In [3]:
```
import torch
print(torch.cuda.is_available())
print(torch.cuda.device_count())
```

```
False
0
```

to znaczy gpu nie dostepne wiec pip install torch --index-url https://download.pytorch.org/whl/cpu

> pip install "langchain>=0.2.12" "langchain-community>=0.2.10" "langchain-huggingface>=0.0.3" sentence-transformers faiss-cpu pypdf transformers accelerate bedzoiemy korzystac z cpu

In [4]:
```
from langchain_community.document_loaders import PyPDFLoader
from langchain.text_splitter import RecursiveCharacterTextSplitter
from langchain.vectorstores import FAISS
from langchain.embeddings import HuggingFaceEmbeddings
from langchain.llms import HuggingFacePipeline
from langchain.chains import ConversationalRetrievalChain
from langchain.prompts import PromptTemplate
```

In [14]:
```
import os
from transformers import AutoTokenizer, AutoModelForSeq2SeqLM, pipeline
```

In [5]:
```
DATA_DIR = "."
EMB_MODEL_NAME = "sentence-transformers/paraphrase-multilingual-mpnet-base-v2"
QA_MODEL_NAME = "google/flan-t5-base"

SCORE_THRESHOLD = 0.38
TOP_K = 4
```

In [7]:
```
def load_pdfs_as_documents(pdf_dir: str):
    docs = []
    for fname in sorted(os.listdir(pdf_dir)):
        if fname.lower().endswith(".pdf"):
            loader = PyPDFLoader(os.path.join(pdf_dir, fname))
            docs.extend(loader.load())
    if not docs:
        raise FileNotFoundError(f"No PDFs found in folder {pdf_dir}.")
    return docs

def split_docs(docs, chunk_size=1000, chunk_overlap=150):
    splitter = RecursiveCharacterTextSplitter(chunk_size=chunk_size, chunk_overlap=
    return splitter.split_documents(docs)

def build_vectorstore(chunks):
    embeddings = HuggingFaceEmbeddings(model_name=EMB_MODEL_NAME, encode_kwargs={"n
    vectordb = FAISS.from_documents(chunks, embedding=embeddings)
    return vectordb

def make_llm(model_name: str):
```

```
    tok = AutoTokenizer.from_pretrained(model_name)
    mdl = AutoModelForSeq2SeqLM.from_pretrained(model_name)
    gen = pipeline(
        "text2text-generation",
        model=mdl,
        tokenizer=tok,
        max_new_tokens=512,
        temperature=0.2,
        repetition_penalty=1.05
    )
    return HuggingFacePipeline(pipeline=gen)

def format_sources(source_docs) -> str:
    lines = []
    for d in source_docs:
        src = os.path.basename(d.metadata.get("source", ""))
        page = int(d.metadata.get("page", 0)) + 1
        lines.append(f"- {src}, p.{page}")
    lines = list(dict.fromkeys(lines))
    return "\n".join(lines)
```

Ten kod pobiera wszystkie pliki PDF z folderu i przekształca je w dokumenty, a następnie dzieli je na mniejsze fragmenty, aby ułatwić pracę z nimi. Następnie tworzy magazyn wektorowy dla tych fragmentów tekstu, aby wyszukać podobne fragmenty, i ładuje model języka, aby odpowiedzieć na pytania. Dostępna jest również funkcja, która wyraźnie pokazuje, z których plików PDF i stron pochodzą informacje.

In [11]:
```
docs = load_pdfs_as_documents(DATA_DIR)
chunks = split_docs(docs, chunk_size=1200, chunk_overlap=200)
vectordb = build_vectorstore(chunks)

retriever = vectordb.as_retriever(
    search_type="similarity_score_threshold",
    search_kwargs={"score_threshold": SCORE_THRESHOLD, "k": TOP_K})
```

```
C:\Users\sofiy\AppData\Local\Temp\ipykernel_26600\450368898.py:16: LangChainDeprecat
ionWarning: 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:`~langc
hain-huggingface package and should be used instead. To use it run `pip install -U :
class:`~langchain-huggingface` and import as `from :class:`~langchain_huggingface im
port HuggingFaceEmbeddings``.
  embeddings = HuggingFaceEmbeddings(model_name=EMB_MODEL_NAME, encode_kwargs={"norm
alize_embeddings": True})
c:\Python312\Lib\site-packages\tqdm\auto.py:21: TqdmWarning: IProgress not found. Pl
ease update jupyter and ipywidgets. See https://ipywidgets.readthedocs.io/en/stable/
user_install.html
  from .autonotebook import tqdm as notebook_tqdm
C:\Users\sofiy\AppData\Roaming\Python\Python312\site-packages\huggingface_hub\file_d
ownload.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:\U
sers\sofiy\.cache\huggingface\hub\models--sentence-transformers--paraphrase-multilin
gual-mpnet-base-v2. Caching files will still work but in a degraded version that mig
ht 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://hu
ggingface.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: h
ttps://docs.microsoft.com/en-us/windows/apps/get-started/enable-your-device-for-deve
lopment
  warnings.warn(message)
Xet Storage is enabled for this repo, but the 'hf_xet' package is not installed. Fal
ling back to regular HTTP download. For better performance, install the package wit
h: `pip install huggingface_hub[hf_xet]` or `pip install hf_xet`
Xet Storage is enabled for this repo, but the 'hf_xet' package is not installed. Fal
ling back to regular HTTP download. For better performance, install the package wit
h: `pip install huggingface_hub[hf_xet]` or `pip install hf_xet`
```

Ten fragment po prostu uruchamia wcześniej napisane funkcje

In [12]:
```
prompt_template = """
You are a careful scientific assistant. Answer ONLY based on the 'Context'.
If the answer is not in the context — honestly say that there is not enough informa

Chat history:
{chat_history}

User question:
{question}

Context:
{context}

Answer:
"""
prompt = PromptTemplate(template=prompt_template, input_variables=["chat_history",
```

In [16]:
```
from typing import List, Tuple
chat_history: List[Tuple[str, str]] = []
```

In [17]:
```
llm = make_llm(QA_MODEL_NAME)

qa_chain = ConversationalRetrievalChain.from_llm(
    llm=llm,
    retriever=retriever,
    combine_docs_chain_kwargs={"prompt": prompt},
    return_source_documents=True,
    verbose=False)
chat_history: List[Tuple[str, str]] = []
```

```
Device set to use cpu
```

In [18]:
```
def ask(question: str):
    candidate_docs = retriever.get_relevant_documents(question)
    if not candidate_docs:
        print("I couldn't find enough content in the database to answer. Try refini
        return

    res = qa_chain.invoke({"question": question, "chat_history": chat_history})
    answer = res["answer"].strip()
    sources = res.get("source_documents", []) or candidate_docs
    cites = format_sources(sources)

    print("\nAnswer:\n" + answer)
    print("\nCitations:\n" + (cites if cites else "- (no relevant sources found)"))
    chat_history.append((question, answer))
```

In [ ]:
```
print("RAG ready. Ask question about pdf like this: ask('your question')")
```

```
RAG ready. Ask question about pdf like this: ask('your question')
```

In [21]:
```
import warnings
warnings.filterwarnings("ignore", category=UserWarning)
warnings.filterwarnings("ignore", category=DeprecationWarning)
```

In [25]:
```
ask("What is late-life well-being")
```

```
Answer:
social orienta- tion is associated with late-life well-being levels, rates of termin
al decline, and delayed onset of terminal decline

Citations:
- pag-pag0000072.pdf, p.8
- pag-pag0000072.pdf, p.13
- pag-pag0000072.pdf, p.2
```

In [24]:
```
ask("Are family goals important for late-life well-being according to this study?")
```

Answer:
it is not entirely clear whether such effects are maintained as individuals enter th
e phase of terminal decline

Citations:
- pag-pag0000072.pdf, p.3
- pag-pag0000072.pdf, p.7
- pag-pag0000072.pdf, p.9
- pag-pag0000072.pdf, p.4

In [27]:
```
ask("Can online gaming increase feelings of loneliness?")
```

Answer:
Yes

Citations:
- s40359-025-02892-w.pdf, p.10
- s40359-025-02892-w.pdf, p.8
- s40359-025-02892-w.pdf, p.9

In [28]:
```
ask("Who is more interested in sports, boys or girls?")
```

Answer:
boys and men are substantially more interested than girls and women in sports, both
in terms of participation and spectating. Moreover, the evidence suggests that this
sex difference in sports interest oc- curs in all or nearly all societies. We also s
howed that there is mounting evidence for a reliable sex difference in sports motiva
tion, with males typically showing greater compet- itiveness and risk taking. In thi
s section, we demonstrate that there is unambiguous evidence for a substantial sex d
ifference in sports participation and spectatorship. We also review several lines of
evidence indicating that these patterns reflect a sex difference in underlying sport
s interest, not merely differences in opportunities for en- gagement. In the second
half of the review, we applied an evolutionary perspective to explain— both in terms
of functional and proximate causal hypotheses—why females are interested in sport, y
et relatively less interested than males.

Citations:
- ebs-ebs0000049.pdf, p.16
- ebs-ebs0000049.pdf, p.2
- ebs-ebs0000049.pdf, p.4
- ebs-ebs0000049.pdf, p.1

In [29]:
```
ask("What is competitiveness?")
```

Answer:
Humans can compete, for instance, by driving a sports car, starting a false rumor, i
nsulting a competitor’s hairstyle, or eating all the cookies before a sibling has a
chance to enter the kitchen. Another important aspect of human competi- tion is that
men and women differ in their use of competitive modes, especially in intrasexual or
same sex competition. Perhaps most notably, men are more likely than women to use hi
gh- stakes physical aggression (e.g., fighting), whereas women’s aggression more fre
quently involves indirect or relational tactics, such as gossiping (Benenson, 2013;
Campbell, 2002).

Citations:
- ebs-ebs0000049.pdf, p.1

In [33]:
```
ask("How does loneliness influence the link between social network addiction and su
```

```
Answer:
heightened loneliness are more inclined to use social media and online games as copi
ng mecha - nisms. This tendency, however, may increase digital dependence and ultima
tely elevate suicidal ideation risk. Our findings support the hypothesis that loneli
ness mod- erates the correlation between social network addiction, online gaming add
iction, and suicidal ideation, positioning loneliness as a central factor in this pa
thway. While social networks and online games are often promoted as tools for social
connection, our results suggest that their excessive use can weaken real-world relat
ionships and lead to intensified loneliness. This out - come aligns with previous re
search indicating this dynamic reflects the experience of thwarted belong - ingness,
which is directly tied to suicidal ideation. Our findings provide support for the hy
pothesis that loneliness mediates the relationship between social network addiction,
online gaming addiction, and suicidal ideation, positioning loneliness as a central
factor in this pathway. While social networks and online games are often promoted as
tools for social connection, our findings suggest that their excessive use can weake
n real-world relationships and lead to intensified loneliness. This out - come align
s with previous research indicating this dynamic reflects the experience of thwarted
belong - ingness, which is directly tied to suicidal ideation. Our findings provide
support for the hypothesis that loneliness mediates the relationship between social
network addiction, online gaming addiction, and suicidal ideation, positioning lonel
iness as a central factor in this pathway. While social networks and online games ar
e often promoted as tools for social connection, our findings suggest that their exc
essive use can weaken real-world relationships and lead to intensified loneliness. T
his out - come aligns with previous research indicating this dynamic reflects the ex
perience of thwarted belong - ingness, which is directly tied to suicidal ideation.
Our findings provide support for the hypothesis that loneliness mediates the relatio
nship between social net - work addiction, online gaming addiction, and suicidal ide
ation, positioning loneliness as a central factor in this pathway. While social netw
orks and online games are often promoted as tools for social connection, our finding
s suggest that their excessive use can weaken real-world relationships and lead to i
ntensified loneliness, positioning loneliness as a central factor in this pathway.

Citations:
- s40359-025-02892-w.pdf, p.10
- s40359-025-02892-w.pdf, p.9
- s40359-025-02892-w.pdf, p.3
```

In [ ]:

conclusion

model RAG PDF, potrafi czytać PDF z artykułami naukowymi, które dodałam, i odpowiadać na pytania na ich podstawie. Model „rozumie" treść artykułów i odpowiada na temat, czasem zaprosto, czasem trochę trudniej, ale generalnie jest zrozumiały.

Krótko o artykułach:

Terminal Decline in Well-Being - o tym, jak aktywne życie społeczne i wartości relacji wpływają na poczucie dobrostanu w późnym wieku.

Social Network and Online Gaming Addiction - o związku uzależnienia od mediów społecznościowych i gier online z samotnością i myślami samobójczymi.

Sex Differences in Sports Interest - o różnicach między mężczyznami a kobietami w zainteresowaniu sportem i motywacji do niego z perspektywy ewolucyjnej.

Na pierwszy artykuł model odpowiadał trochę gorzej, natomiast na drugi i trzeci radzi sobie bardzo dobrze - szczegółowo opisuje zależności i różnice, np. w uzależnieniach czy zainteresowaniu sportem. Ogólnie model działa i daje sensowne odpowiedzi, choć można go jeszcze ulepszyć.