**⚙️ 2. Backend Setup (FastAPI + BLIP)**

mkdir ImageCaptioner && cd ImageCaptioner

mkdir backend && cd backend

python -m venv venv

venv\Scripts\activate # On Windows

PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner> mkdir backend

Directory: C:\Users\pc\OneDrive\Desktop\ImageCaptioner

Mode LastWriteTime Length Name

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

d----- 6/23/2025 6:47 AM backend

PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner> cd backend

PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\backend> python -m venv venv

PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\backend> venv\Scripts\activate # On Windows

(venv) PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\backend>

**✅ B. Install Dependencies**

pip install fastapi uvicorn python-multipart transformers torch pillow

(venv) PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\backend> pip install fastapi uvicorn python-multipart transformers torch pillow

Collecting fastapi

Using cached fastapi-0.115.13-py3-none-any.whl.metadata (27 kB)

Collecting uvicorn

Using cached uvicorn-0.34.3-py3-none-any.whl.metadata (6.5 kB)

Collecting python-multipart

Using cached python\_multipart-0.0.20-py3-none-any.whl.metadata (1.8 kB)

Collecting transformers

Using cached transformers-4.52.4-py3-none-any.whl.metadata (38 kB)

Collecting torch

Using cached torch-2.7.1-cp313-cp313-win\_amd64.whl.metadata (28 kB)

Collecting pillow

Using cached pillow-11.2.1-cp313-cp313-win\_amd64.whl.metadata (9.1 kB)

Collecting starlette<0.47.0,>=0.40.0 (from fastapi)

Using cached starlette-0.46.2-py3-none-any.whl.metadata (6.2 kB)

Collecting pydantic!=1.8,!=1.8.1,!=2.0.0,!=2.0.1,!=2.1.0,<3.0.0,>=1.7.4 (from fastapi)

Using cached pydantic-2.11.7-py3-none-any.whl.metadata (67 kB)

Collecting typing-extensions>=4.8.0 (from fastapi)

Using cached typing\_extensions-4.14.0-py3-none-any.whl.metadata (3.0 kB)

Collecting click>=7.0 (from uvicorn)

Using cached click-8.2.1-py3-none-any.whl.metadata (2.5 kB)

Collecting h11>=0.8 (from uvicorn)

Using cached h11-0.16.0-py3-none-any.whl.metadata (8.3 kB)

Collecting filelock (from transformers)

Using cached filelock-3.18.0-py3-none-any.whl.metadata (2.9 kB)

Collecting huggingface-hub<1.0,>=0.30.0 (from transformers)

Using cached huggingface\_hub-0.33.0-py3-none-any.whl.metadata (14 kB)

Collecting numpy>=1.17 (from transformers)

Using cached numpy-2.3.1-cp313-cp313-win\_amd64.whl.metadata (60 kB)

Collecting packaging>=20.0 (from transformers)

Using cached packaging-25.0-py3-none-any.whl.metadata (3.3 kB)

Collecting pyyaml>=5.1 (from transformers)

Using cached PyYAML-6.0.2-cp313-cp313-win\_amd64.whl.metadata (2.1 kB)

Collecting regex!=2019.12.17 (from transformers)

Using cached regex-2024.11.6-cp313-cp313-win\_amd64.whl.metadata (41 kB)

Collecting requests (from transformers)

Using cached requests-2.32.4-py3-none-any.whl.metadata (4.9 kB)

Collecting tokenizers<0.22,>=0.21 (from transformers)

Using cached tokenizers-0.21.1-cp39-abi3-win\_amd64.whl.metadata (6.9 kB)

Collecting safetensors>=0.4.3 (from transformers)

Using cached safetensors-0.5.3-cp38-abi3-win\_amd64.whl.metadata (3.9 kB)

Collecting tqdm>=4.27 (from transformers)

Using cached tqdm-4.67.1-py3-none-any.whl.metadata (57 kB)

Collecting sympy>=1.13.3 (from torch)

Using cached sympy-1.14.0-py3-none-any.whl.metadata (12 kB)

Collecting networkx (from torch)

Using cached networkx-3.5-py3-none-any.whl.metadata (6.3 kB)

Collecting jinja2 (from torch)

Using cached jinja2-3.1.6-py3-none-any.whl.metadata (2.9 kB)

Collecting fsspec (from torch)

Using cached fsspec-2025.5.1-py3-none-any.whl.metadata (11 kB)

Collecting setuptools (from torch)

Using cached setuptools-80.9.0-py3-none-any.whl.metadata (6.6 kB)

Collecting colorama (from click>=7.0->uvicorn)

Using cached colorama-0.4.6-py2.py3-none-any.whl.metadata (17 kB)

Collecting annotated-types>=0.6.0 (from pydantic!=1.8,!=1.8.1,!=2.0.0,!=2.0.1,!=2.1.0,<3.0.0,>=1.7.4->fastapi)

Using cached annotated\_types-0.7.0-py3-none-any.whl.metadata (15 kB)

Collecting pydantic-core==2.33.2 (from pydantic!=1.8,!=1.8.1,!=2.0.0,!=2.0.1,!=2.1.0,<3.0.0,>=1.7.4->fastapi)

Using cached pydantic\_core-2.33.2-cp313-cp313-win\_amd64.whl.metadata (6.9 kB)

Collecting typing-inspection>=0.4.0 (from pydantic!=1.8,!=1.8.1,!=2.0.0,!=2.0.1,!=2.1.0,<3.0.0,>=1.7.4->fastapi)

Using cached typing\_inspection-0.4.1-py3-none-any.whl.metadata (2.6 kB)

Collecting anyio<5,>=3.6.2 (from starlette<0.47.0,>=0.40.0->fastapi)

Using cached anyio-4.9.0-py3-none-any.whl.metadata (4.7 kB)

Collecting mpmath<1.4,>=1.1.0 (from sympy>=1.13.3->torch)

Using cached mpmath-1.3.0-py3-none-any.whl.metadata (8.6 kB)

Collecting MarkupSafe>=2.0 (from jinja2->torch)

Using cached MarkupSafe-3.0.2-cp313-cp313-win\_amd64.whl.metadata (4.1 kB)

Collecting charset\_normalizer<4,>=2 (from requests->transformers)

Using cached charset\_normalizer-3.4.2-cp313-cp313-win\_amd64.whl.metadata (36 kB)

Collecting idna<4,>=2.5 (from requests->transformers)

Using cached idna-3.10-py3-none-any.whl.metadata (10 kB)

Collecting urllib3<3,>=1.21.1 (from requests->transformers)

Using cached urllib3-2.5.0-py3-none-any.whl.metadata (6.5 kB)

Collecting certifi>=2017.4.17 (from requests->transformers)

Using cached certifi-2025.6.15-py3-none-any.whl.metadata (2.4 kB)

Collecting sniffio>=1.1 (from anyio<5,>=3.6.2->starlette<0.47.0,>=0.40.0->fastapi)

Using cached sniffio-1.3.1-py3-none-any.whl.metadata (3.9 kB)

Using cached fastapi-0.115.13-py3-none-any.whl (95 kB)

Using cached uvicorn-0.34.3-py3-none-any.whl (62 kB)

Using cached python\_multipart-0.0.20-py3-none-any.whl (24 kB)

Using cached transformers-4.52.4-py3-none-any.whl (10.5 MB)

Using cached torch-2.7.1-cp313-cp313-win\_amd64.whl (216.1 MB)

Using cached pillow-11.2.1-cp313-cp313-win\_amd64.whl (2.7 MB)

Using cached click-8.2.1-py3-none-any.whl (102 kB)

Using cached h11-0.16.0-py3-none-any.whl (37 kB)

Using cached huggingface\_hub-0.33.0-py3-none-any.whl (514 kB)

Using cached fsspec-2025.5.1-py3-none-any.whl (199 kB)

Using cached numpy-2.3.1-cp313-cp313-win\_amd64.whl (12.7 MB)

Using cached packaging-25.0-py3-none-any.whl (66 kB)

Using cached pydantic-2.11.7-py3-none-any.whl (444 kB)

Using cached pydantic\_core-2.33.2-cp313-cp313-win\_amd64.whl (2.0 MB)

Using cached PyYAML-6.0.2-cp313-cp313-win\_amd64.whl (156 kB)

Using cached regex-2024.11.6-cp313-cp313-win\_amd64.whl (273 kB)

Using cached safetensors-0.5.3-cp38-abi3-win\_amd64.whl (308 kB)

Using cached starlette-0.46.2-py3-none-any.whl (72 kB)

Using cached sympy-1.14.0-py3-none-any.whl (6.3 MB)

Using cached tokenizers-0.21.1-cp39-abi3-win\_amd64.whl (2.4 MB)

Using cached tqdm-4.67.1-py3-none-any.whl (78 kB)

Using cached typing\_extensions-4.14.0-py3-none-any.whl (43 kB)

Using cached filelock-3.18.0-py3-none-any.whl (16 kB)

Using cached jinja2-3.1.6-py3-none-any.whl (134 kB)

Using cached networkx-3.5-py3-none-any.whl (2.0 MB)

Using cached requests-2.32.4-py3-none-any.whl (64 kB)

Using cached setuptools-80.9.0-py3-none-any.whl (1.2 MB)

Using cached annotated\_types-0.7.0-py3-none-any.whl (13 kB)

Using cached anyio-4.9.0-py3-none-any.whl (100 kB)

Using cached certifi-2025.6.15-py3-none-any.whl (157 kB)

Using cached charset\_normalizer-3.4.2-cp313-cp313-win\_amd64.whl (105 kB)

Using cached idna-3.10-py3-none-any.whl (70 kB)

Using cached MarkupSafe-3.0.2-cp313-cp313-win\_amd64.whl (15 kB)

Using cached mpmath-1.3.0-py3-none-any.whl (536 kB)

Using cached typing\_inspection-0.4.1-py3-none-any.whl (14 kB)

Using cached urllib3-2.5.0-py3-none-any.whl (129 kB)

Using cached colorama-0.4.6-py2.py3-none-any.whl (25 kB)

Using cached sniffio-1.3.1-py3-none-any.whl (10 kB)

Installing collected packages: mpmath, urllib3, typing-extensions, sympy, sniffio, setuptools, safetensors, regex, pyyaml, python-multipart, pillow, packaging, numpy, networkx, MarkupSafe, idna, h11, fsspec, filelock, colorama, charset\_normalizer, certifi, annotated-types, typing-inspection, tqdm, requests, pydantic-core, jinja2, click, anyio, uvicorn, torch, starlette, pydantic, huggingface-hub, tokenizers, fastapi, transformers

Successfully installed MarkupSafe-3.0.2 annotated-types-0.7.0 anyio-4.9.0 certifi-2025.6.15 charset\_normalizer-3.4.2 click-8.2.1 colorama-0.4.6 fastapi-0.115.13 filelock-3.18.0 fsspec-2025.5.1 h11-0.16.0 huggingface-hub-0.33.0 idna-3.10 jinja2-3.1.6 mpmath-1.3.0 networkx-3.5 numpy-2.3.1 packaging-25.0 pillow-11.2.1 pydantic-2.11.7 pydantic-core-2.33.2 python-multipart-0.0.20 pyyaml-6.0.2 regex-2024.11.6 requests-2.32.4 safetensors-0.5.3 setuptools-80.9.0 sniffio-1.3.1 starlette-0.46.2 sympy-1.14.0 tokenizers-0.21.1 torch-2.7.1 tqdm-4.67.1 transformers-4.52.4 typing-extensions-4.14.0 typing-inspection-0.4.1 urllib3-2.5.0 uvicorn-0.34.3

[notice] A new release of pip is available: 25.0.1 -> 25.1.1

[notice] To update, run: python.exe -m pip install --upgrade pip

(venv) PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\backend> mkdir uploads

Directory: C:\Users\pc\OneDrive\Desktop\ImageCaptioner\backend

Mode LastWriteTime Length Name

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

d----- 6/23/2025 6:54 AM uploads

(venv) PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\backend> uvicorn main:app --reload

INFO: Will watch for changes in these directories: ['C:\\Users\\pc\\OneDrive\\Desktop\\ImageCaptioner\\backend']

INFO: Uvicorn running on http://127.0.0.1:8000 (Press CTRL+C to quit)

INFO: Started reloader process [18188] using StatReload

Using a slow image processor as `use\_fast` is unset and a slow processor was saved with this model. `use\_fast=True` will be the default behavior in v4.52, even if the model was saved with a slow processor. This will result in minor differences in outputs. You'll still be able to use a slow processor with `use\_fast=False`.

preprocessor\_config.json: 100%|███████████████████████████████████████████████████████████████████████████████████████████| 287/287 [00:00<00:00, 920kB/s]

C:\Users\pc\OneDrive\Desktop\ImageCaptioner\backend\venv\Lib\site-packages\huggingface\_hub\file\_download.py:143: UserWarning: `huggingface\_hub` cache-system uses symlinks by default to efficiently store duplicated files but your machine does not support them in C:\Users\pc\.cache\huggingface\hub\models--Salesforce--blip-image-captioning-base. Caching files will still work but in a degraded version that might require more space on your disk. This warning can be disabled by setting the `HF\_HUB\_DISABLE\_SYMLINKS\_WARNING` environment variable. For more details, see https://huggingface.co/docs/huggingface\_hub/how-to-cache#limitations.

To support symlinks on Windows, you either need to activate Developer Mode or to run Python as an administrator. In order to activate developer mode, see this article: https://docs.microsoft.com/en-us/windows/apps/get-started/enable-your-device-for-development

warnings.warn(message)

tokenizer\_config.json: 100%|█████████████████████████████████████████████████████████████████████████████████████████████| 506/506 [00:00<00:00, 1.63MB/s]

vocab.txt: 100%|████████████████████████████████████████████████████████████████████████████████████████████████████████| 232k/232k [00:00<00:00, 581kB/s]

tokenizer.json: 100%|██████████████████████████████████████████████████████████████████████████████████████████████████| 711k/711k [00:00<00:00, 1.40MB/s]

special\_tokens\_map.json: 100%|████████████████████████████████████████████████████████████████████████████████████████████| 125/125 [00:00<00:00, 644kB/s]

config.json: 100%|███████████████████████████████████████████████████████████████████████████████████████████████████| 4.56k/4.56k [00:00<00:00, 23.8MB/s]

pytorch\_model.bin: 100%|███████████████████████████████████████████████████████████████████████████████████████████████| 990M/990M [09:02<00:00, 1.82MB/s]

INFO: Started server process [14156]

INFO: Waiting for application startup.

INFO: Application startup complete.

model.safetensors: 28%|██████████████████████████▏ | 273M/990M [02:29<06:38, 1.80MB/s]INFO: 127.0.0.1:57029 - "GET /docs HTTP/1.1" 200 OK

INFO: 127.0.0.1:57029 - "GET /openapi.json HTTP/1.1" 200 OK

model.safetensors: 100%|███████████████████████████████████████████████████████████████████████████████████████████████| 990M/990M [09:11<00:00, 1.80MB/s]

backend\captioner.py

from transformers import BlipProcessor, BlipForConditionalGeneration

from PIL import Image

# Load model & processor once

processor = BlipProcessor.from\_pretrained("Salesforce/blip-image-captioning-base")

model = BlipForConditionalGeneration.from\_pretrained("Salesforce/blip-image-captioning-base")

def generate\_caption(image\_path: str) -> str:

    image = Image.open(image\_path).convert("RGB")

    inputs = processor(image, return\_tensors="pt")

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

    caption = processor.decode(output[0], skip\_special\_tokens=True)

    return caption

backend\main.py

from fastapi import FastAPI, UploadFile, File

from fastapi.middleware.cors import CORSMiddleware

from captioner import generate\_caption

import shutil

import os

app = FastAPI()

app.add\_middleware(

    CORSMiddleware,

    allow\_origins=["\*"],  # Allow frontend

    allow\_credentials=True,

    allow\_methods=["\*"],

    allow\_headers=["\*"],

)

@app.post("/caption/")

async def caption\_image(file: UploadFile = File(...)):

    try:

        file\_location = f"uploads/{file.filename}"

        with open(file\_location, "wb") as buffer:

            shutil.copyfileobj(file.file, buffer)

        caption = generate\_caption(file\_location)

        os.remove(file\_location)

        return {"caption": caption}

    except Exception as e:

        return {"error": str(e)}

**✅ E. Create uploads/ Folder**

mkdir uploads

**F. Run the Server**

uvicorn main:app --reload

(venv) PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\backend> uvicorn main:app --reload

INFO: Will watch for changes in these directories: ['C:\\Users\\pc\\OneDrive\\Desktop\\ImageCaptioner\\backend']

INFO: Uvicorn running on http://127.0.0.1:8000 (Press CTRL+C to quit)

INFO: Started reloader process [14844] using StatReload

Using a slow image processor as `use\_fast` is unset and a slow processor was saved with this model. `use\_fast=True` will be the default behavior in v4.52, even if the model was saved with a slow processor. This will result in minor differences in outputs. You'll still be able to use a slow processor with `use\_fast=False`.

INFO: Started server process [41892]

INFO: Waiting for application startup.

INFO: Application startup complete.

INFO: 127.0.0.1:60829 - "GET /docs HTTP/1.1" 200 OK

INFO: 127.0.0.1:60829 - "GET /openapi.json HTTP/1.1" 200 OK

**Ok backend is working now .**

**🎨 3. Frontend Setup (React + Tailwind)**

(venv) PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\backend> cd ..

(venv) PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner> npm create vite@latest frontend -- --template react-ts

> npx

> create-vite frontend react-ts

│

◇ Select a framework:

│ React

│

◇ Select a variant:

│ TypeScript

│

◇ Scaffolding project in C:\Users\pc\OneDrive\Desktop\ImageCaptioner\frontend...

│

└ Done. Now run:

cd frontend

npm install

npm run dev

(venv) PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner>

PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner> cd frontend

PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\frontend> npm install

added 190 packages, and audited 191 packages in 32s

48 packages are looking for funding

run `npm fund` for details

found 0 vulnerabilities

PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\frontend> npm install -D tailwindcss@3.4.17 postcss@8.5.4 autoprefixer

added 77 packages, changed 1 package, and audited 268 packages in 8s

72 packages are looking for funding

run `npm fund` for details

found 0 vulnerabilities

PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\frontend> npx tailwindcss init -p

Created Tailwind CSS config file: tailwind.config.js

Created PostCSS config file: postcss.config.js

PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\frontend>

frontend\src\index.css

@tailwind base;

@tailwind components;

@tailwind utilities;

frontend\src\App.tsx

import { useState } from "react";

function App() {

  const [file, setFile] = useState<File | null>(null);

  const [caption, setCaption] = useState("");

  const [loading, setLoading] = useState(false);

  const handleUpload = async () => {

    if (!file) return;

    setLoading(true);

    const formData = new FormData();

    formData.append("file", file);

    const res = await fetch("http://127.0.0.1:8000/caption/", {

      method: "POST",

      body: formData,

    });

    const data = await res.json();

    setCaption(data.caption || "Error generating caption");

    setLoading(false);

  };

  return (

    <div className="min-h-screen flex flex-col items-center justify-center bg-gray-100 p-4">

      <h1 className="text-2xl font-bold mb-4">🖼️ AI Image Caption Generator</h1>

      <input type="file" accept="image/\*" onChange={(e) => setFile(e.target.files?.[0] || null)} />

      <button

        className="mt-4 px-4 py-2 bg-blue-500 text-white rounded"

        onClick={handleUpload}

        disabled={!file || loading}

      >

        {loading ? "Generating..." : "Generate Caption"}

      </button>

      {caption && <p className="mt-6 text-xl text-center max-w-lg">{caption}</p>}

    </div>

  );

}

export default App;

frontend\tailwind.config.js

/\*\* @type {import('tailwindcss').Config} \*/

export default {

  content: ["./index.html", "./src/\*\*/\*.{js,ts,jsx,tsx}"],

  theme: {

    extend: {},

  },

  plugins: [],

}

Successfully run the backend

PS C:\Users\pc\OneDrive\Desktop\ImageCaptioner\frontend> npm run dev

> frontend@0.0.0 dev

> vite

  VITE v6.3.5  ready in 2169 ms

  ➜  Local:   http://localhost:5173/

  ➜  Network: use --host to expose

  ➜  press h + enter to show help