

```
!pip install gradio langchain langchain_community langchain_milvus transformers pymilvus
```

```
Requirement already satisfied: gradio in /usr/local/lib/python3.11/dist-packages (5.39.0)
Requirement already satisfied: langchain in /usr/local/lib/python3.11/dist-packages (0.3.27)
Collecting langchain_community
  Downloading langchain_community-0.3.27-py3-none-any.whl.metadata (2.9 kB)
Collecting langchain_milvus
  Downloading langchain_milvus-0.2.1-py3-none-any.whl.metadata (3.8 kB)
Requirement already satisfied: transformers in /usr/local/lib/python3.11/dist-packages (4.54.1)
Collecting pymilvus
  Downloading pymilvus-2.6.0-py3-none-any.whl.metadata (5.8 kB)
Requirement already satisfied: aiofiles<25.0,>=22.0 in /usr/local/lib/python3.11/dist-packages (from gradio) (24.1.0)
Requirement already satisfied: anyio<5.0,>=3.0 in /usr/local/lib/python3.11/dist-packages (from gradio) (4.10.0)
Requirement already satisfied: brotli>=1.1.0 in /usr/local/lib/python3.11/dist-packages (from gradio) (1.1.0)
Requirement already satisfied: fastapi<1.0,>=0.115.2 in /usr/local/lib/python3.11/dist-packages (from gradio) (0.116.1)
Requirement already satisfied: ffmpeg in /usr/local/lib/python3.11/dist-packages (from gradio) (0.6.1)
Requirement already satisfied: gradio-client==1.11.0 in /usr/local/lib/python3.11/dist-packages (from gradio) (1.11.0)
Requirement already satisfied: groovy~=0.1 in /usr/local/lib/python3.11/dist-packages (from gradio) (0.1.2)
Requirement already satisfied: httpx<1.0,>=0.24.1 in /usr/local/lib/python3.11/dist-packages (from gradio) (0.28.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.33.5 in /usr/local/lib/python3.11/dist-packages (from gradio) (0.34.3)
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) (3.11.1)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-packages (from gradio) (25.0)
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) (11.3.0)
Requirement already satisfied: pydantic<2.12,>=2.0 in /usr/local/lib/python3.11/dist-packages (from gradio) (2.11.7)
Requirement already satisfied: pydub in /usr/local/lib/python3.11/dist-packages (from gradio) (0.25.1)
Requirement already satisfied: python-multipart==0.0.18 in /usr/local/lib/python3.11/dist-packages (from gradio) (0.0.20)
Requirement already satisfied: pyyaml<7.0,>=5.0 in /usr/local/lib/python3.11/dist-packages (from gradio) (6.0.2)
Requirement already satisfied: ruff>0.9.3 in /usr/local/lib/python3.11/dist-packages (from gradio) (0.12.7)
Requirement already satisfied: safehttpx<0.2.0,>=0.1.6 in /usr/local/lib/python3.11/dist-packages (from gradio) (0.1.6)
Requirement already satisfied: semantic-version==2.0 in /usr/local/lib/python3.11/dist-packages (from gradio) (2.10.0)
Requirement already satisfied: starlette<1.0,>=0.40.0 in /usr/local/lib/python3.11/dist-packages (from gradio) (0.47.2)
Requirement already satisfied: tomkit<0.14.0,>=0.12.0 in /usr/local/lib/python3.11/dist-packages (from gradio) (0.13.3)
Requirement already satisfied: typer<1.0,>=0.12 in /usr/local/lib/python3.11/dist-packages (from gradio) (0.16.0)
Requirement already satisfied: typing-extensions~=4.0 in /usr/local/lib/python3.11/dist-packages (from gradio) (4.14.1)
Requirement already satisfied: uvicorn>=0.14.0 in /usr/local/lib/python3.11/dist-packages (from gradio) (0.35.0)
Requirement already satisfied: fsspec in /usr/local/lib/python3.11/dist-packages (from gradio-client==1.11.0->gradio) (2025.3.0)
Requirement already satisfied: websockets<16.0,>=10.0 in /usr/local/lib/python3.11/dist-packages (from gradio-client==1.11.0->gradio) (13.1)
Requirement already satisfied: langchain-core<1.0.0,>=0.3.72 in /usr/local/lib/python3.11/dist-packages (from langchain) (0.3.72)
Requirement already satisfied: langchain-text-splitters<1.0.0,>=0.3.9 in /usr/local/lib/python3.11/dist-packages (from langchain) (0.3.9)
Requirement already satisfied: langsmith==0.1.17 in /usr/local/lib/python3.11/dist-packages (from langchain) (0.4.10)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in /usr/local/lib/python3.11/dist-packages (from langchain) (2.0.42)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.11/dist-packages (from langchain) (2.32.3)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /usr/local/lib/python3.11/dist-packages (from langchain_community) (3.12.15)
Requirement already satisfied: tenacity!=8.4.0,<10,>=8.1.0 in /usr/local/lib/python3.11/dist-packages (from langchain_community) (8.4.0)
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-2.10.1-py3-none-any.whl.metadata (3.4 kB)
Collecting httpx-sse<1.0.0,>=0.4.0 (from langchain_community)
  Downloading httpx_sse-0.4.1-py3-none-any.whl.metadata (9.4 kB)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from transformers) (3.18.0)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11/dist-packages (from transformers) (2024.11.6)
Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.21.4)
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.5.3)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.11/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: setuptools<69 in /usr/local/lib/python3.11/dist-packages (from pymilvus) (75.2.0)
```

```
!pip install replicate
```

```
Collecting replicate
  Downloading replicate-1.0.7-py3-none-any.whl.metadata (29 kB)
Requirement already satisfied: httpx<1,>=0.21.0 in /usr/local/lib/python3.11/dist-packages (from replicate) (0.28.1)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-packages (from replicate) (25.0)
Requirement already satisfied: pydantic>1.10.7 in /usr/local/lib/python3.11/dist-packages (from replicate) (2.11.7)
Requirement already satisfied: typing_extensions>=4.5.0 in /usr/local/lib/python3.11/dist-packages (from replicate) (4.14.1)
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.21.0->replicate) (4.10.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.21.0->replicate) (2025.8.3)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.21.0->replicate) (1.0.9)
Requirement already satisfied: idna in /usr/local/lib/python3.11/dist-packages (from httpx<1,>=0.21.0->replicate) (3.10)
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-packages (from httpcore==1.*->httpx<1,>=0.21.0->replicate) (0.14.0)
Requirement already satisfied: annotated-types==0.6.0 in /usr/local/lib/python3.11/dist-packages (from pydantic>1.10.7->replicate) (0.6.0)
Requirement already satisfied: pydantic-core==2.33.2 in /usr/local/lib/python3.11/dist-packages (from pydantic>1.10.7->replicate) (2.33.2)
Requirement already satisfied: typing-inspection>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from pydantic>1.10.7->replicate) (0.4.0)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-packages (from anyio->httpx<1,>=0.21.0->replicate) (1.3.1)
Downloading replicate-1.0.7-py3-none-any.whl (48 kB)
48.6/48.6 kB 650.6 kB/s eta 0:00:00
Installing collected packages: replicate
Successfully installed replicate-1.0.7
```

```
!pip install sentence-transformers
```



```

207.5/207.5 MB 9.2 MB/s eta 0:00:00
Downloading nvidia_nccl_cu12-2.21.5-py3-none-manylinux2014_x86_64.whl (188.7 MB)
188.7/188.7 MB 7.1 MB/s eta 0:00:00
21.1/21.1 MB 62.2 MB/s eta 0:00:00
Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (21.1 MB)
Installing collected packages: nvidia-nvjitlink-cu12, nvidia-nccl-cu12, nvidia-curand-cu12, nvidia-cufft-cu12, nvidia-cuda-runtime-cu12
Attempting uninstall: nvidia-nvjitlink-cu12
Found existing installation: nvidia-nvjitlink-cu12 12.5.82
Uninstalling nvidia-nvjitlink-cu12-12.5.82:
Successfully uninstalled nvidia-nvjitlink-cu12-12.5.82
Attempting uninstall: nvidia-nccl-cu12
Found existing installation: nvidia-nccl-cu12 2.23.4
Uninstalling nvidia-nccl-cu12-2.23.4:
Successfully uninstalled nvidia-nccl-cu12-2.23.4
Attempting uninstall: nvidia-curand-cu12
Found existing installation: nvidia-curand-cu12 10.3.6.82
Uninstalling nvidia-curand-cu12-10.3.6.82:
Successfully uninstalled nvidia-curand-cu12-10.3.6.82
Attempting uninstall: nvidia-cufft-cu12
Found existing installation: nvidia-cufft-cu12 11.2.3.61
Uninstalling nvidia-cufft-cu12-11.2.3.61:
Successfully uninstalled nvidia-cufft-cu12-11.2.3.61
Attempting uninstall: nvidia-cuda-runtime-cu12
Found existing installation: nvidia-cuda-runtime-cu12 12.5.82
Uninstalling nvidia-cuda-runtime-cu12-12.5.82:
Successfully uninstalled nvidia-cuda-runtime-cu12-12.5.82
Attempting uninstall: nvidia-cuda-nvrtc-cu12
Found existing installation: nvidia-cuda-nvrtc-cu12 12.5.82
Uninstalling nvidia-cuda-nvrtc-cu12-12.5.82:
Successfully uninstalled nvidia-cuda-nvrtc-cu12-12.5.82
Attempting uninstall: nvidia-cuda-cupti-cu12
Found existing installation: nvidia-cuda-cupti-cu12 12.5.82
Uninstalling nvidia-cuda-cupti-cu12-12.5.82:
Successfully uninstalled nvidia-cuda-cupti-cu12-12.5.82
Attempting uninstall: nvidia-cublas-cu12
Found existing installation: nvidia-cublas-cu12 12.5.3.2
Uninstalling nvidia-cublas-cu12-12.5.3.2:
Successfully uninstalled nvidia-cublas-cu12-12.5.3.2
Attempting uninstall: nvidia-cusparse-cu12
Found existing installation: nvidia-cusparse-cu12 12.5.1.3
Uninstalling nvidia-cusparse-cu12-12.5.1.3:
Successfully uninstalled nvidia-cusparse-cu12-12.5.1.3
Attempting uninstall: nvidia-cudnn-cu12
Found existing installation: nvidia-cudnn-cu12 9.3.0.75
Uninstalling nvidia-cudnn-cu12-9.3.0.75:
Successfully uninstalled nvidia-cudnn-cu12-9.3.0.75
Attempting uninstall: nvidia-cusolver-cu12
Found existing installation: nvidia-cusolver-cu12 11.6.3.83
Uninstalling nvidia-cusolver-cu12-11.6.3.83:
Successfully uninstalled nvidia-cusolver-cu12-11.6.3.83
Successfully installed nvidia-cublas-cu12-12.4.5.8 nvidia-cuda-cupti-cu12-12.4.127 nvidia-cuda-nvrtc-cu12-12.4.127 nvidia-cuda-runtime-cu12-12.4.127

```

```

import os
import gradio as gr
import tempfile
from langchain_milvus import Milvus
from langchain_community.llms import Replicate
from langchain_community.document_loaders import TextLoader
from langchain.text_splitter import CharacterTextSplitter
from transformers import AutoTokenizer
from langchain.prompts import PromptTemplate
from langchain.chains import RetrievalQA, LLMChain
from langchain.chains.combine_documents.stuff import StuffDocumentsChain
from langchain_community.embeddings import HuggingFaceEmbeddings

# Replicate token and model setup
os.environ['REPLICATE_API_TOKEN'] = "r8_3HHrig1gtQF2CZjxSVBiyGXIHrx33Yc2ee1BI"
model_path = "ibm-granite/granite-3.3-8b-instruct"
tokenizer = AutoTokenizer.from_pretrained(model_path)
model = Replicate(model=model_path, replicate_api_token=os.environ['REPLICATE_API_TOKEN'])

```

```

# 📦 Temporary Milvus DB
db_file = tempfile.NamedTemporaryFile(prefix="milvus_", suffix=".db", delete=False).name
embeddings = HuggingFaceEmbeddings(model_name="all-MiniLM-L6-v2")

```

```

vector_db = Milvus(
    embedding_function=embeddings,
    connection_args={"uri": db_file},
    auto_id=True,
    index_params={"index_type": "AUTOINDEX"},
)

```

```

# 📄 Financial Literacy Content
filename = "finance_literacy.txt"
with open(filename, "w") as f:
    f.write("""

```

UPI: Unified Payments Interface (UPI) is a real-time payment system developed by NPCI. It allows users to link bank accounts and make instant payments.

Avoiding Online Scams: Never share your OTP, UPI PIN, or passwords. Beware of links promising free rewards. Use official banking apps only.

Interest Rates: Interest is the cost of borrowing money. Loans usually have an interest rate. Compare rates from different banks before applying.

Budgeting: Create a monthly budget by listing your income and expenses. Prioritize needs over wants. Track your spending to save money.

budgeting: Create a monthly budget, listing your income and expenses. Prioritize needs over wants. Track your spending to save money.

Digital Wallets: Apps like PhonePe, Google Pay, Paytm allow you to pay bills, recharge, and send money securely.

Safe Loan Practices: Always read terms before taking a loan. Choose loans with lower interest and no hidden fees.

How to use BHIM app: Download, register with your mobile number linked to your bank, set UPI PIN, and start sending or requesting money.

Fraud Protection: Use two-factor authentication. Don't save card details on unknown websites. Report suspicious activity to your bank immediately.

```
# Load & Split Content
loader = TextLoader(filename)
documents = loader.load()
splitter = CharacterTextSplitter.from_huggingface_tokenizer(
    tokenizer=tokenizer,
    chunk_size=tokenizer.model_max_length // 2,
    chunk_overlap=0,
)
texts = splitter.split_documents(documents)
for i, doc in enumerate(texts):
    doc.metadata["doc_id"] = i + 1
vector_db.add_documents(texts)

# Financial Literacy Prompt
template = """
You are a helpful Digital Financial Literacy Assistant.
User Question: {question}
Give a simple, culturally-inclusive explanation in easy language.
Include tips to stay safe, explain tools like UPI, wallets, loans, interest, scams, budgeting.
"""
prompt = PromptTemplate(template=template, input_variables=["question"])

# Chains
llm_chain = LLMChain(llm=model, prompt=prompt)
combine_chain = StuffDocumentsChain(llm_chain=llm_chain)

rag_chain = RetrievalQA(
    retriever=vector_db.as_retriever(),
    combine_documents_chain=combine_chain,
    return_source_documents=False
)

# Gradio Interface
def ask_finance_agent(query):
    try:
        response = rag_chain.run(query)
        return response
    except Exception as e:
        return f"❌ Error: {str(e)}"

iface = gr.Interface(
    fn=ask_finance_agent,
    inputs=gr.Textbox(label="Ask a question about digital finance", placeholder="e.g. How to use UPI?"),
    outputs=gr.Textbox(label="Answer"),
    title="AI Agent for Digital Financial Literacy",
    description="Ask about UPI, loans, scams, budgeting, and digital tools. This assistant helps you stay safe and informed.",
    theme="default"
)

iface.launch(share=True)
```



```
/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 se
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(

tokenizer_config.json:      9.93k/? [00:00<00:00, 278kB/s]

vocab.json:      777k/? [00:00<00:00, 1.66MB/s]

merges.txt:      442k/? [00:00<00:00, 7.55MB/s]

tokenizer.json:      3.48M/? [00:00<00:00, 13.4MB/s]

added_tokens.json: 100%                                207/207 [00:00<00:00, 4.67kB/s]

special_tokens_map.json: 100%                            801/801 [00:00<00:00, 29.7kB/s]
/tmp/ipython-input-3987640662.py:22: LangChainDeprecationWarning: The class `HuggingFaceEmbeddings` was deprecated in LangChain 0.2.2
  embeddings = HuggingFaceEmbeddings(model_name="all-MiniLM-L6-v2")
modules.json: 100%                                     349/349 [00:00<00:00, 21.5kB/s]

config_sentence_transformers.json: 100%                  116/116 [00:00<00:00, 12.2kB/s]

README.md:      10.5k/? [00:00<00:00, 955kB/s]

sentence_bert_config.json: 100%                        53.0/53.0 [00:00<00:00, 4.91kB/s]

config.json: 100%                                     612/612 [00:00<00:00, 66.7kB/s]

model.safetensors: 100%                               90.9M/90.9M [00:03<00:00, 36.6MB/s]

tokenizer_config.json: 100%                            350/350 [00:00<00:00, 38.6kB/s]

vocab.txt:      232k/? [00:00<00:00, 11.8MB/s]

tokenizer.json:      466k/? [00:00<00:00, 24.6MB/s]

special_tokens_map.json: 100%                          112/112 [00:00<00:00, 12.7kB/s]

config.json: 100%                                     190/190 [00:00<00:00, 17.5kB/s]
/usr/local/lib/python3.11/dist-packages/torch/nn/modules/module.py:1750: FutureWarning: `encoder_attention_mask` is deprecated and will
  return forward_call(*args, **kwargs)
/tmp/ipython-input-3987640662.py:75: LangChainDeprecationWarning: The class `LLMChain` was deprecated in LangChain 0.1.17 and will be
  llm_chain = LLMChain(llm=model, prompt=prompt)
/tmp/ipython-input-3987640662.py:76: LangChainDeprecationWarning: This class is deprecated. Use the `create_stuff_documents_chain` co
  combine_chain = StuffDocumentsChain(llm_chain=llm_chain)
/tmp/ipython-input-3987640662.py:78: LangChainDeprecationWarning: This class is deprecated. Use the `create_retrieval_chain` construc
  rag_chain = RetrievalQA(
Colab notebook detected. To show errors in colab notebook, set debug=True in launch()
* Running on public URL: https://c10a38c6115e602088.gradio.live
```

This share link expires in 1 week. For free permanent hosting and GPU upgrades, run `gradio deploy` from the terminal in the working directory.

# AI Agent for Digital Financial Literacy

Ask about UPI, loans, scams, budgeting, and digital tools. This assistant helps you stay safe and informed.

Ask a question about digital finance

how to set up upi

Clear

Submit

Answer

**\*\*Using PhonePe and Google Pay:\*\***

PhonePe and Google Pay are digital wallet apps that help you manage your finances easily and securely. Here's a simple guide on how to use them:

**\*\*PhonePe:\*\***

- \*\*Download and Register:\*\*** Download the PhonePe app from Google Play Store or Apple App Store. Register using your mobile number, which should be linked to your bank account.
- \*\*Link Bank Account:\*\*** Verify your bank account by entering the details asked. This could be through an OTP (One Time Password) sent to your registered mobile number or through net banking.