, in \_inner\_fn

validate\_repo\_id(arg\_value)

~~~~~~~~~~~~~~~~^^^^^^^^^^^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 160, in validate\_repo\_id

raise HFValidationError(

...<3 lines>...

)

huggingface\_hub.errors.HFValidationError: Repo id must use alphanumeric chars or '-', '\_', '.', '--' and '..' are forbidden, '-' and '.' cannot start or end the name, max length is 96: './gemma-2b'.

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

File "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b\chat\_api.py", line 7, in <module>

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\models\auto\tokenization\_auto.py", line 950, in from\_pretrained

tokenizer\_config = get\_tokenizer\_config(pretrained\_model\_name\_or\_path, \*\*kwargs)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\models\auto\tokenization\_auto.py", line 782, in get\_tokenizer\_config

resolved\_config\_file = cached\_file(

pretrained\_model\_name\_or\_path,

...<12 lines>...

\_commit\_hash=commit\_hash,

)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 312, in cached\_file

file = cached\_files(path\_or\_repo\_id=path\_or\_repo\_id, filenames=[filename], \*\*kwargs)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 523, in cached\_files

\_get\_cache\_file\_to\_return(path\_or\_repo\_id, filename, cache\_dir, revision) for filename in full\_filenames

~~~~~~~~~~~~~~~~~~~~~~~~~^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\transformers\utils\hub.py", line 140, in \_get\_cache\_file\_to\_return

resolved\_file = try\_to\_load\_from\_cache(path\_or\_repo\_id, full\_filename, cache\_dir=cache\_dir, revision=revision)

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 106, in \_inner\_fn

validate\_repo\_id(arg\_value)

~~~~~~~~~~~~~~~~^^^^^^^^^^^

File "C:\Users\Stella Barbarella\AppData\Roaming\Python\Python313\site-packages\huggingface\_hub\utils\\_validators.py", line 160, in validate\_repo\_id

raise HFValidationError(

...<3 lines>...

)

huggingface\_hub.errors.HFValidationError: Repo id must use alphanumeric chars or '-', '\_', '.', '--' and '..' are forbidden, '-' and '.' cannot start or end the name, max length is 96: './gemma-2b'.

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b>

Copilot said: Тази грешка означава, че Huggingface Transformers/Hub \*\*не

Тази грешка означава, че Huggingface Transformers/Hub не приема пътя до модела във вида ./gemma-2b или вероятно с друга невалидна стойност.

Причина

Huggingface Hub очаква или:

име на модел от hub-а (например "google/gemma-2b"), или

абсолютен или валиден относителен път към локална папка.

Път като ./gemma-2b или такъв с грешни символи (например край с точка или тире, двойни точки и пр.) не се приема.

Как да го оправиш

1. Използвай абсолютен път до модела

Промени:

Python

model\_path = "./gemma-2b"

или нещо подобно

на:

Python

model\_path = "C:/Users/Stella Barbarella/MyProjects/stella-projects-final/gemma-2b"

или (Windows стил, също работи):

Python

model\_path = r"C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

2. Провери дали използваш правилната променлива навсякъде

Във всички извиквания за зареждане на модел/токенайзер трябва да използваш този абсолютен път:

Python

tokenizer = AutoTokenizer.from\_pretrained(model\_path)

model = AutoModelForCausalLM.from\_pretrained(model\_path)

3. Не използвай ./ или ../ в пътя, ако има проблеми!

Ако скриптът ти е в друга папка, абсолютният път е най-сигурен.

Пример – как трябва да изглежда:

chat\_api.py

v4

from flask import Flask, request, jsonify

from transformers import AutoModelForCausalLM, AutoTokenizer

import torch

app = Flask(\_\_name\_\_)

Рестартирай скрипта след тази промяна!

Ако има нова грешка – копирай я тук!

You said: дай рестарт

дай рестарт

Copilot said: За да рестартираш Flask сървъра (или скрипта

За да рестартираш Flask сървъра (или скрипта си), направи следното:

1. В терминала (PowerShell) натисни Ctrl + C

Това ще спре текущия работещ Python сървър.

2. Стартирай отново с:

PowerShell

python chat\_api.py

Ако имаш нужда от автоматичен рестарт при промяна на кода, можеш да стартираш Flask с debug режим. Промени последния ред така:

Python

app.run(host="0.0.0.0", port=5000, debug=True)

Така Flask ще се рестартира автоматично при всяка промяна на файла.

Готово! Ако имаш грешка или искаш примерен код за автоматичен рестарт – кажи!

You said: den, '-' and '.' cannot start or

den, '-' and '.' cannot start or end the name, max length is 96: './gemma-2b'.

PS C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b> python chat\_api.py

Loading checkpoint shards: 100%|█████████████████████████████████████████████████████████| 2/2 [00:04<00:00, 2.25s/it]

\* Serving Flask app 'chat\_api'

\* Debug mode: off

WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.

\* Running on all addresses (0.0.0.0)

\* Running on http://127.0.0.1:5000

\* Running on http://192.168.1.32:5000

Press CTRL+C to quit

Copilot said: Чудесно! Твоят Flask сървър вече работи успешно!

Чудесно! Твоят Flask сървър вече работи успешно! 🎉

Виждаш тези редове:

Code

\* Running on all addresses (0.0.0.0)

\* Running on http://127.0.0.1:5000

\* Running on http://192.168.1.32:5000

Press CTRL+C to quit

Това означава:

API-то ти е достъпно на адресите:

http://127.0.0.1:5000 (локално само за твоя компютър)

http://192.168.1.32:5000 (от други устройства в твоята местна мрежа)

Какво можеш да правиш сега?

1. Тествай API-то от браузър или Postman

Отвори браузър и посети:

http://127.0.0.1:5000

(Ще видиш съобщение „Not Found“ – това е нормално, защото API endpoint-ът ти е POST към /chat.)

2. Тествай с Postman, Insomnia, или curl

Изпрати POST заявка към:

http://127.0.0.1:5000/chat

с body (JSON):

JSON

{ "prompt": "Здравей, модел!" }

3. Свържи сайта си (front-end)

Използвай JavaScript (fetch/AJAX) за да пратиш заявка към този адрес.

Пример с HTML+JS:

HTML

<form id="chatForm">

<input type="text" id="prompt" placeholder="Въведи въпрос..." />

<button type="submit">Прати</button>

</form>

<div id="response"></div>

<script>

document.getElementById('chatForm').onsubmit = async function(e) {

e.preventDefault();

const prompt = document.getElementById('prompt').value;

const res = await fetch('http://127.0.0.1:5000/chat', {

method: 'POST',

headers: { 'Content-Type': 'application/json' },

body: JSON.stringify({ prompt })

});

const data = await res.json();

document.getElementById('response').innerText = data.response;

};

</script>

Ако искаш достъп от друг компютър (в LAN)

Отвори: http://192.168.1.32:5000/chat от друг компютър в твоята мрежа.

Ако не работи: изключи защитната стена за порт 5000 или разреши входящи връзки.

Ако искаш примерен сайт, имаш грешка или въпрос – кажи!

Uploaded image: image.png

You said: Describe this image

Describe this image

Copilot said: На изображението се вижда браузър, отворил адрес

На изображението се вижда браузър, отворил адрес http://192.168.1.32:5000, и получената страница е с грешка 404 и съобщение на български:

"Не е намерено"

Текстът гласи:

Заявеният URL адрес не е намерен на сървъра. Ако сте въвели URL адреса ръчно, моля, проверете правописа си и опитайте отново.

Това означава, че няма дефиниран "route" за началната страница ("/") във Flask приложението ти. Това е стандартната 404 страница, когато посетиш адрес, за който няма обработка в Python кода.

Решение:

За да получиш резултат, използвай крайна точка (endpoint), която съществува — например /chat с POST заявка. Ако искаш да има нещо и на "/", трябва да добавиш route за това във Flask приложението си.

Ако имаш нужда от примерен код или искаш да направим начална страница — кажи!

Uploaded image: image.png

You said: Describe this image

Describe this image

Copilot said: На изображението се вижда браузър, отворил адрес

На изображението се вижда браузър, отворил адрес http://127.0.0.1:5000. Показва се съобщение "Not Found" с текст:

The requested URL was not found on the server. If you entered the URL manually please check your spelling and try again.

Това означава, че Flask сървърът работи, но няма дефиниран маршрут (route) за началната страница ("/"). Това е стандартната 404 грешка на Flask, когато няма обработка за даден URL.

Извод:

Сървърът работи правилно, просто трябва да ползваш дефинирания endpoint (например /chat с POST заявка), а не началната страница. Ако искаш да има специална начална страница, трябва да добавиш нов route във Flask приложението си.

You said: дай пълния пъп..хъг

дай пълния пъп..хъг

Copilot said: Ето пълния примерен Python/Flask API, който зарежда

Ето пълния примерен Python/Flask API, който зарежда локален HuggingFace модел (от папка, без изтегляне от интернет), връща отговор от модела и е готов за интеграция с front-end.

chat\_api.py

v5

from flask import Flask, request, jsonify

from transformers import AutoModelForCausalLM, AutoTokenizer

import torch

app = Flask(\_\_name\_\_)

Как да го ползваш

Увери се, че всички нужни библиотеки са инсталирани:

PowerShell

pip install flask torch transformers

Сложи този файл в папката на модела.

Стартирай:

PowerShell

python chat\_api.py

Изпращай POST заявки до http://127.0.0.1:5000/chat с JSON като:

JSON

{ "prompt": "Твоят въпрос..." }

Ако отвориш http://127.0.0.1:5000 ще видиш потвърждение, че API-то работи.

Ако искаш примерен front-end или имаш грешка – пиши!

You said: дай пълния адресне работи

дай пълния адресне работи

Copilot said: \*\*Ето пълния път, който трябва да използваш

Ето пълния път, който трябва да използваш в Python кода си, за да работи локалното зареждане на модела:

Python

model\_path = r"C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

Провери, че в chat\_api.py имаш точно този ред:

Python

MODEL\_PATH = r"C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

tokenizer = AutoTokenizer.from\_pretrained(MODEL\_PATH, local\_files\_only=True)

model = AutoModelForCausalLM.from\_pretrained(MODEL\_PATH, local\_files\_only=True)

Защо така?

Абсолютният път гарантира, че HuggingFace ще намери папката ти, независимо къде стартираш скрипта.

local\_files\_only=True гарантира, че няма да опитва да тегли от интернет.

Ако все още не работи:

Увери се, че в тази папка има файлове като:

config.json, tokenizer.json, model-00001-of-00002.safetensors, и т.н.

Увери се, че нямаш излишни точки, тирета или невалидни символи в името на папката.

Стартирай скрипта отново.

Ако пак има грешка, копирай точния текст на грешката тук, за да помогна!

Ако искаш – дай и съдържанието на началните редове на твоя chat\_api.py.

You said: { "architectures": [

{

"architectures": [

"GemmaForCausalLM"

],

"attention\_bias": false,

"attention\_dropout": 0.0,

"bos\_token\_id": 2,

"eos\_token\_id": 1,

"head\_dim": 256,

"hidden\_act": "gelu",

"hidden\_size": 2048,

"initializer\_range": 0.02,

"intermediate\_size": 16384,

"max\_position\_embeddings": 8192,

"model\_type": "gemma",

"num\_attention\_heads": 8,

"num\_hidden\_layers": 18,

"num\_key\_value\_heads": 1,

"pad\_token\_id": 0,

"rms\_norm\_eps": 1e-06,

"rope\_scaling": null,

"rope\_theta": 10000.0,

"torch\_dtype": "bfloat16",

"transformers\_version": "4.38.0.dev0",

"use\_cache": true,

"vocab\_size": 256000

}

Copilot said: Това е съдържанието на твоя `config.json` файл

Това е съдържанието на твоя config.json файл за модела Gemma-2B.

Този файл изглежда коректно и е разпознаваем от HuggingFace Transformers за локално зареждане на модела.

Важно:

За да работи зареждането с HuggingFace, в папката "C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b" трябва да има още няколко файла, например:

tokenizer.json

tokenizer\_config.json

model-00001-of-00002.safetensors (или подобно име)

pytorch\_model.bin (ако не е safetensors)

special\_tokens\_map.json

Ако искаш да заредиш този модел с Flask API, твоята начална част на кода трябва да изглежда така:

chat\_api.py

v6

from flask import Flask, request, jsonify

from transformers import AutoModelForCausalLM, AutoTokenizer

import torch

app = Flask(\_\_name\_\_)

Ако все още не работи:

Провери дали всички нужни файлове са в папката!

Ако има грешка – копирай я буквално тук и изпрати съдържанието на директорията (можеш да напишеш dir в PowerShell и да го копираш).

Готов съм за следваща стъпка или помощ по интеграцията!

Uploaded image: image.png

You said: Describe this image

Describe this image

Copilot said: На изображението се вижда браузър (Microsoft Edge

На изображението се вижда браузър (Microsoft Edge или подобен), отворил адрес http://127.0.0.1:5000. Лявата част на екрана е DevTools (инструментите за разработчици), а вдясно се вижда стандартна страница с надпис "Not Found".

В конзолата на DevTools има съобщение:

Code

Failed to load resource: the server responded with a status of 404 (NOT FOUND) 127.0.0.1/:1

Това означава, че браузърът е направил GET заявка към началната страница (/), но Flask сървърът не връща съдържание за този адрес и връща 404 грешка.

Обобщение:

Сървърът работи, но няма дефиниран маршрут за началната страница.

Това е нормално поведение, ако не си добавил route за '/' във Flask приложението си.

Всякакви опити за достъп до / ще връщат тази страница, докато не добавиш такъв маршрут.

Решение:

Ако искаш да има начална страница, трябва да добавиш във Flask:

Python

@app.route("/", methods=["GET"])

def index():

return "<h2>🟢 Gemma-2b API работи! Изпрати POST към /chat</h2>"

За да използваш модела, пращай POST заявка към /chat (например с Postman, curl или JavaScript).

You said: ... from transformers import AutoModelForCausalLM, AutoTokenizer ... ... app

... from transformers import AutoModelForCausalLM, AutoTokenizer

...

... app = Flask(\_\_name\_\_)

...

... model\_path = "./gemma-2b"

... tokenizer = AutoTokenizer.from\_pretrained(model\_path)

... model = AutoModelForCausalLM.from\_pretrained(model\_path)

...

... @app.route("/generate", methods=["POST"])

... def generate():

... prompt = request.json.get("prompt", "Здравей! Как мога да помогна?")

... inputs = tokenizer(prompt, return\_tensors="pt")

... outputs = model.generate(\*\*inputs, max\_new\_tokens=100)

... response = tokenizer.decode(outputs[0], skip\_special\_tokens=True)

... return jsonify({"response": response})

...

... if \_\_name\_\_ == "\_\_main\_\_":

... app.run(port=5000)

... @app.route("/", methods=["GET"])

... def index():

... return "<h2>?? Gemma-2b API работи! Изпрати POST към /chat</h2>"

...

Copilot said: В кода ти има няколко проблема: 1. \*\*Route

В кода ти има няколко проблема:

Route за "/" е поставен вътре в блока if \_\_name\_\_ == "\_\_main\_\_":, което не е правилно.

Има route /generate, но не и /chat, както се казва в съобщението за началната страница.

Route-ът за "/" трябва да е над if \_\_name\_\_ == "\_\_main\_\_":.

Пътят към модела (model\_path) е относителен. Препоръчвам абсолютен път за по-малко проблеми!

Ето коригиран и препоръчителен вариант на твоя Flask API:

chat\_api.py

v7

from flask import Flask, request, jsonify

from transformers import AutoModelForCausalLM, AutoTokenizer

app = Flask(\_\_name\_\_)

# За по-сигурно използвай абсолютен път!

Как да тестваш

Начална страница:

Отвори в браузър: http://127.0.0.1:5000/

Генериране:

Изпрати POST заявка до http://127.0.0.1:5000/generate

с JSON:

JSON

{ "prompt": "Здравей, модел!" }

Ако имаш нужда от примерен front-end или има грешка – пиши!

You said: Not Found The requested URL was not found

Not Found

The requested URL was not found on the server. If you entered the URL manually please check your spelling and try again.

Copilot said: Това съобщение означава, че си опитал/а да

Това съобщение означава, че си опитал/а да отвориш адрес (URL) на твоя Flask сървър, за който няма дефиниран маршрут (route) в Python кода.

Най-чести причини:

Опитваш да отвориш началната страница /, а в кода няма @app.route("/").

Опитваш да достъпиш грешен endpoint (например /chat вместо /generate).

Как да го оправиш

Провери какви маршрути (routes) имаш в кода! Например, ако имаш само:

Python

@app.route("/generate", methods=["POST"])

def generate():

...

Тогава само /generate работи (с POST заявка)!

За да имаш начална страница, добави:

Python

@app.route("/", methods=["GET"])

def index():

return "<h2>🟢 Gemma-2b API работи! Изпрати POST към /generate</h2>"

Използвай POST до правилния endpoint! За да тестваш генериране, използвай Postman или curl с:

Code

POST http://127.0.0.1:5000/generate

Content-Type: application/json

{ "prompt": "Здравей, модел!" }

Примерен работещ Flask API

chat\_api.py

v8

from flask import Flask, request, jsonify

from transformers import AutoModelForCausalLM, AutoTokenizer

app = Flask(\_\_name\_\_)

model\_path = r"C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

Ако все още имаш проблем – кажи точно кой адрес отваряш и какъв е твоят код!

You said: http://127.0.0.1:5000... @app.route("/generate", methods=["POST"]) ... def generate(): ...

http://127.0.0.1:5000... @app.route("/generate", methods=["POST"])

... def generate():

... prompt = request.json.get("prompt", "Здравей! Как мога да помогна?")

... inputs = tokenizer(prompt, return\_tensors="pt")

... outputs = model.generate(\*\*inputs, max\_new\_tokens=100)

... response = tokenizer.decode(outputs[0], skip\_special\_tokens=True)

... return jsonify({"response": response})

...

... if \_\_name\_\_ == "\_\_main\_\_":

... app.run(port=5000)

... @app.route("/", methods=["GET"])

... def index():

... return "<h2>?? Gemma-2b API работи! Изпрати POST към /chat</h2>"

... @app.route("/generate", methods=["POST"])

... def generate():

... .../

Copilot said: В кода ти има няколко структурни грешки,

В кода ти има няколко структурни грешки, които пречат на Flask да регистрира правилно маршрутите (routes).

Ето какво трябва да поправиш:

1. Всички @app.route... трябва да са извън блока if \_\_name\_\_ == "\_\_main\_\_":

Това гарантира, че Flask "вижда" всички маршрути, когато стартира сървъра.

2. Не трябва да има дублиране на функции с едно и също име (generate).

3. Трябва да имаш само един @app.route("/generate", methods=["POST"]) и един @app.route("/", methods=["GET"]).

Ето коригиран и работещ пример:

chat\_api.py

v9

from flask import Flask, request, jsonify

from transformers import AutoModelForCausalLM, AutoTokenizer

app = Flask(\_\_name\_\_)

# Използвай абсолютен път за модела!

Как да тестваш

Начална страница:

http://127.0.0.1:5000/ — ще видиш потвърждение, че API-то работи.

Генериране:

Изпрати POST заявка до http://127.0.0.1:5000/generate с body:

JSON

{ "prompt": "Напиши нещо" }

Ако пак има проблем — копирай точната грешка тук!

You said: from flask import Flask, request, jsonify from transformers

from flask import Flask, request, jsonify

from transformers import AutoModelForCausalLM, AutoTokenizer

app = Flask(\_\_name\_\_)

# Използвай абсолютен път за модела!

model\_path = r"C:\Users\Stella Barbarella\MyProjects\stella-projects-final\gemma-2b"

tokenizer = AutoTokenizer.from\_pretrained(model\_path, local\_files\_only=True)

model = AutoModelForCausalLM.from\_pretrained(model\_path, local\_files\_only=True)

@app.route("/generate", methods=["POST"])

def generate():

prompt = request.json.get("prompt", "Здравей! Как мога да помогна?")

inputs = tokenizer(prompt, return\_tensors="pt")

outputs = model.generate(\*\*inputs, max\_new\_tokens=100)

response = tokenizer.decode(outputs[0], skip\_special\_tokens=True)

return jsonify({"response": response})

@app.route("/", methods=["GET"])

def index():

return "<h2>🟢 Gemma-2b API работи! Изпрати POST към /generate</h2>"

if \_\_name\_\_ == "\_\_main\_\_":

app.run(host="0.0.0.0", port=5000)

Copilot said: Твоят код е напълно коректен! С

Твоят код е напълно коректен!

С това Flask API можеш да:

Виждаш потвърждение при отваряне на http://127.0.0.1:5000/ (или http://localhost:5000/):

🟢 Gemma-2b API работи! Изпрати POST към /generate

Изпращаш POST заявка към http://127.0.0.1:5000/generate с JSON:

JSON

{ "prompt": "Твоят въпрос тук" }

и получаваш отговор от модела.

Ако видиш "Not Found"

Това ще стане само ако отвориш несъществуващ адрес (например http://127.0.0.1:5000/chat).

За началната страница / и за /generate (POST), всичко ще работи.

Как да тестваш POST заявка

С Postman или Insomnia:

Метод: POST

URL: http://127.0.0.1:5000/generate

Body → raw → JSON:

JSON

{ "prompt": "Колко е 2+2?" }

С curl (от терминал):

sh

curl -X POST http://127.0.0.1:5000/generate -H "Content-Type: application/json" -d "{\"prompt\": \"Колко е 2+2?\"}"

Ако имаш грешка или въпрос – копирай тук цялото съобщение от терминала или браузъра!

Този код работи и е добър старт за твоя API 🎉

Uploaded image: image.png

You said: виж

виж

Copilot said: На изображението се вижда отвореният файл `config.json`

На изображението се вижда отвореният файл config.json в папката на модела gemma-2b. Файлът съдържа основната конфигурация за HuggingFace модел (архитектура, размери, hyperparameters и др.) и допълнително — полета като "prompt", "inputs", "parameters".

Важно:

Стандартният config.json за модел НЕ трябва да съдържа полета като "prompt", "inputs" или "parameters".

Тези полета не са част от модела и могат да причинят проблеми при зареждане с transformers.

Правилно съдържание: Първата част на файла (до "vocab\_size": 256000) е стандартна и правилна.

JSON

{

"architectures": ["GemmaForCausalLM"],

"attention\_bias": false,

"attention\_dropout": 0.0,

"bos\_token\_id": 2,

"eos\_token\_id": 1,

"head\_dim": 256,

"hidden\_act": "gelu",

"hidden\_size": 2048,

"initializer\_range": 0.02,

"intermediate\_size": 16384,

"max\_position\_embeddings": 8192,

"model\_type": "gemma",

"num\_attention\_heads": 8,

"num\_hidden\_layers": 18,

"num\_key\_value\_heads": 1,

"pad\_token\_id": 0,

"rms\_norm\_eps": 1e-06,

"rope\_scaling": null,

"rope\_theta": 10000.0,

"torch\_dtype": "bfloat16",

"transformers\_version": "4.38.0.dev0",

"use\_cache": true,

"vocab\_size": 256000

}

Грешните части са:

JSON

"prompt": "Колко е 2+2?",

"inputs": "Колко е 2+2?",

"parameters": {

"max\_new\_tokens": 50,

"temperature": 0.7,

"top\_p": 0.9,

"top\_k": 50,

"repetition\_penalty": 1.1

}

Те трябва да бъдат премахнати от config.json!

Как да поправиш:

Остави само параметрите, които са за структурата на модела (от "architectures" до "vocab\_size").

Изтрий всичко след последното валидно поле, така че файлът да завършва с:

JSON

"vocab\_size": 256000

}

Пояснение:

Променливите prompt, inputs, parameters трябва да се използват в Python кода ти, а не вътре в модела!

Пример за коректен config.json:

JSON

{

"architectures": ["GemmaForCausalLM"],

"attention\_bias": false,

"attention\_dropout": 0.0,

"bos\_token\_id": 2,

"eos\_token\_id": 1,

"head\_dim": 256,

"hidden\_act": "gelu",

"hidden\_size": 2048,

"initializer\_range": 0.02,

"intermediate\_size": 16384,

"max\_position\_embeddings": 8192,

"model\_type": "gemma",

"num\_attention\_heads": 8,

"num\_hidden\_layers": 18,

"num\_key\_value\_heads": 1,

"pad\_token\_id": 0,

"rms\_norm\_eps": 1e-06,

"rope\_scaling": null,

"rope\_theta": 10000.0,

"torch\_dtype": "bfloat16",

"transformers\_version": "4.38.0.dev0",

"use\_cache": true,

"vocab\_size": 256000

}

Ако имаш нужда от помощ за автоматично почистване на файла — кажи!

You said: как да продължа

как да продължа

Error

Access denied. You do not have permission to view this.