

# 🔧 Multi-Agent RAG Research Assistant

An interactive and modular Retrieval-Augmented Generation (RAG) pipeline that combines PDF/text ingestion, FAISS vector retrieval, and LLM-based summarization using Together.ai's LLaMA-4. Built with LangGraph agents and deployed via Gradio for real-time, user-customizable research summarization.

## ⚙️ Tech Stack

- **LangGraph** – Multi-agent state machine orchestration
- **FAISS** – Fast Approximate Nearest Neighbor vector store
- **Hugging Face Sentence Transformers** – Text embeddings (MiniLM)
- **Together.ai** – Hosted LLaMA-4 summarization
- **SerpAPI** – Google Search snippets
- **Gradio** – User-friendly interface

```
# Basic NLP and retrieval tools
```

```
!pip install -q sentence-transformers langchain faiss-cpu serpapi
```

```
# LLaMA model loader (this is the slow one that may need C++  
compilation)
```

```
!pip install -q llama-cpp-python
```

```
Installing build dependencies ... ents to build wheel ... etadata  
(pyproject.toml) ... a-cpp-python (pyproject.toml) ...
```

```
!pip install -U langchain-community
```

```
Collecting langchain-community
```

```
  Downloading langchain_community-0.3.21-py3-none-any.whl.metadata  
(2.4 kB)
```

```
Collecting langchain-core<1.0.0,>=0.3.51 (from langchain-community)
```

```
  Downloading langchain_core-0.3.51-py3-none-any.whl.metadata (5.9 kB)
```

```
Collecting langchain<1.0.0,>=0.3.23 (from langchain-community)
```

```
  Downloading langchain-0.3.23-py3-none-any.whl.metadata (7.8 kB)
```

```
Requirement already satisfied: SQLAlchemy<3,>=1.4 in
```

```
/usr/local/lib/python3.11/dist-packages (from langchain-community)  
(2.0.40)
```

```
Requirement already satisfied: requests<3,>=2 in
```

```
/usr/local/lib/python3.11/dist-packages (from langchain-community)  
(2.32.3)
```

```
Requirement already satisfied: PyYAML<=5.3 in
```

```
/usr/local/lib/python3.11/dist-packages (from langchain-community)  
(6.0.2)
```

```
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in
```

```
/usr/local/lib/python3.11/dist-packages (from langchain-community)  
(3.11.15)
```

```
Requirement already satisfied: tenacity!=8.4.0,<10,>=8.1.0 in
```

```
/usr/local/lib/python3.11/dist-packages (from langchain-community)
(9.1.2)
Collecting dataclasses-json<0.7,>=0.5.7 (from langchain-community)
  Downloading dataclasses_json-0.6.7-py3-none-any.whl.metadata (25 kB)
Collecting pydantic-settings<3.0.0,>=2.4.0 (from langchain-community)
  Downloading pydantic_settings-2.8.1-py3-none-any.whl.metadata (3.5
kB)
Requirement already satisfied: langsmith<0.4,>=0.1.125 in
/usr/local/lib/python3.11/dist-packages (from langchain-community)
(0.3.22)
Collecting httpx-sse<1.0.0,>=0.4.0 (from langchain-community)
  Downloading httpx_sse-0.4.0-py3-none-any.whl.metadata (9.0 kB)
Requirement already satisfied: numpy<3,>=1.26.2 in
/usr/local/lib/python3.11/dist-packages (from langchain-community)
(2.0.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (1.5.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (6.3.1)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (0.3.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
>langchain-community) (1.18.3)
Collecting marshmallow<4.0.0,>=3.18.0 (from dataclasses-
json<0.7,>=0.5.7->langchain-community)
  Downloading marshmallow-3.26.1-py3-none-any.whl.metadata (7.3 kB)
Collecting typing-inspect<1,>=0.4.0 (from dataclasses-
json<0.7,>=0.5.7->langchain-community)
  Downloading typing_inspect-0.9.0-py3-none-any.whl.metadata (1.5 kB)
Collecting langchain-text-splitters<1.0.0,>=0.3.8 (from
langchain<1.0.0,>=0.3.23->langchain-community)
  Downloading langchain_text_splitters-0.3.8-py3-none-any.whl.metadata
(1.9 kB)
Requirement already satisfied: pydantic<3.0.0,>=2.7.4 in
/usr/local/lib/python3.11/dist-packages (from
langchain<1.0.0,>=0.3.23->langchain-community) (2.11.1)
```

Requirement already satisfied: jsonpatch<2.0,>=1.33 in  
/usr/local/lib/python3.11/dist-packages (from langchain-  
core<1.0.0,>=0.3.51->langchain-community) (1.33)

Requirement already satisfied: packaging<25,>=23.2 in  
/usr/local/lib/python3.11/dist-packages (from langchain-  
core<1.0.0,>=0.3.51->langchain-community) (24.2)

Requirement already satisfied: typing-extensions>=4.7 in  
/usr/local/lib/python3.11/dist-packages (from langchain-  
core<1.0.0,>=0.3.51->langchain-community) (4.13.0)

Requirement already satisfied: httpx<1,>=0.23.0 in  
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.125-  
>langchain-community) (0.28.1)

Requirement already satisfied: orjson<4.0.0,>=3.9.14 in  
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.125-  
>langchain-community) (3.10.16)

Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in  
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.125-  
>langchain-community) (1.0.0)

Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in  
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.125-  
>langchain-community) (0.23.0)

Collecting python-dotenv>=0.21.0 (from pydantic-  
settings<3.0.0,>=2.4.0->langchain-community)

Downloading python\_dotenv-1.1.0-py3-none-any.whl.metadata (24 kB)

Requirement already satisfied: charset-normalizer<4,>=2 in  
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2-  
>langchain-community) (3.4.1)

Requirement already satisfied: idna<4,>=2.5 in  
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2-  
>langchain-community) (3.10)

Requirement already satisfied: urllib3<3,>=1.21.1 in  
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2-  
>langchain-community) (2.3.0)

Requirement already satisfied: certifi>=2017.4.17 in  
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2-  
>langchain-community) (2025.1.31)

Requirement already satisfied: greenlet>=1 in  
/usr/local/lib/python3.11/dist-packages (from SQLAlchemy<3,>=1.4-  
>langchain-community) (3.1.1)

Requirement already satisfied: anyio in  
/usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.23.0-  
>langsmith<0.4,>=0.1.125->langchain-community) (4.9.0)

Requirement already satisfied: httpcore==1.\* in  
/usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.23.0-  
>langsmith<0.4,>=0.1.125->langchain-community) (1.0.7)

Requirement already satisfied: h11<0.15,>=0.13 in  
/usr/local/lib/python3.11/dist-packages (from httpcore==1.\*-  
>httpx<1,>=0.23.0->langsmith<0.4,>=0.1.125->langchain-community)  
(0.14.0)

```

Requirement already satisfied: jsonpointer>=1.9 in
/usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33-
>langchain-core<1.0.0,>=0.3.51->langchain-community) (3.0.0)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4-
>langchain<1.0.0,>=0.3.23->langchain-community) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4-
>langchain<1.0.0,>=0.3.23->langchain-community) (2.33.0)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4-
>langchain<1.0.0,>=0.3.23->langchain-community) (0.4.0)
Collecting mypy_extensions>=0.3.0 (from typing-inspect<1,>=0.4.0-
>dataclasses-json<0.7,>=0.5.7->langchain-community)
  Downloading mypy_extensions-1.0.0-py3-none-any.whl.metadata (1.1 kB)
Requirement already satisfied: sniffio>=1.1 in
/usr/local/lib/python3.11/dist-packages (from anyio->httpx<1,>=0.23.0-
>langsmith<0.4,>=0.1.125->langchain-community) (1.3.1)
Downloading langchain_community-0.3.21-py3-none-any.whl (2.5 MB)
----- 2.5/2.5 MB 28.3 MB/s eta
0:00:00
----- 1.0/1.0 MB 42.2 MB/s eta
0:00:00
----- 423.3/423.3 kB 25.5 MB/s eta
0:00:00
arshmallow-3.26.1-py3-none-any.whl (50 kB)
----- 50.9/50.9 kB 3.4 MB/s eta
0:00:00
ypy_extensions-1.0.0-py3-none-any.whl (4.7 kB)
Installing collected packages: python-dotenv, mypy_extensions,
marshmallow, httpx-sse, typing-inspect, pydantic-settings,
dataclasses-json, langchain-core, langchain-text-splitters, langchain,
langchain-community
  Attempting uninstall: langchain-core
    Found existing installation: langchain-core 0.3.49
    Uninstalling langchain-core-0.3.49:
      Successfully uninstalled langchain-core-0.3.49
  Attempting uninstall: langchain-text-splitters
    Found existing installation: langchain-text-splitters 0.3.7
    Uninstalling langchain-text-splitters-0.3.7:
      Successfully uninstalled langchain-text-splitters-0.3.7
  Attempting uninstall: langchain
    Found existing installation: langchain 0.3.22
    Uninstalling langchain-0.3.22:
      Successfully uninstalled langchain-0.3.22
Successfully installed dataclasses-json-0.6.7 httpx-sse-0.4.0
langchain-0.3.23 langchain-community-0.3.21 langchain-core-0.3.51
langchain-text-splitters-0.3.8 marshmallow-3.26.1 mypy_extensions-
1.0.0 pydantic-settings-2.8.1 python-dotenv-1.1.0 typing-inspect-0.9.0

```

```
!pip install llama-cpp-python[server]
```

```
Requirement already satisfied: llama-cpp-python[server] in
/usr/local/lib/python3.11/dist-packages (0.3.8)
Requirement already satisfied: typing-extensions>=4.5.0 in
/usr/local/lib/python3.11/dist-packages (from llama-cpp-
python[server]) (4.13.0)
Requirement already satisfied: numpy>=1.20.0 in
/usr/local/lib/python3.11/dist-packages (from llama-cpp-
python[server]) (2.0.2)
Requirement already satisfied: diskcache>=5.6.1 in
/usr/local/lib/python3.11/dist-packages (from llama-cpp-
python[server]) (5.6.3)
Requirement already satisfied: jinja2>=2.11.3 in
/usr/local/lib/python3.11/dist-packages (from llama-cpp-
python[server]) (3.1.6)
Collecting uvicorn>=0.22.0 (from llama-cpp-python[server])
  Downloading uvicorn-0.34.0-py3-none-any.whl.metadata (6.5 kB)
Collecting fastapi>=0.100.0 (from llama-cpp-python[server])
  Downloading fastapi-0.115.12-py3-none-any.whl.metadata (27 kB)
Requirement already satisfied: pydantic-settings>=2.0.1 in
/usr/local/lib/python3.11/dist-packages (from llama-cpp-
python[server]) (2.8.1)
Collecting sse-starlette>=1.6.1 (from llama-cpp-python[server])
  Downloading sse_starlette-2.2.1-py3-none-any.whl.metadata (7.8 kB)
Collecting starlette-context<0.4,>=0.3.6 (from llama-cpp-
python[server])
  Downloading starlette_context-0.3.6-py3-none-any.whl.metadata (4.3
kB)
Requirement already satisfied: PyYAML>=5.1 in
/usr/local/lib/python3.11/dist-packages (from llama-cpp-
python[server]) (6.0.2)
Collecting starlette<0.47.0,>=0.40.0 (from fastapi>=0.100.0->llama-
cpp-python[server])
  Downloading starlette-0.46.1-py3-none-any.whl.metadata (6.2 kB)
Requirement already satisfied: pydantic!=1.8,!<1.8.1,!<2.0.0,!<2.0.1,!
<2.1.0,<3.0.0,>=1.7.4 in /usr/local/lib/python3.11/dist-packages (from
fastapi>=0.100.0->llama-cpp-python[server]) (2.11.1)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2>=2.11.3->llama-
cpp-python[server]) (3.0.2)
Requirement already satisfied: python-dotenv>=0.21.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic-
settings>=2.0.1->llama-cpp-python[server]) (1.1.0)
Requirement already satisfied: anyio>=4.7.0 in
/usr/local/lib/python3.11/dist-packages (from sse-starlette>=1.6.1-
>llama-cpp-python[server]) (4.9.0)
Requirement already satisfied: click>=7.0 in
/usr/local/lib/python3.11/dist-packages (from uvicorn>=0.22.0->llama-
cpp-python[server]) (8.1.8)
```

```

Requirement already satisfied: h11>=0.8 in
/usr/local/lib/python3.11/dist-packages (from uvicorn>=0.22.0->llama-
cpp-python[server]) (0.14.0)
Requirement already satisfied: idna>=2.8 in
/usr/local/lib/python3.11/dist-packages (from anyio>=4.7.0->sse-
starlette>=1.6.1->llama-cpp-python[server]) (3.10)
Requirement already satisfied: sniffio>=1.1 in
/usr/local/lib/python3.11/dist-packages (from anyio>=4.7.0->sse-
starlette>=1.6.1->llama-cpp-python[server]) (1.3.1)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic!=1.8,!1.8.1,!
=2.0.0,!2.0.1,!2.1.0,<3.0.0,>=1.7.4->fastapi>=0.100.0->llama-cpp-
python[server]) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic!=1.8,!1.8.1,!
=2.0.0,!2.0.1,!2.1.0,<3.0.0,>=1.7.4->fastapi>=0.100.0->llama-cpp-
python[server]) (2.33.0)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic!=1.8,!1.8.1,!
=2.0.0,!2.0.1,!2.1.0,<3.0.0,>=1.7.4->fastapi>=0.100.0->llama-cpp-
python[server]) (0.4.0)
Downloading fastapi-0.115.12-py3-none-any.whl (95 kB)
_____ 95.2/95.2 kB 6.1 MB/s eta
0:00:00
_____ 62.3/62.3 kB 4.2 MB/s eta
0:00:00
_____ 72.0/72.0 kB 4.6 MB/s eta
0:00:00

!pip install together

Collecting together
  Downloading together-1.5.5-py3-none-any.whl.metadata (14 kB)
Requirement already satisfied: aiohttp<4.0.0,>=3.9.3 in
/usr/local/lib/python3.11/dist-packages (from together) (3.11.15)
Requirement already satisfied: click<9.0.0,>=8.1.7 in
/usr/local/lib/python3.11/dist-packages (from together) (8.1.8)
Collecting eval-type-backport<0.3.0,>=0.1.3 (from together)
  Downloading eval_type_backport-0.2.2-py3-none-any.whl.metadata (2.2
kB)
Requirement already satisfied: filelock<4.0.0,>=3.13.1 in
/usr/local/lib/python3.11/dist-packages (from together) (3.18.0)
Requirement already satisfied: numpy>=1.23.5 in
/usr/local/lib/python3.11/dist-packages (from together) (2.0.2)
Requirement already satisfied: pillow<12.0.0,>=11.1.0 in
/usr/local/lib/python3.11/dist-packages (from together) (11.1.0)
Requirement already satisfied: pyarrow>=10.0.1 in
/usr/local/lib/python3.11/dist-packages (from together) (18.1.0)
Requirement already satisfied: pydantic<3.0.0,>=2.6.3 in
/usr/local/lib/python3.11/dist-packages (from together) (2.11.1)

```

Requirement already satisfied: requests<3.0.0,>=2.31.0 in  
/usr/local/lib/python3.11/dist-packages (from together) (2.32.3)

Requirement already satisfied: rich<14.0.0,>=13.8.1 in  
/usr/local/lib/python3.11/dist-packages (from together) (13.9.4)

Requirement already satisfied: tabulate<0.10.0,>=0.9.0 in  
/usr/local/lib/python3.11/dist-packages (from together) (0.9.0)

Requirement already satisfied: tqdm<5.0.0,>=4.66.2 in  
/usr/local/lib/python3.11/dist-packages (from together) (4.67.1)

Requirement already satisfied: typer<0.16,>=0.9 in  
/usr/local/lib/python3.11/dist-packages (from together) (0.15.2)

Requirement already satisfied: aiohappyeyeballs>=2.3.0 in  
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.9.3->together) (2.6.1)

Requirement already satisfied: aiosignal>=1.1.2 in  
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.9.3->together) (1.3.2)

Requirement already satisfied: attrs>=17.3.0 in  
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.9.3->together) (25.3.0)

Requirement already satisfied: frozenlist>=1.1.1 in  
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.9.3->together) (1.5.0)

Requirement already satisfied: multidict<7.0,>=4.5 in  
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.9.3->together) (6.3.1)

Requirement already satisfied: propcache>=0.2.0 in  
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.9.3->together) (0.3.1)

Requirement already satisfied: yarl<2.0,>=1.17.0 in  
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.9.3->together) (1.18.3)

Requirement already satisfied: annotated-types>=0.6.0 in  
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.6.3->together) (0.7.0)

Requirement already satisfied: pydantic-core==2.33.0 in  
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.6.3->together) (2.33.0)

Requirement already satisfied: typing-extensions>=4.12.2 in  
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.6.3->together) (4.13.0)

Requirement already satisfied: typing-inspection>=0.4.0 in  
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.6.3->together) (0.4.0)

Requirement already satisfied: charset-normalizer<4,>=2 in  
/usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.31.0->together) (3.4.1)

Requirement already satisfied: idna<4,>=2.5 in  
/usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.31.0->together) (3.10)

Requirement already satisfied: urllib3<3,>=1.21.1 in  
/usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.31.0-  
>together) (2.3.0)  
Requirement already satisfied: certifi>=2017.4.17 in  
/usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.31.0-  
>together) (2025.1.31)  
Requirement already satisfied: markdown-it-py>=2.2.0 in  
/usr/local/lib/python3.11/dist-packages (from rich<14.0.0,>=13.8.1-  
>together) (3.0.0)  
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in  
/usr/local/lib/python3.11/dist-packages (from rich<14.0.0,>=13.8.1-  
>together) (2.18.0)  
Requirement already satisfied: shellingham>=1.3.0 in  
/usr/local/lib/python3.11/dist-packages (from typer<0.16,>=0.9-  
>together) (1.5.4)  
Requirement already satisfied: mdurl~=0.1 in  
/usr/local/lib/python3.11/dist-packages (from markdown-it-py>=2.2.0-  
>rich<14.0.0,>=13.8.1->together) (0.1.2)  
Downloading together-1.5.5-py3-none-any.whl (87 kB)  
87.9/87.9 kB 2.1 MB/s eta  
0:00:00

!pip install langgraph

Collecting langgraph  
 Downloading langgraph-0.3.25-py3-none-any.whl.metadata (7.7 kB)  
Requirement already satisfied: langchain-core<0.4,>=0.1 in  
/usr/local/lib/python3.11/dist-packages (from langgraph) (0.3.51)  
Collecting langgraph-checkpoint<3.0.0,>=2.0.10 (from langgraph)  
 Downloading langgraph\_checkpoint-2.0.24-py3-none-any.whl.metadata  
(4.6 kB)  
Collecting langgraph-prebuilt<0.2,>=0.1.1 (from langgraph)  
 Downloading langgraph\_prebuilt-0.1.8-py3-none-any.whl.metadata (5.0  
kB)  
Collecting langgraph-sdk<0.2.0,>=0.1.42 (from langgraph)  
 Downloading langgraph\_sdk-0.1.61-py3-none-any.whl.metadata (1.8 kB)  
Collecting xxhash<4.0.0,>=3.5.0 (from langgraph)  
 Downloading xxhash-3.5.0-cp311-cp311-  
manylinux\_2\_17\_x86\_64.manylinux2014\_x86\_64.whl.metadata (12 kB)  
Requirement already satisfied: langsmith<0.4,>=0.1.125 in  
/usr/local/lib/python3.11/dist-packages (from langchain-  
core<0.4,>=0.1->langgraph) (0.3.22)  
Requirement already satisfied: tenacity!=8.4.0,<10.0.0,>=8.1.0 in  
/usr/local/lib/python3.11/dist-packages (from langchain-  
core<0.4,>=0.1->langgraph) (9.1.2)  
Requirement already satisfied: jsonpatch<2.0,>=1.33 in  
/usr/local/lib/python3.11/dist-packages (from langchain-  
core<0.4,>=0.1->langgraph) (1.33)  
Requirement already satisfied: PyYAML>=5.3 in  
/usr/local/lib/python3.11/dist-packages (from langchain-



```
core<0.4,>=0.1->langgraph) (6.0.2)
Requirement already satisfied: packaging<25,>=23.2 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<0.4,>=0.1->langgraph) (24.2)
Requirement already satisfied: typing-extensions>=4.7 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<0.4,>=0.1->langgraph) (4.13.0)
Requirement already satisfied: pydantic<3.0.0,>=2.5.2 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<0.4,>=0.1->langgraph) (2.11.1)
Collecting ormsgpack<2.0.0,>=1.8.0 (from langgraph-
checkpoint<3.0.0,>=2.0.10->langgraph)
  Downloading ormsgpack-1.9.1-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (43 kB)
----- 43.5/43.5 kB 2.0 MB/s eta
```

0:00:00

```
ent already satisfied: httpx>=0.25.2 in
/usr/local/lib/python3.11/dist-packages (from langgraph-
sdk<0.2.0,>=0.1.42->langgraph) (0.28.1)
Requirement already satisfied: orjson>=3.10.1 in
/usr/local/lib/python3.11/dist-packages (from langgraph-
sdk<0.2.0,>=0.1.42->langgraph) (3.10.16)
Requirement already satisfied: anyio in
/usr/local/lib/python3.11/dist-packages (from httpx>=0.25.2-
>langgraph-sdk<0.2.0,>=0.1.42->langgraph) (4.9.0)
Requirement already satisfied: certifi in
/usr/local/lib/python3.11/dist-packages (from httpx>=0.25.2-
>langgraph-sdk<0.2.0,>=0.1.42->langgraph) (2025.1.31)
Requirement already satisfied: httpcore==1.* in
/usr/local/lib/python3.11/dist-packages (from httpx>=0.25.2-
>langgraph-sdk<0.2.0,>=0.1.42->langgraph) (1.0.7)
Requirement already satisfied: idna in /usr/local/lib/python3.11/dist-
packages (from httpx>=0.25.2->langgraph-sdk<0.2.0,>=0.1.42->langgraph)
(3.10)
Requirement already satisfied: h11<0.15,>=0.13 in
/usr/local/lib/python3.11/dist-packages (from httpcore==1.*-
>httpx>=0.25.2->langgraph-sdk<0.2.0,>=0.1.42->langgraph) (0.14.0)
Requirement already satisfied: jsonpointer>=1.9 in
/usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33-
>langchain-core<0.4,>=0.1->langgraph) (3.0.0)
Requirement already satisfied: requests<3,>=2 in
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.125-
>langchain-core<0.4,>=0.1->langgraph) (2.32.3)
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.125-
>langchain-core<0.4,>=0.1->langgraph) (1.0.0)
Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.125-
>langchain-core<0.4,>=0.1->langgraph) (0.23.0)
```

```

Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.5.2-
>langchain-core<0.4,>=0.1->langgraph) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.5.2-
>langchain-core<0.4,>=0.1->langgraph) (2.33.0)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.5.2-
>langchain-core<0.4,>=0.1->langgraph) (0.4.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2-
>langsmith<0.4,>=0.1.125->langchain-core<0.4,>=0.1->langgraph) (3.4.1)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2-
>langsmith<0.4,>=0.1.125->langchain-core<0.4,>=0.1->langgraph) (2.3.0)
Requirement already satisfied: sniffio>=1.1 in
/usr/local/lib/python3.11/dist-packages (from anyio->httpx>=0.25.2-
>langgraph-sdk<0.2.0,>=0.1.42->langgraph) (1.3.1)
Downloading langgraph-0.3.25-py3-none-any.whl (142 kB)
_____ 142.4/142.4 kB 4.3 MB/s eta
0:00:00
_____ 42.0/42.0 kB 2.7 MB/s eta
0:00:00
_____ 47.2/47.2 kB 2.9 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
_____ 194.8/194.8 kB 12.1 MB/s eta
0:00:00
sgpack-1.9.1-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (223 kB)
_____ 223.6/223.6 kB 9.2 MB/s eta
0:00:00
sgpack, langgraph-sdk, langgraph-checkpoint, langgraph-prebuilt,
langgraph
Successfully installed langgraph-0.3.25 langgraph-checkpoint-2.0.24
langgraph-prebuilt-0.1.8 langgraph-sdk-0.1.61 ormsgpack-1.9.1 xxhash-
3.5.0

!pip install grandalf

Collecting grandalf
  Downloading grandalf-0.8-py3-none-any.whl.metadata (1.7 kB)
Requirement already satisfied: pyparsing in
/usr/local/lib/python3.11/dist-packages (from grandalf) (3.2.3)
Downloading grandalf-0.8-py3-none-any.whl (41 kB)
_____ 41.8/41.8 kB 1.7 MB/s eta
0:00:00

!pip install gradio
!python app.py

```

```
Collecting gradio
  Downloading gradio-5.23.3-py3-none-any.whl.metadata (16 kB)
Collecting aiofiles<24.0,>=22.0 (from gradio)
  Downloading aiofiles-23.2.1-py3-none-any.whl.metadata (9.7 kB)
Requirement already satisfied: anyio<5.0,>=3.0 in
/usr/local/lib/python3.11/dist-packages (from gradio) (4.9.0)
Requirement already satisfied: fastapi<1.0,>=0.115.2 in
/usr/local/lib/python3.11/dist-packages (from gradio) (0.115.12)
Collecting ffmpeg (from gradio)
  Downloading ffmpeg-0.5.0-py3-none-any.whl.metadata (3.0 kB)
Collecting gradio-client==1.8.0 (from gradio)
  Downloading gradio_client-1.8.0-py3-none-any.whl.metadata (7.1 kB)
Collecting groovy~=0.1 (from gradio)
  Downloading groovy-0.1.2-py3-none-any.whl.metadata (6.1 kB)
Requirement already satisfied: httpx>=0.24.1 in
/usr/local/lib/python3.11/dist-packages (from gradio) (0.28.1)
Requirement already satisfied: huggingface-hub>=0.28.1 in
/usr/local/lib/python3.11/dist-packages (from gradio) (0.30.1)
Requirement already satisfied: jinja2<4.0 in
/usr/local/lib/python3.11/dist-packages (from gradio) (3.1.6)
Requirement already satisfied: markupsafe<4.0,>=2.0 in
/usr/local/lib/python3.11/dist-packages (from gradio) (3.0.2)
Requirement already satisfied: numpy<3.0,>=1.0 in
/usr/local/lib/python3.11/dist-packages (from gradio) (2.0.2)
Requirement already satisfied: orjson~=3.0 in
/usr/local/lib/python3.11/dist-packages (from gradio) (3.10.16)
Requirement already satisfied: packaging in
/usr/local/lib/python3.11/dist-packages (from gradio) (24.2)
Requirement already satisfied: pandas<3.0,>=1.0 in
/usr/local/lib/python3.11/dist-packages (from gradio) (2.2.2)
Requirement already satisfied: pillow<12.0,>=8.0 in
/usr/local/lib/python3.11/dist-packages (from gradio) (11.1.0)
Requirement already satisfied: pydantic<2.12,>=2.0 in
/usr/local/lib/python3.11/dist-packages (from gradio) (2.11.1)
Collecting pydub (from gradio)
  Downloading pydub-0.25.1-py2.py3-none-any.whl.metadata (1.4 kB)
Collecting python-multipart>=0.0.18 (from gradio)
  Downloading python_multipart-0.0.20-py3-none-any.whl.metadata (1.8
kB)
Requirement already satisfied: pyyaml<7.0,>=5.0 in
/usr/local/lib/python3.11/dist-packages (from gradio) (6.0.2)
Collecting ruff>=0.9.3 (from gradio)
  Downloading ruff-0.11.4-py3-none-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (25 kB)
Collecting safehttpx<0.2.0,>=0.1.6 (from gradio)
  Downloading safehttpx-0.1.6-py3-none-any.whl.metadata (4.2 kB)
Collecting semantic-version~=2.0 (from gradio)
  Downloading semantic_version-2.10.0-py2.py3-none-any.whl.metadata
(9.7 kB)
Requirement already satisfied: starlette<1.0,>=0.40.0 in
```

```
/usr/local/lib/python3.11/dist-packages (from gradio) (0.46.1)
Collecting tomkit<0.14.0,>=0.12.0 (from gradio)
  Downloading tomkit-0.13.2-py3-none-any.whl.metadata (2.7 kB)
Requirement already satisfied: typer<1.0,>=0.12 in
/usr/local/lib/python3.11/dist-packages (from gradio) (0.15.2)
Requirement already satisfied: typing-extensions~=4.0 in
/usr/local/lib/python3.11/dist-packages (from gradio) (4.13.0)
Requirement already satisfied: uvicorn>=0.14.0 in
/usr/local/lib/python3.11/dist-packages (from gradio) (0.34.0)
Requirement already satisfied: fsspec in
/usr/local/lib/python3.11/dist-packages (from gradio-client==1.8.0-
>gradio) (2025.3.2)
Requirement already satisfied: websockets<16.0,>=10.0 in
/usr/local/lib/python3.11/dist-packages (from gradio-client==1.8.0-
>gradio) (15.0.1)
Requirement already satisfied: idna>=2.8 in
/usr/local/lib/python3.11/dist-packages (from anyio<5.0,>=3.0->gradio)
(3.10)
Requirement already satisfied: sniffio>=1.1 in
/usr/local/lib/python3.11/dist-packages (from anyio<5.0,>=3.0->gradio)
(1.3.1)
Requirement already satisfied: certifi in
/usr/local/lib/python3.11/dist-packages (from httpx>=0.24.1->gradio)
(2025.1.31)
Requirement already satisfied: httpcore==1.* in
/usr/local/lib/python3.11/dist-packages (from httpx>=0.24.1->gradio)
(1.0.7)
Requirement already satisfied: h11<0.15,>=0.13 in
/usr/local/lib/python3.11/dist-packages (from httpcore==1.*-
>httpx>=0.24.1->gradio) (0.14.0)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.28.1-
>gradio) (3.18.0)
Requirement already satisfied: requests in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.28.1-
>gradio) (2.32.3)
Requirement already satisfied: tqdm>=4.42.1 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.28.1-
>gradio) (4.67.1)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.11/dist-packages (from pandas<3.0,>=1.0-
>gradio) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.11/dist-packages (from pandas<3.0,>=1.0-
>gradio) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.11/dist-packages (from pandas<3.0,>=1.0-
>gradio) (2025.2)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<2.12,>=2.0-
```

```

>gradio) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<2.12,>=2.0-
>gradio) (2.33.0)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<2.12,>=2.0-
>gradio) (0.4.0)
Requirement already satisfied: click>=8.0.0 in
/usr/local/lib/python3.11/dist-packages (from typer<1.0,>=0.12-
>gradio) (8.1.8)
Requirement already satisfied: shellingham>=1.3.0 in
/usr/local/lib/python3.11/dist-packages (from typer<1.0,>=0.12-
>gradio) (1.5.4)
Requirement already satisfied: rich>=10.11.0 in
/usr/local/lib/python3.11/dist-packages (from typer<1.0,>=0.12-
>gradio) (13.9.4)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2-
>pandas<3.0,>=1.0->gradio) (1.17.0)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/usr/local/lib/python3.11/dist-packages (from rich>=10.11.0-
>typer<1.0,>=0.12->gradio) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.11/dist-packages (from rich>=10.11.0-
>typer<1.0,>=0.12->gradio) (2.18.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->huggingface-
hub>=0.28.1->gradio) (3.4.1)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests->huggingface-
hub>=0.28.1->gradio) (2.3.0)
Requirement already satisfied: mdurl~=0.1 in
/usr/local/lib/python3.11/dist-packages (from markdown-it-py>=2.2.0-
>rich>=10.11.0->typer<1.0,>=0.12->gradio) (0.1.2)
Downloading gradio-5.23.3-py3-none-any.whl (46.5 MB)
_____ 46.5/46.5 MB 15.6 MB/s eta
0:00:00
_____ 322.2/322.2 kB 14.9 MB/s eta
0:00:00
ultipart-0.0.20-py3-none-any.whl (24 kB)
Downloading ruff-0.11.4-py3-none-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (11.3 MB)
_____ 11.3/11.3 MB 51.7 MB/s eta
0:00:00
antic_version-2.10.0-py2.py3-none-any.whl (15 kB)
Downloading tomlkit-0.13.2-py3-none-any.whl (37 kB)
Downloading ffmpeg-0.5.0-py3-none-any.whl (6.0 kB)
Downloading pydub-0.25.1-py2.py3-none-any.whl (32 kB)
Installing collected packages: pydub, tomlkit, semantic-version, ruff,
python-multipart, groovy, ffmpeg, aiofiles, safehttpx, gradio-client,

```

```
gradio
Successfully installed aiofiles-23.2.1 ffmpeg-0.5.0 gradio-5.23.3
gradio-client-1.8.0 groovy-0.1.2 pydub-0.25.1 python-multipart-0.0.20
ruff-0.11.4 safehttpx-0.1.6 semantic-version-2.10.0 tomlkit-0.13.2
python3: can't open file '/content/app.py': [Errno 2] No such file or
directory
```

```
import faiss
import numpy as np
import requests
from langchain_community.embeddings import HuggingFaceEmbeddings
from langchain_community.vectorstores.faiss import FAISS
from langchain.docstore.document import Document
```

## □ Pipeline Flow

1. □ **SearchAgent** – Uses SerpAPI to fetch top Google results
2. □ **PDFLoaderAgent** – Loads and splits uploaded PDFs into text
3. □ **EmbedAgent** – Embeds all documents using MiniLM
4. □ **RetrieveAgent** – Uses FAISS to find top-k relevant chunks
5. □ **SummarizeAgent** – Calls Together.ai LLaMA-4 with user prompt

The entire workflow is managed via LangGraph's stateful graph execution.

## □ How to Use

1. Enter your **research query** (e.g., "AutoML in 2024 (500-word summary)")
2. Optionally upload a **PDF file**
3. Enter your **SerpAPI** and **Together.ai API keys**
4. Click **Submit** and wait for the generated summary

```
def search_serpapi(query, api_key):
    params = {
        "engine": "google",
        "q": query,
        "api_key": api_key,
        "num": 20 # Ask for more results
    }
    response = requests.get("https://serpapi.com/search",
params=params)
    data = response.json()
    results = []
    for item in data.get("organic_results", []):
        results.append(item.get("snippet", ""))
    return results

# -----
# 2. Query SerpAPI for NAS content
# -----
```

```
SERPAPI_KEY =  
"e63948d7aaba37cb53c7f84f9a8e8e2e2985470ad2bb00937fc2be10aabf13ad"  
query = "NAS (Neural Architecture Search) Techniques in 2024"  
snippets = search_serpapi(query, SERPAPI_KEY)
```

```
print(" Retrieved snippets:")  
for i, s in enumerate(snippets, 1):  
    print(f"{i}. {s}")  
print(f"\n Total snippets retrieved: {len(snippets)}")
```

Retrieved snippets:

1. One famous implementation is Google's AutoML, where RL agents design neural networks for tasks like image classification.
2. This paper delves into the multifaceted aspects of NAS, elaborating on its recent advances, applications, tools, benchmarks and prospective research directions.
3. This work presents a novel approach to neural architecture search (NAS) that aims to increase carbon efficiency for the model design process ...
4. Neural architecture search (NAS) is an automated machine learning method that aims to find optimal model structures by searching the neural network architecture ...
5. Neural Architecture Search (NAS) is a technique in machine learning that automates the process of designing neural network architectures.
6. NAS with Reinforcement Learning. Reinforcement learning-based NAS methods frame the architecture search as a sequential decision-making process.
7. We propose a novel and effective method called Evolutionary Gradient-Based Neural Architecture Search (EG-NAS).
8. Explore Neural Architecture Search (NAS) techniques for optimizing computer vision models and enhancing performance.
9. We propose a novel neural architecture search scheme for binary neural networks, named NAS-BNN. We first carefully design a search space based on the unique ...
10. In this paper, we propose an attention-based progressive partially connected neural architecture search method (PPCAtt-NAS) to address these two issues.
11. The document details the shift from expert-driven design to algorithm-driven processes, exploring initial methodologies like reinforcement learning and ...
12. Neural Architecture Search (NAS) aims to automate deep neural network design across various applications, while a good search space design is core to NAS ...
13. Neural Architecture Search (NAS) methods seek effective optimization toward performance metrics regarding model accuracy and generalization while facing ...
14. This article takes a close look at how to blend transfer learning ideas with NAS to make the search quicker and improve how models perform.
15. Neural architecture search (NAS) is a technique for automating the

design of artificial neural networks (ANN), a widely used model in the field of machine ...

16. Recent federated NAS methods not only reduce manual effort but also provide more accuracy than traditional FL methods like FedAvg. Despite the success, existing ...

17. This paper introduces TrajectoryNAS, a novel neural architecture search (NAS) method designed to develop an efficient and more accurate LiDAR-based trajectory ...

18. Neural architecture search (NAS) is a technique for automating the design of artificial neural networks (ANN), a widely used model in the field of machine ...

19. We introduce DfuseNAS, a novel NAS methodology rooted in diffusion processes. DfuseNAS brings substantial improvements in both NAS search efficiency and the ...

□ Total snippets retrieved: 19

```
docs = [Document(page_content=s) for s in snippets if s.strip() != ""]

# Ensure we have enough to train IVF index (e.g., at least nlist)
if len(docs) < 5:
    raise ValueError("Not enough snippets to train IndexIVFFlat.
Please retrieve more results.")
```

## □ API Keys Required

- SERPAPI\_KEY from [serpapi.com](https://serpapi.com)
- TOGETHER\_API\_KEY from [platform.together.xyz](https://platform.together.xyz)

Keys can be provided in the Gradio UI or loaded securely in a `.env` file (if running locally).

```
from langchain.docstore import InMemoryDocstore
from langchain_community.vectorstores.faiss import FAISS
from langchain.docstore.document import Document
from langchain_community.embeddings import HuggingFaceEmbeddings
import faiss
import numpy as np

# 1. Documents & embeddings
docs = [Document(page_content=s) for s in snippets if s.strip() != ""]
embedding_model = HuggingFaceEmbeddings(model_name="all-MiniLM-L6-v2")
vectors = embedding_model.embed_documents([doc.page_content for doc in docs])
vectors_np = np.array(vectors).astype("float32")

# 2. Build IVF index
d = vectors_np.shape[1]
nlist = 5
quantizer = faiss.IndexFlatL2(d)
index = faiss.IndexIVFFlat(quantizer, d, nlist)
```



```

index.train(vectors_np)
index.add(vectors_np)

# 3. LangChain expects this mapping:
index_to_docstore_id = {i: str(i) for i in range(len(docs))}
docstore = InMemoryDocstore({str(i): doc for i, doc in
enumerate(docs)})

# 4. Now construct FAISS object manually
vectorstore = FAISS(
    embedding_function=embedding_model.embed_query,
    index=index,
    docstore=docstore,
    index_to_docstore_id=index_to_docstore_id
)

# 5. Run similarity search
query_text = "recent advancements in neural architecture search"
results = vectorstore.similarity_search(query_text, k=3)

print("\n Top relevant documents:")
for i, doc in enumerate(results, 1):
    print(f"\n--- Retrieved {i} ---\n{doc.page_content}")

<ipython-input-16-58ea2205578f>:10: LangChainDeprecationWarning: The
class `HuggingFaceEmbeddings` was deprecated in LangChain 0.2.2 and
will be removed in 1.0. An updated version of the class exists in
the :class:`~langchain-huggingface` package and should be used instead.
To use it run `pip install -U :class:`~langchain-huggingface` and
import as `from :class:`~langchain_huggingface import
HuggingFaceEmbeddings`.
  embedding_model = HuggingFaceEmbeddings(model_name="all-MiniLM-L6-
v2")
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py
:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your
settings tab (https://huggingface.co/settings/tokens), set it as
secret in your Google Colab and restart your session.
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to
access public models or datasets.
  warnings.warn(

{"model_id": "03a1834f1a384d07b246aa876ac7419e", "version_major": 2, "vers
ion_minor": 0}

{"model_id": "fa9646dd3ee94ec2953b29fb4f8a6275", "version_major": 2, "vers
ion_minor": 0}

```

```
{"model_id":"52fef3a837264554bdc8f15325d13050","version_major":2,"version_minor":0}

{"model_id":"801d864245a942b587110069be47daa4","version_major":2,"version_minor":0}

{"model_id":"08de60e80b71404da4b41bf1d46d8cc9","version_major":2,"version_minor":0}

{"model_id":"cb5a6dbe58aa4552aff8f4fbe1c46e03","version_major":2,"version_minor":0}

{"model_id":"a977215cce954c93ba345fa5484244b9","version_major":2,"version_minor":0}

{"model_id":"2ff186d113a84917ab3ade27172066a8","version_major":2,"version_minor":0}

{"model_id":"199f1ec9cf6a4c928a4837f02b1b975a","version_major":2,"version_minor":0}

{"model_id":"1b05141f96c446d699f0001b651529c1","version_major":2,"version_minor":0}

{"model_id":"47028faab4ed45158d12f65926f6fca5","version_major":2,"version_minor":0}
```

WARNING:langchain\_community.vectorstores.faiss:`embedding\_function` is expected to be an Embeddings object, support for passing in a function will soon be removed.

□ Top relevant documents:

--- Retrieved 1 ---

We propose a novel and effective method called Evolutionary Gradient-Based Neural Architecture Search (EG-NAS).

--- Retrieved 2 ---

Neural Architecture Search (NAS) methods seek effective optimization toward performance metrics regarding model accuracy and generalization while facing ...

--- Retrieved 3 ---

Neural Architecture Search (NAS) aims to automate deep neural network design across various applications, while a good search space design is core to NAS ...

```
from together import Together
```

```
client =
```

```
Together(api_key="f686a109d1f4267d0019648d3b74d2f014a5c8585ec58e9734e9
```

```

8c6b38f1f658")

retrieved_texts = [doc.page_content for doc in results]
prompt = "Summarize the following research snippets:\n\n" + "\n\n".join(retrieved_texts)

response = client.chat.completions.create(
    model="meta-llama/Llama-4-Maverick-17B-128E-Instruct-FP8",
    messages=[
        {"role": "system", "content": "You are a helpful research assistant."},
        {"role": "user", "content": prompt}
    ],
    max_tokens=512,
    temperature=0.7,
)

print("\n Summary:\n", response.choices[0].message.content)

```

Summary:

Here is a summary of the research snippets:

The snippets discuss Neural Architecture Search (NAS), a method that aims to automate the design of deep neural networks. The researchers propose a new NAS method called Evolutionary Gradient-Based Neural Architecture Search (EG-NAS), suggesting that existing NAS methods face challenges in optimizing model accuracy and generalization. A key aspect of NAS is designing a good search space, which is crucial for its effectiveness.

## API Keys Required

- SERPAPI\_KEY from [serpapi.com](https://serpapi.com)
- TOGETHER\_API\_KEY from [platform.together.xyz](https://platform.together.xyz)

Keys can be provided in the Gradio UI or loaded securely in a `.env` file (if running locally).