AI Course Copilot – Final Prototype (Phase 05)

Team: Oscar Cortez, Judith Barrios, Ruben Valenzuela Alvarado

Course: ITAI-2277 – Artificial Intelligence

Description:

This notebook contains the working RAG prototype, final integration,

documentation links, and instructions for reproduction as required for Phase 05.

📦 Complete Project Package – Index

This project package includes:

1. Source Code

- RAG pipeline implementation
- Retrieval system (FAISS)
- Embedding generation
- Re-ranking
- Gradio UI prototype

2. Documentation

- Technical Report (Phase 03)
- System Integration Summary (Phase 04)
- User Guide (Phase 05)
- Architecture diagrams

3. Data Artifacts

- Embeddings
- Vector index
- Chunked course documents

4. Prototype

- Fully working Gradio interface
- Query → Retrieval → Answer pipeline

5. Reproducibility Files

- requirements.txt
- config.yaml
- Instructions for running the system

▶️ How to Run the Prototype

1. Install dependencies
2. Load FAISS index
3. Initialize retriever
4. Run Gradio interface

All steps are included in this notebook for testing and evaluation.

```
!pip install sentence-transformers faiss-cpu gradio
```

```
Requirement already satisfied: sentence-transformers in /usr/local/lib/python3
Collecting faiss-cpu
  Downloading faiss_cpu-1.13.0-cp39-abi3-manylinux_2_27_x86_64.manylinux_2_28_
Requirement already satisfied: gradio in /usr/local/lib/python3.12/dist-packag
Requirement already satisfied: transformers<5.0.0,>=4.41.0 in /usr/local/lib/p
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-packages
Requirement already satisfied: torch>=1.11.0 in /usr/local/lib/python3.12/dist
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.12/dist-
Requirement already satisfied: scipy in /usr/local/lib/python3.12/dist-package
Requirement already satisfied: huggingface-hub>=0.20.0 in /usr/local/lib/pytho
Requirement already satisfied: Pillow in /usr/local/lib/python3.12/dist-packag
Requirement already satisfied: typing_extensions>=4.5.0 in /usr/local/lib/pyth
Requirement already satisfied: numpy<3.0,>=1.25.0 in /usr/local/lib/python3.12
Requirement already satisfied: packaging in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: aiofiles<25.0,>=22.0 in /usr/local/lib/python3.
Requirement already satisfied: anyio<5.0,>=3.0 in /usr/local/lib/python3.12/di
Requirement already satisfied: brotli>=1.1.0 in /usr/local/lib/python3.12/dist
Requirement already satisfied: fastapi<1.0,>=0.115.2 in /usr/local/lib/python3
Requirement already satisfied: ffmpy in /usr/local/lib/python3.12/dist-package
Requirement already satisfied: gradio-client==1.13.3 in /usr/local/lib/python3
Requirement already satisfied: groovy~=0.1 in /usr/local/lib/python3.12/dist-p
Requirement already satisfied: httpx<1.0,>=0.24.1 in /usr/local/lib/python3.12
Requirement already satisfied: jinja2<4.0 in /usr/local/lib/python3.12/dist-pa
Requirement already satisfied: markupsafe<4.0,>=2.0 in /usr/local/lib/python3.
Requirement already satisfied: orjson~=3.0 in /usr/local/lib/python3.12/dist-p
Requirement already satisfied: pandas<3.0,>=1.0 in /usr/local/lib/python3.12/d
Requirement already satisfied: pydantic<2.12,>=2.0 in /usr/local/lib/python3.1
Requirement already satisfied: pydub in /usr/local/lib/python3.12/dist-package
Requirement already satisfied: python-multipart>=0.0.18 in /usr/local/lib/pyth
Requirement already satisfied: pyyaml<7.0,>=5.0 in /usr/local/lib/python3.12/d
Requirement already satisfied: ruff>=0.9.3 in /usr/local/lib/python3.12/dist-p
Requirement already satisfied: safehttpx<0.2.0,>=0.1.6 in /usr/local/lib/pytho
Requirement already satisfied: semantic-version~=2.0 in /usr/local/lib/python3
Requirement already satisfied: starlette<1.0,>=0.40.0 in /usr/local/lib/python
Requirement already satisfied: tomlkit<0.14.0,>=0.12.0 in /usr/local/lib/pytho
Requirement already satisfied: typer<1.0,>=0.12 in /usr/local/lib/python3.12/d
Requirement already satisfied: uvicorn>=0.14.0 in /usr/local/lib/python3.12/di
Requirement already satisfied: fsspec in /usr/local/lib/python3.12/dist-packag
Requirement already satisfied: websockets<16.0,>=13.0 in /usr/local/lib/python
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.12/dist-
Requirement already satisfied: certifi in /usr/local/lib/python3.12/dist-packa
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.12/dist
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.12/dist-pac
Requirement already satisfied: filelock in /usr/local/lib/python3.12/dist-pack
Requirement already satisfied: requests in /usr/local/lib/python3.12/dist-pack
Requirement already satisfied: hf-xet<2.0.0,>=1.1.3 in /usr/local/lib/python3.
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dis
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python
Requirement already satisfied: pydantic-core==2.33.2 in /usr/local/lib/python3
Requirement already satisfied: typing-inspection>=0.4.0 in /usr/local/lib/pyth
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-pa
Requirement already satisfied: sympy>=1.13.3 in /usr/local/lib/python3.12/dist
Requirement already satisfied: networkx>=2.5.1 in /usr/local/lib/python3.12/di
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.6.77 in /usr/local/l
```

```
from sentence_transformers import SentenceTransformer
import faiss
import numpy as np
import gradio as gr
import textwrap

# === Real mini-corpus for ITAI-2277 (Capstone) ===
documents = [
    {
        "id": "doc1",
        "course": "ITAI 2277",
        "week": "Course Overview",
        "doc_type": "Syllabus",
        "title": "ITAI 2277 – Course Vision and Welcome",
        "text": """
Houston Community College’s vision is to deliver relevant, high-quality educati
success for all students, the community, and the economy. ITAI 2277 – Artificia
(Capstone) is taught fully online (WW – Online Anytime).

The course is led by Professor Anna Devarakonda (Annapurna Rachapudi). The clas
an AI Applications Capstone Project where students design and deploy real-worlc
domains such as healthcare, finance, sustainability, and more. Students work wi
TensorFlow, PyTorch, and cloud platforms, and practice AI application developme
and professional collaboration.

The course does not use a traditional textbook. Instead, all instructional mate
through Canvas, using curated, up-to-date articles, papers, videos, and resourc
encouraged to check the course site and their HCC email at least once per day a
the Modules section in Canvas.
"""
    },
    {
        "id": "doc2",
        "course": "ITAI 2277",
        "week": "Course Requirements",
        "doc_type": "Syllabus",
        "title": "ITAI 2277 – Assignments, Grading and Workload",
        "text": """
ITAI 2277 uses multiple graded components:

- Module Group Assignments (15%): change according to the lecture topic. Format
  written Word/PDF documents, PowerPoint presentations, or multimedia submissic

- Case Study Analysis (20%): tied to lecture topics and focused on applying cor

- Exams/Quizzes (20%): 4 separate assessments covering key concepts, using mult
  true/false, and short-answer questions.

- Midterm (20%): group case study analysis that tests analytical skills relatec
  philosophical, and practical applications of AI in a specific industry.

- Final Project (25%): students, working in groups, create a proposal for integ
  new or existing process within a chosen sector. This is the main capstone-sty

There is also an optional Extra Credit Portfolio (5%) for uploading course work
continue building an AI portfolio.

The HCC grading system uses A (90–100), B (80–89), C (70–79), D (60–69), F or F
W (withdrawn), and I (incomplete), according to standard HCC policies.
"""
    },
    {
        "id": "doc3",
        "course": "ITAI 2277",
        "week": "Policies",
        "doc_type": "Syllabus",
        "title": "ITAI 2277 – Incompletes, Attendance, Make-Up Work, Academic I
        "text": """
Incomplete grades (“I”) are only considered if the student has completed at lea
in the course, and the instructor still has the discretion to decline the reque

Make-up exams and assignments are allowed only for documented emergencies, such
or auto accidents. They do not apply to reasons like forgetting the due date or
Documentation must be provided as soon as possible. All missed grades are recor
no approved make-up is arranged.

Online students must show satisfactory progress in the course. Students may be
miss turning in assignments that total more than 12.5% of the course work befor
Students are responsible for contacting the instructor if they are having a pro

Academic Integrity: Scholastic dishonesty results in referral to the Dean of St
Group work is allowed, but groups must not share the same files and then make m
submit as their own. Using copied work or unauthorized collaboration may result
assignment and a disciplinary referral. Students must follow HCC academic integ
"""
    },
    {
        "id": "doc4",
        "course": "ITAI 2277",
        "week": "Final Project",
        "doc_type": "Assignment",
        "title": "ITAI 2277 – Capstone Final Project and Presentation",
        "text": """
The Capstone Final Assignment for ITAI 2277 is titled “Capstone Project 2025.”
develop, and submit a GitHub repository as their final class project, along wit

This capstone course allows students to synthesize knowledge from the entire As
Applied Technologies – AI and Robotics. Working in teams, students build a subs
integrates multiple technologies from computer vision, natural language process
machine learning, deep learning, and related areas.

By the end of the course, students should be able to:
1. Plan and execute a comprehensive project that integrates multiple AI/ML tech
2. Apply knowledge from core courses to implement solutions to real-world probl
3. Collaborate using industry-standard tools and professional practices.
4. Document and communicate technical work through professional reports and pre
5. Evaluate and refine solutions using technical metrics, stakeholder requireme
6. Deliver a complete, portfolio-ready project that demonstrates readiness for

Final presentation requirements include:
- A public GitHub repo with a clear README and installation instructions.
- A formal project presentation (around 20 minutes per team).
- A live demonstration and Q&A session showcasing the system and its impact.
"""
    },
]
```

```
# load model
model = SentenceTransformer("sentence-transformers/all-MiniLM-L6-v2")

# Convert text embedding
corpus_texts = [doc["text"] for doc in documents]
embeddings = model.encode(corpus_texts, convert_to_numpy=True, show_progress_ba

# Create faiss
dim = embeddings.shape[1]
index = faiss.IndexFlatL2(dim)
index.add(embeddings)

print("Index size:", index.ntotal)
```

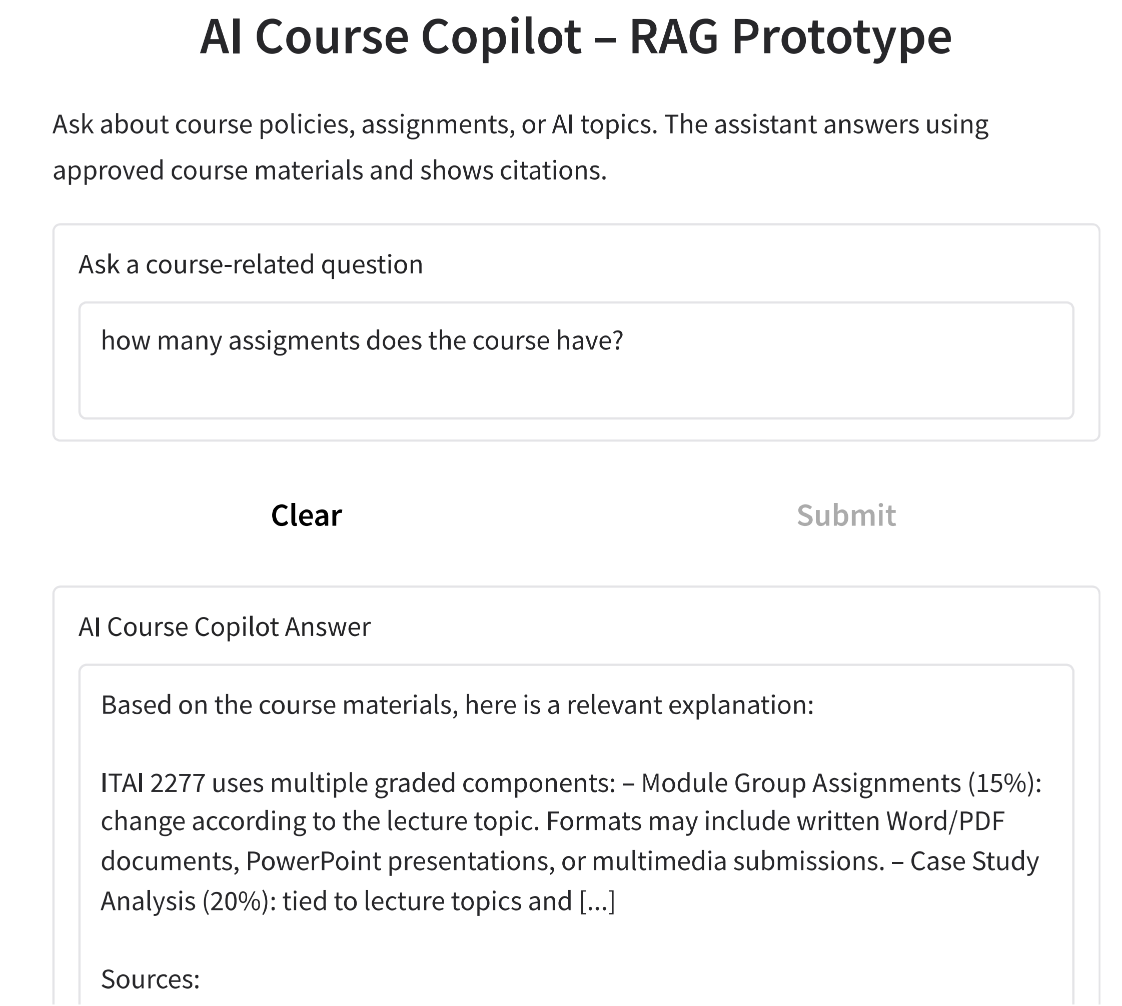
```
/usr/local/lib/python3.12/dist-packages/huggingface_hub/utils/_auth.py:94: UserW
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access publ
  warnings.warn(
modules.json:  100%                                     349/349  [00:00<00:00,  30.1kB/s]
config_sentence_transformers.json:  100%                116/116  [00:00<00:00,  11.5kB/s]
README.md:        10.5k/?  [00:00<00:00,  612kB/s]
sentence_bert_config.json:  100%                        53.0/53.0  [00:00<00:00,  5.49kB/s]
config.json:  100%                                      612/612  [00:00<00:00,  66.4kB/s]
model.safetensors:  100%                                90.9M/90.9M  [00:02<00:00,  51.5MB/s]
tokenizer_config.json:  100%                            350/350  [00:00<00:00,  33.0kB/s]
vocab.txt:        232k/?  [00:00<00:00,  9.38MB/s]
tokenizer.json:        466k/?  [00:00<00:00,  26.1MB/s]
special_tokens_map.json:  100%                          112/112  [00:00<00:00,  12.4kB/s]
config.json:  100%                                      190/190  [00:00<00:00,  18.1kB/s]
Batches:  100%                                          1/1  [00:00<00:00,   1.58it/s]
Index size: 4
```

```
def retrieve(query, k=3):
    """Return top-k docs for a query with similarity scores."""
    query_emb = model.encode([query], convert_to_numpy=True)
    distances, indices = index.search(query_emb, k)
    results = []
    for rank, (idx, dist) in enumerate(zip(indices[0], distances[0]), start=1):
        doc = documents[idx]
        results.append({
            "rank": rank,
            "score": float(dist),
            "id": doc["id"],
            "title": doc["title"],
            "course": doc["course"],
            "week": doc["week"],
            "doc_type": doc["doc_type"],
            "snippet": textwrap.shorten(doc["text"].replace("\n", " "), width=2
        })
    return results


def format_answer(query, k=3):
    """Generate a simple answer using retrieved snippets + citations."""
    results = retrieve(query, k=k)
    if not results:
        return "I couldn't find any relevant course materials for this question

    # Parte tipo "respuesta" (extractive)
    top = results[0]
    answer_intro = (
        f"Based on the course materials, here is a relevant explanation:\n\n"
        f"{top['snippet']}\n\n"
    )

    # Citations
    citations_lines = []
    for r in results:
        citations_lines.append(
            f"[{r['rank']}] {r['title']} "
            f"({r['course']}, {r['week']}, {r['doc_type']})"
        )
    citations_text = "Sources:\n" + "\n".join(citations_lines)

    return answer_intro + citations_text


# Prueba rápida en la consola
print(format_answer("What is the late work policy in ITAI 1370?"))
```

```
Based on the course materials, here is a relevant explanation:

ITAI 2277 uses multiple graded components: – Module Group Assignments (15%): cha

Sources:
[1] ITAI 2277 – Assignments, Grading and Workload (ITAI 2277, Course Requirement
[2] ITAI 2277 – Incompletes, Attendance, Make-Up Work, Academic Integrity (ITAI
[3] ITAI 2277 – Capstone Final Project and Presentation (ITAI 2277, Final Projec
```

```
def rag_chat(query):
    return format_answer(query, k=3)

demo = gr.Interface(
    fn=rag_chat,
    inputs=gr.Textbox(lines=2, label="Ask a course-related question"),
    outputs=gr.Textbox(lines=12, label="AI Course Copilot Answer"),
    title="AI Course Copilot – RAG Prototype",
    description="Ask about course policies, assignments, or AI topics. The assi
)

demo.launch()
```

```
It looks like you are running Gradio on a hosted Jupyter notebook, which require

Colab notebook detected. To show errors in colab notebook, set debug=True in lau
* Running on public URL: https://6739650c6da26ebec3.gradio.live

This share link expires in 1 week. For free permanent hosting and GPU upgrades,
```



AI Course Copilot – RAG Prototype

Ask about course policies, assignments, or AI topics. The assistant answers using approved course materials and shows citations.

Ask a course-related question

how many assigments does the course have?

AI Course Copilot Answer

Based on the course materials, here is a relevant explanation:

ITAI 2277 uses multiple graded components: – Module Group Assignments (15%): change according to the lecture topic. Formats may include written Word/PDF documents, PowerPoint presentations, or multimedia submissions. – Case Study Analysis (20%): tied to lecture topics and [...]

Sources:

Next steps: Deploy to Cloud Run