01_agentic_wikipedia_aimpoint_interview_project (2)

February 21, 2026

0.1 # Aimpoint Digital AI Engineering Assignment

0.2 Objective

Your assignment is to design, build, and explain a novel agentic workflow that utilizes a subset of the Wikipedia dataset. As part of this, you will need to define a distinctive GenAI use case that your system is intended to solve. The aim is to showcase not just your technical implementation skills, but also your ability to apply agentic system design innovatively and practically. You will implement your workflow in the Databricks Free Edition, starting from the provided notebook `01_agentic_wikipedia_aimpoint_interview.ipynb`.

To get you started, we pre-installed LangChain and LangGraph which are open source GenAI orchestration frameworks that work well in a Databricks workspace. In addition, we have provided you with a basic setup to access the data source using a LangChain dataloader (https://python.langchain.com/docs/integrations/document_loaders/wikipedia/).

You may use coding assistants for this assignment, but you must provide your own custom prompts and demonstrate your own critical thinking. Large language models must not be used to generate responses for the open-response questions in Part B of this notebook.

Note: This assignment uses serverless clusters. At the time of creating this notebook, all components run successfully. However, you may need to address package dependency issues in the future to ensure your GenAI solution continues to function properly.

0.3 Deliverables

1. Reference Architecture
   - This should highlight your approach to addressing your use case or problem in either a pdf or image format; include technical agentic workflow details here.
2. Databricks Notebook(s)
   - Includes primary notebook `01_agentic_wikipedia_aimpoint_interview`.ipynb and any supplemental notebooks required to run the agent
   - In the `01_agentic_wikipedia_aimpoint_interview`.ipynb notebook complete the **GenAI Application Development** and **Reflection** sections. The GenAI Application Development section is where you add your own custom logic to create and run your agentic workflow. The Reflection section is writing a markdown response to answer the two questions.
   - To reduce your development time, we created the logic for you to have a FAISS vector store and made the LLM accessible as well.

- Before finalizing, make sure your code runs correctly by using "Run All" to validate functionality. Then go to "File" → "Export" → "HTML" to download as HTML file. Next, open this HTML file. Finally save as a PDF see instructions below. **Note: In your submissions this must be a PDF file format**
  Save HTML as PDF
  - Windows: (ctrl + P) → Save as PDF → Save
  - MacOS: (+ P) → Save as PDF → Save

0.4 Data Source

The Wikipedia Loader ingests documents from the Wikipedia API and converts them into LangChain document objects. The page content includes the first sections of the Wikipedia articles and the metadata is described in detail below.

Recommendation: If you are using the LangChain document loader we recommend filtering down to 10k or fewer documents. The `query_terms` argument below can be upated to update the search term used to search wikipedia. Make sure you update this based on the use case you defined.

In the metadata of the LangChain document object; we have the following information:

Column	Definition
title	The Wikipedia page title (e.g., "Quantum Computing").
summary	A short extract or condensed description from the page content.
source	The URL link to the original Wikipedia article.

[ ]:
```
%pip install -U -qqqq backoff databricks-langchain langgraph==0.5.3 uv␣
↪databricks-agents mlflow-skinny[databricks] chromadb sentence-transformers␣
↪langchain-huggingface langchain-chroma wikipedia faiss-cpu
dbutils.library.restartPython()
```

```
Note: you may need to restart the kernel using %restart_python ordbutils.library.restartPython
```

[ ]:
```
␣
↪#############################################################################################
###### Python Package Imports for this notebook                                    ␣
↪                    ######

␣
↪#############################################################################################

from langchain.document_loaders import WikipediaLoader
import faiss
from langchain_community.docstore.in_memory import InMemoryDocstore
from langchain_community.vectorstores import FAISS
# from langchain.embeddings import DatabricksEmbeddings

from databricks_langchain import (
    ChatDatabricks,
    DatabricksEmbeddings,
```

```
    UCFunctionToolkit,
    VectorSearchRetrieverTool,
)



␣
 ↪#############################################################################################
 ###### Config (Define LLMs, Embeddings, Vector Store, Data Loader specs)      ␣
 ↪                  ######
␣
 ↪#############################################################################################

# DataLoader Config
query_terms = ["sport", "football", "soccer", "basketball","baseball",␣
 ↪"track","swimming", "gymnastics"] #TODO: update to match your use case␣
 ↪requirements
max_docs = 10 #TODO: recommend starting with a smaller number for testing␣
 ↪purposes

# Retriever Config
k = 8 # number of documents to return
EMBEDDING_MODEL = "databricks-bge-large-en" # Embedding model endpoint name


# LLM Config
LLM_ENDPOINT_NAME = "databricks-meta-llama-3-1-8b-instruct" # Model Serving␣
 ↪endpoint name; other option see "Serving" under AI/ML tab (e.g.␣
 ↪databricks-gpt-oss-20b)


example_question = "describe soccer?"
```

[ ]:
```
␣
 ↪#############################################################################################
 ###### Wikipedia Data Loader                                                  ␣
 ↪                  ######
␣
 ↪#############################################################################################

docs = WikipediaLoader(query=query_terms, load_max_docs=max_docs).load() # Load␣
 ↪in documents from Wikipedia takes about 10 minutes for 1K articles

#############################################################################################
###### FAISS Retriever: Using DBX embedding model                             ␣
 ↪                  ######␣
 ↪#############################################################################################
```

```
# Define the embeddings and the FAISS vector store
embeddings = DatabricksEmbeddings(endpoint=EMBEDDING_MODEL) # Use to generate␣
 ↪embeddings
vector_store = FAISS.from_documents(docs, embeddings)

# Example of how to invoke the vector store
results = vector_store.similarity_search(
    "What is the most popular sport in the US?",
    k=k
)
for res in results:
    print(f"* {res.page_content} [{res.metadata}]")

#########################################################################################################
###### LLM: Using DBX Foundation Model                                                                  ␣
 ↪                  ######␣
 ↪#########################################################################################################

llm = ChatDatabricks(endpoint=LLM_ENDPOINT_NAME)

response = llm.invoke("What is the most popular sport in the US?")

print("\n",response.content)
```

```
* Simone Arianne Biles Owens (née Biles; born March 14, 1997) is an American
artistic gymnast. Her 11 Olympic medals and 30 World Championship medals make
her the most decorated gymnast in history. She is widely regarded as one of the
greatest gymnasts of all time, and one of the greatest female athletes in
history. With 11 Olympic medals, she is tied with Věra Čáslavská as the second-
most decorated female Olympic gymnast behind Larisa Latynina, and has the most
Olympic medals earned by a U.S. gymnast.
At the Olympic Games, Biles is a two-time gold medalist in the individual all-
around (2016, 2024). She is also a two-time champion on vault (2016, 2024), the
2016 champion and 2024 silver medalist on floor exercise, and a two-time bronze
medalist on balance beam (2016, 2020). Biles led the gold medal-winning United
States teams in 2016, dubbed the "Final Five," and in 2024, dubbed the "Golden
Girls". At the 2020 Summer Olympics, where she was favored to win at least four
of the six available gold medals, she withdrew from most of the competition
after the qualification round due to "the twisties", a temporary loss of air
awareness while performing twisting elements. She won a silver medal with the
United States team nicknamed the "Fighting Four".
At the World Championships, she is the most decorated - male or female -
artistic gymnast of all time with 30 total medals in which 23 of them are Gold.
Biles is a six-time individual all-around champion (2013, 2014, 2015, 2018, 2019
and 2023), six-time floor exercise champion (2013-2015, 2018-2019, 2023), and
four-time balance beam champion (2014-2015, 2019, 2023), all record-high totals.
```

She is also a two-time vault champion (2018-2019) and a member of a record-high five gold medal-winning United States teams (2014-2015, 2018-2019, 2023). She is also a four-time World silver medalist (2013-2014 and 2023 on vault, 2018 on uneven bars), a three-time World bronze medalist (2015 on vault, 2013 and 2018 on balance beam).
Domestically, Biles has won a record-high nine United States national all-around championships (2013-2016, 2018-2019, 2021, 2023-2024); her win in 2024 made her the oldest female gymnast to ever win the title. She is also a seven-time champion on vault, balance beam, and floor exercise, a two-time uneven bars champion, and the only woman to win all five gold medals in a single championships twice (2018, 2024).
Biles is the sixth woman to win an individual all-around title at both the Olympics and the World Championships and the first since Lilia Podkopayeva in 1996 to hold both titles simultaneously. She is the tenth female gymnast and first American female gymnast to win a World medal on every event, and the first female gymnast since Daniela Silivaș in 1988 to win a medal on every event at a single Olympics or World Championships. Biles is the originator of the most difficult skill on women's vault, balance beam, and floor exercise and the only gymnast to attempt each skill to date.
In 2022, President Joe Biden awarded her the Presidential Medal of Freedom. In 2023, she won her eighth U.S. Gymnastics title, breaking the 90-year-old U.S. Gymnastics title record previously held by Alfred Jochim. Biles has won the Laureus World Sportswoman of the Year four times (2017, 2019, 2020, 2025) and Comeback of the Year once (2024).

== Early life and education ==
Biles was born on March 14, 1997, in Columbus, Ohio, the third of four siblings. Her birth mother, Shanon Biles, was unable to care for Simone or her other children. All four went in and out of foster care.
In 2000, Biles's maternal grandfather, Ron Biles, and his second wife, Nellie Cayetano Biles, began caring temporarily for Shanon's children in the north Houston suburb of Spring, Texas, after learning his grandchildren were in foster care. In 2003, the couple formally adopted Simone and her younger sister Adria. Ron's sister, Shanon's aunt Harriet, adopted the two oldest children. Simone holds Belizean citizenship thro [{'title': 'Simone Biles', 'summary': 'Simone Arianne Biles Owens (née Biles; born March 14, 1997) is an American artistic gymnast. Her 11 Olympic medals and 30 World Championship medals make her the most decorated gymnast in history. She is widely regarded as one of the greatest gymnasts of all time, and one of the greatest female athletes in history. With 11 Olympic medals, she is tied with Věra Čáslavská as the second-most decorated female Olympic gymnast behind Larisa Latynina, and has the most Olympic medals earned by a U.S. gymnast.\nAt the Olympic Games, Biles is a two-time gold medalist in the individual all-around (2016, 2024). She is also a two-time champion on vault (2016, 2024), the 2016 champion and 2024 silver medalist on floor exercise, and a two-time bronze medalist on balance beam (2016, 2020). Biles led the gold medal-winning United States teams in 2016, dubbed the "Final Five," and in 2024, dubbed the "Golden Girls". At the 2020 Summer Olympics,

where she was favored to win at least four of the six available gold medals, she withdrew from most of the competition after the qualification round due to "the twisties", a temporary loss of air awareness while performing twisting elements. She won a silver medal with the United States team nicknamed the "Fighting Four".\nAt the World Championships, she is the most decorated - male or female - artistic gymnast of all time with 30 total medals in which 23 of them are Gold. Biles is a six-time individual all-around champion (2013, 2014, 2015, 2018, 2019 and 2023), six-time floor exercise champion (2013-2015, 2018-2019, 2023), and four-time balance beam champion (2014-2015, 2019, 2023), all record-high totals. She is also a two-time vault champion (2018-2019) and a member of a record-high five gold medal-winning United States teams (2014-2015, 2018-2019, 2023). She is also a four-time World silver medalist (2013-2014 and 2023 on vault, 2018 on uneven bars), a three-time World bronze medalist (2015 on vault, 2013 and 2018 on balance beam).\nDomestically, Biles has won a record-high nine United States national all-around championships (2013-2016, 2018-2019, 2021, 2023-2024); her win in 2024 made her the oldest female gymnast to ever win the title. She is also a seven-time champion on vault, balance beam, and floor exercise, a two-time uneven bars champion, and the only woman to win all five gold medals in a single championships twice (2018, 2024).\nBiles is the sixth woman to win an individual all-around title at both the Olympics and the World Championships and the first since Lilia Podkopayeva in 1996 to hold both titles simultaneously. She is the tenth female gymnast and first American female gymnast to win a World medal on every event, and the first female gymnast since Daniela Silivaș in 1988 to win a medal on every event at a single Olympics or World Championships. Biles is the originator of the most difficult skill on women\'s vault, balance beam, and floor exercise and the only gymnast to attempt each skill to date.\nIn 2022, President Joe Biden awarded her the Presidential Medal of Freedom. In 2023, she won her eighth U.S. Gymnastics title, breaking the 90-year-old U.S. Gymnastics title record previously held by Alfred Jochim. Biles has won the Laureus World Sportswoman of the Year four times (2017, 2019, 2020, 2025) and Comeback of the Year once (2024). \n\n', 'source':
'https://en.wikipedia.org/wiki/Simone_Biles'}]
* Gymnastics events have been contested at every Summer Olympic Games since the birth of the modern Olympic movement at the 1896 Summer Olympics in Athens. For 32 years, only men were allowed to compete. Beginning at the 1928 Summer Olympics in Amsterdam, women were allowed to compete in artistic gymnastics events as well. Rhythmic gymnastics events were introduced at the 1984 Summer Olympics in Los Angeles, and trampoline events were added at the 2000 Summer Olympics in Sydney.

== Summary ==

== Artistic gymnastics ==

=== Men's events ===

Past events

=== Women's events ===

Past events

Team, portable apparatus: 1952, 1956

=== Medal table ===
(1896-2024)

=== Nations ===
Nations competing in artistic gymnastics, and the number of gymnasts (male and female) each nation brought to each Olympics, are shown below.

== Rhythmic gymnastics ==

=== Events ===

=== Medal table ===
(1984-2024)

=== Nations ===
Nations competing in rhythmic gymnastics and the number of gymnasts each nation brought to each Olympics, are shown below.

== Trampoline ==

=== Events ===

=== Medal table ===
(2000-2024)

=== Nations ===
Nations competing in trampoline gymnastics and the number of gymnasts each nation brought to each Olympics, are shown below.

== Overall medal table ==
Last updated after the 2024 Summer Olympic Games

== See also ==

Gymnastics at the Youth Olympic Games
List of Olympic venues in gymnastics
World Gymnastics Championships
Gymnastics at the Alternate Olympics

== Notes ==
A.^ ^ ^ As ROC.

== References ==

== External links ==
Official Olympic Report (archived)
gymnasticsresults.com (archived)
gymn-forum.net [{'title': 'Gymnastics at the Summer Olympics', 'summary': 'Gymnastics events have been contested at every Summer Olympic Games since the birth of the modern Olympic movement at the 1896 Summer Olympics in Athens. For 32 years, only men were allowed to compete. Beginning at the 1928 Summer Olympics in Amsterdam, women were allowed to compete in artistic gymnastics events as well. Rhythmic gymnastics events were introduced at the 1984 Summer Olympics in Los Angeles, and trampoline events were added at the 2000 Summer Olympics in Sydney.\n\n', 'source': 'https://en.wikipedia.org/wiki/Gymnastics_at_the_Summer_Olympics'}]
* Gymnastics is a group of sport that includes physical exercises requiring balance, strength, flexibility, agility, coordination, artistry and endurance. The movements involved in gymnastics contribute to the development of the arms, legs, shoulders, back, chest, and abdominal muscle groups. Gymnastics evolved from exercises used by the ancient Greeks that included skills for mounting and dismounting a horse.
The most common form of competitive gymnastics is artistic gymnastics; for women, the events include floor, vault, uneven bars, and balance beam; for men, besides floor and vault, it includes rings, pommel horse, parallel bars, and horizontal bar.
The governing body for competition in gymnastics throughout the world is World Gymnastics. Eight sports are governed by the FIG, including gymnastics for all, men's and women's artistic gymnastics, rhythmic gymnastics (women's branch only), trampolining (including double mini-trampoline), tumbling, acrobatic, aerobic, parkour and para-gymnastics. Disciplines not currently recognized by

FIG include wheel gymnastics, aesthetic group gymnastics, TeamGym, men's rhythmic gymnastics (both the Spanish form which is identical to the women's version and the Japanese version which is a different sport) and mallakhamba. Participants in gymnastics-related sports include young children, recreational-level athletes, and competitive athletes at all skill levels.

== Etymology ==
The word gymnastics derives from the common Greek adjective (gymnos), by way of the related verb (gymnazo), whose meaning is to "train naked", "train in gymnastic exercise", generally "to train, to exercise". The verb had this meaning because athletes in ancient times exercised and competed without clothing.

== History ==

Gymnastics can be traced to exercises performed in Ancient Greece, specifically in Sparta and Athens. Exercise of that time was documented by Philostratus' work Gymnastics: The Ethics of an Athletic Aesthetic. The original term for the practice of gymnastics is from the related Greek verb (gumnázō), which translates as "to train naked or nude," because young men exercised without clothing. In ancient Greece, physical fitness was highly valued among both men and women. It was not until after the Romans conquered Greece in 146 BC that gymnastics became more formalized and was used to train men in warfare. On Philostratus' claim that gymnastics is a form of wisdom, comparable to philosophy, poetry, music, geometry, and astronomy, the people of Athens combined this more physical training with the education of the mind. At the Palestra, a physical education training center, the disciplines of educating the body and the mind were combined, allowing for a form of gymnastics that was more aesthetic and individual and that left behind the focus on strictness, discipline, the emphasis on defeating records, and a focus on strength.

Don Francisco Amorós y Ondeano-a Spanish colonel born on 19 February 1770, in Valencia, who died on 8 August 1848, in Paris-was the first person to introduce educative gymnastics in France. The German Friedrich Ludwig Jahn began the German gymnastics movement in 1811 in Berlin, which led to the invention of the parallel bars, rings, the horizontal bar, the pommel horse and the vault horse. Germans Charles Beck and Charles Follen and American John Neal brought the first wave of gymnastics to the United States in the 1820s. Beck opened the first gymnasium in the US in 1825 at the Round Hill School in Northampton, Massachusetts. Follen opened the first college gymnasium and the first public gymnasium in the US in 1826 at Harvard University and in Boston, Massachusetts, respectively. Neal was the first American to open a public gymnasium in the US, in Portland, Maine, in 1827. He also documented and promoted these early efforts in the American Journal of Education and The Yankee, helping to establish the American branch of the movement.

The Federation of Intern [{'title': 'Gymnastics', 'summary': "Gymnastics is a group of sport that includes physical exercises requiring balance, strength, flexibility, agility, coordination, artistry and endurance. The movements involved in gymnastics contribute to the development of the arms, legs, shoulders, back, chest, and abdominal muscle groups. Gymnastics evolved from exercises used by the ancient Greeks that included skills for mounting and dismounting a horse.\nThe most common form of competitive gymnastics is artistic gymnastics; for women, the events include floor, vault, uneven bars, and balance beam; for men, besides floor and vault, it includes rings, pommel horse, parallel bars, and horizontal bar.\nThe governing body for competition in gymnastics throughout the world is World Gymnastics. Eight sports are governed by the FIG, including gymnastics for all, men's and women's artistic gymnastics, rhythmic gymnastics (women's branch only), trampolining (including double mini-trampoline), tumbling, acrobatic, aerobic, parkour and para-gymnastics. Disciplines not currently recognized by FIG include wheel gymnastics, aesthetic group gymnastics, TeamGym, men's rhythmic gymnastics (both the Spanish form which is identical to the women's version and the Japanese version which is a different sport) and mallakhamba.\nParticipants in gymnastics-related sports include young children, recreational-level athletes, and competitive athletes at all skill levels.", 'source': 'https://en.wikipedia.org/wiki/Gymnastics'}]
* Rhythmic gymnastics is a sport in which gymnasts perform individually or in groups on a floor with an apparatus: hoop, ball, clubs, ribbon and rope. The sport combines elements of gymnastics, dance and calisthenics; gymnasts must be strong, flexible, agile, dexterous and coordinated. Rhythmic gymnastics is governed by World Gymnastics, which first recognized it as a sport in 1963. At the international level, rhythmic gymnastics is a women-only sport. Rhythmic gymnastics became an Olympic sport in 1984, when the individual all-around event was first competed, and the group competition was added to the Olympics in 1996. The most prestigious competitions, besides the Olympic Games, are the World Championships, World Games, European Championships, European Games, the World Cup Series and the Grand Prix Series. Gymnasts are judged on their artistry, execution of skills, and difficulty of skills, for which they gain points. They perform leaps, balances, and rotations (spins) along with handling the apparatus.

== History ==

=== Aesthetic gymnastics ===

Rhythmic gymnastics grew out of the ideas of Jean-Georges Noverre (1727–1810), François Delsarte (1811–1871), and Rudolf Bode (1881–1970), who all believed in movement expression, where one used to dance to express oneself and exercise various body parts. From 1834, Pehr Henrik Ling further developed this idea in his 19th-century Swedish system of free exercise, which promoted "aesthetic gymnastics", in which students expressed their feelings and emotions through body movement.

Swedish-style group gymnastics became increasingly popular for women from the mid-19th century through to the early 20th century. Although sports became associated with masculinity, group gymnastics were performed in indoor, private spaces and focused on correctly performing movements before an instructor, which fit societal ideals for women. Women's gymnastics began to focus on qualities perceived as feminine, such as grace and expressiveness.
Ling's ideas were extended by Catharine Beecher, who founded the Western Female Institute in Cincinnati, Ohio, United States, in 1837. She developed a program where pupils exercised to music, moving from simple calisthenics that could be done in a classroom to more strenuous activities. While she promoted the exercises as being for all children, she emphasized that girls were especially lacking in exercise and that their health suffered for it.

=== Harmonic gymnastics ===

François Delsarte created a system of movement which was focused on creating expressive acting with natural poses, but it became a popular form of women's gymnastics for developing grace. In 1885, an American student of Delsarte, Genevieve Stebbins, published her first book, The Delsarte System of Expression. She soon began to perform popular solo dances, and she went on to combine Delsarte's ideas with Ling's and to develop her own gymnastics system. Dubbed "harmonic gymnastics", it encouraged late nineteenth-century American women to engage in physical culture and expression, in defiance of traditional gender norms. Stebbins provided the means, rationale, and model for exercise for middle-class women.
During the 1880s, Émile Jaques-Dalcroze of Switzerland developed eurhythmics, a form of physical training for musicians and dancers. Robert Bode trained at the Dalcroze Eurythmic College and went on to found his own school. George Demeny of France created exercises to music that were designed to promote grace of movement, muscular flexibility, and good posture, and some exercises included apparatuses.
The dancer Isadora Duncan was significant in the development of rhythmic gymnastics. Influenced by Delsarte, Jaques-Dalcroze, and possibly by Stebbins, she developed her own theory of dance that departed from more rigid traditions like that of ballet. Her free dancing style incorporated running and jumping movements.

=== Modern gymnastics ===

In 1929, Hinrich Medau, who graduated from the Bode Sc [{'title': 'Rhythmic gymnastics', 'summary': 'Rhythmic gymnastics is a sport in which gymnasts perform individually or in groups on a floor with an apparatus: hoop, ball, clubs, ribbon and rope. The sport combines elements of gymnastics, dance and calisthenics; gymnasts must be strong, flexible, agile, dexterous and coordinated. Rhythmic gymnastics is governed by World Gymnastics, which first recognized it as a sport in 1963. At the international level, rhythmic

gymnastics is a women-only sport.\nRhythmic gymnastics became an Olympic sport in 1984, when the individual all-around event was first competed, and the group competition was added to the Olympics in 1996. The most prestigious competitions, besides the Olympic Games, are the World Championships, World Games, European Championships, European Games, the World Cup Series and the Grand Prix Series. Gymnasts are judged on their artistry, execution of skills, and difficulty of skills, for which they gain points. They perform leaps, balances, and rotations (spins) along with handling the apparatus.', 'source': 'https://en.wikipedia.org/wiki/Rhythmic_gymnastics'}]
* Gymnastics events have been staged at the Olympic Games since 1896. Since then, 30 female gymnasts have won at least five total medals. The country with the most athletes on this list is the Soviet Union, with nine. Romania (6), United States (6), Hungary (4), East Germany (2), Russia (2), Brazil (1), and Czechoslovakia (1) are also represented.
Nine female gymnasts have won at least eight medals at the Olympic Games: Larisa Latynina (18), Věra Čáslavská (11), Simone Biles (11), Ágnes Keleti (10), Polina Astakhova (10), Nadia Comăneci (9), Ludmilla Tourischeva (9), Margit Korondi (8) and Sofia Muratova (8).
Larisa Latynina and Polina Astakhova each competed for the Soviet Union in 1956, 1960, and 1964. Latynina has the most medals of any female athlete in Olympic history, with 18. She won six medals in each Olympic Games that she competed in, winning the individual all-around titles in 1956 and 1960. Astakhova won two medals in 1956, four medals in 1960, and four medals in 1964. She won the uneven bars golds in 1960 and 1964. Sofia Muratova was Latynina's and Astakhova's teammate in 1956 and 1960. Muratova won a total of eight medals. Ludmilla Tourischeva also competed for the Soviet Union. She won one medal in 1968, four medals in 1972, and four medals in 1976.
Ágnes Keleti and Margit Korondi both competed for Hungary in 1956 and 1960. Keleti won 10 medals, including two golds on floor exercise. Korondi won eight total medals.
Czechoslovakia's Věra Čáslavská won 11 total Olympic medals, the second-most of any female gymnast. She won one in 1960, four in 1964, and six in 1968. She won the individual all-around golds in 1964 and 1968. She holds the record for the most individual gold medals (with 7, all her golds are individual). She remains the only gymnast, male or female, to have won an Olympic gold medal in each individual event.
Nadia Comăneci, who competed for Romania in 1976 and 1980, won nine medals. In 1976, she became the first gymnast to earn a perfect 10 at the Olympic Games and eventually achieved that mark seven times during the Games. She also won the individual all-around gold that year.

== See also ==

List of multiple Olympic gold medalists
List of multiple Olympic medalists
List of Olympic medalists in gymnastics (women)
List of top female medalists at major artistic gymnastics events

List of female artistic gymnasts with the most appearances at Olympic Games

== References == [{'title': "List of Olympic medal leaders in women's gymnastics", 'summary': "Gymnastics events have been staged at the Olympic Games since 1896. Since then, 30 female gymnasts have won at least five total medals. The country with the most athletes on this list is the Soviet Union, with nine. Romania (6), United States (6), Hungary (4), East Germany (2), Russia (2), Brazil (1), and Czechoslovakia (1) are also represented.\nNine female gymnasts have won at least eight medals at the Olympic Games: Larisa Latynina (18), Věra Čáslavská (11), Simone Biles (11), Ágnes Keleti (10), Polina Astakhova (10), Nadia Comăneci (9), Ludmilla Tourischeva (9), Margit Korondi (8) and Sofia Muratova (8).\nLarisa Latynina and Polina Astakhova each competed for the Soviet Union in 1956, 1960, and 1964. Latynina has the most medals of any female athlete in Olympic history, with 18. She won six medals in each Olympic Games that she competed in, winning the individual all-around titles in 1956 and 1960. Astakhova won two medals in 1956, four medals in 1960, and four medals in 1964. She won the uneven bars golds in 1960 and 1964. Sofia Muratova was Latynina's and Astakhova's teammate in 1956 and 1960. Muratova won a total of eight medals. Ludmilla Tourischeva also competed for the Soviet Union. She won one medal in 1968, four medals in 1972, and four medals in 1976.\nÁgnes Keleti and Margit Korondi both competed for Hungary in 1956 and 1960. Keleti won 10 medals, including two golds on floor exercise. Korondi won eight total medals.\nCzechoslovakia's Věra Čáslavská won 11 total Olympic medals, the second-most of any female gymnast. She won one in 1960, four in 1964, and six in 1968. She won the individual all-around golds in 1964 and 1968. She holds the record for the most individual gold medals (with 7, all her golds are individual). She remains the only gymnast, male or female, to have won an Olympic gold medal in each individual event.\nNadia Comăneci, who competed for Romania in 1976 and 1980, won nine medals. In 1976, she became the first gymnast to earn a perfect 10 at the Olympic Games and eventually achieved that mark seven times during the Games. She also won the individual all-around gold that year.", 'source': 'https://en.wikipedia.org/wiki/List_of_Olympic_medal_leaders_in_women%27s_gymnastics'}]
* Artistic gymnastics is a discipline of gymnastics in which athletes perform short routines on different types of apparatus. The sport is governed by World Gymnastics, which assigns the Code of Points used to score performances and regulates all aspects of elite international competition. Within individual countries, gymnastics is regulated by national federations such as British Gymnastics and USA Gymnastics. Artistic gymnastics is a popular spectator sport at many competitions, including the Summer Olympic Games.

== History ==
The gymnastic system was mentioned in writings by ancient authors, including Homer, Aristotle, and Plato. It included many disciplines that later became independent sports, such as swimming, racing, wrestling, boxing, and horse riding. It was also used for military training.

Gymnastics evolved in Bohemia and what later became Germany at the beginning of the 19th century. The term "artistic gymnastics" was introduced to distinguish freestyle performances from those used by the military. The German educator Friedrich Ludwig Jahn, who was known as the father of gymnastics, invented several apparatus, including the horizontal bar and parallel bars. Two of the first gymnastics clubs were Turnvereins and Sokols.
The Fédération Internationale de Gymnastique (FIG), the French name for the International Gymnastics Federation, was founded in 1881 and remains the governing body of international gymnastics. The organization began with three countries and was called the European Gymnastics Federation until 1921, when the first non-European countries joined, and it was reorganized into its modern form.
Gymnastics was included in the 1896 Summer Olympics, but female gymnasts were not allowed to participate in the Olympics until 1928. The World Artistic Gymnastics Championships, held since 1903, were only open to men until 1934. Since then, two branches of artistic gymnastics have developed: women's artistic gymnastics (WAG) and men's artistic gymnastics (MAG). Unlike men's and women's branches of many other sports, WAG and MAG differ significantly in technique and apparatuses used at major competitions.

=== Women's artistic gymnastics ===
As a team event, women's gymnastics entered the Olympics in 1928 and the World Championships in 1950. Individual women were recognized in the all-around as early as the 1934 World Championships. The existing women's program-all-around and event finals on the vault, uneven bars, balance beam, and floor exercise-was introduced at the 1950 World Championships and at the 1952 Summer Olympics.
The earliest champions in women's gymnastics tended to be in their 20s, and most had studied ballet for years before entering the sport. Larisa Latynina, the first great Soviet gymnast, won her first Olympic all-around medal at age 22 and her second at 26; she became world champion in 1958 while pregnant. Věra Čáslavská of Czechoslovakia, who followed Latynina and became a two-time Olympic all-around champion, was 22 when she won her first Olympic gold medal. The rules for the 1952 and 1956 Summer Olympics specified that gymnasts must turn 18 the year of the competition, with female gymnasts turning 16 or 17 only allowed to compete with a doctor's clearance.
In the 1970s, the average age of Olympic gymnasts began to decrease. While it was not unheard of for teenagers to compete in the 1960s - Ludmilla Tourischeva was 16 at her first Olympics in 1968 - younger female gymnasts slowly became the norm as the sport's difficulty increased. Smaller, lighter girls generally excelled in the more challenging acrobatic elements required by the redesigned Code of Points. The 58th Congress of the FIG - held in July 1980, just before the Olympics - decided to raise the minimum age for senior international competition from 14 to 15. However, the change, which came into effect two years later, did not eliminate the problem. By the time of the 1992 Summer Olympics, elite gymnasts consisted almost exclusively of [{'title': 'Artistic gymnastics', 'summary': 'Artistic gymnastics is a discipline of gymnastics in which athletes perform short routines on different types of apparatus. The sport

is governed by World Gymnastics, which assigns the Code of Points used to score performances and regulates all aspects of elite international competition. Within individual countries, gymnastics is regulated by national federations such as British Gymnastics and USA Gymnastics. Artistic gymnastics is a popular spectator sport at many competitions, including the Summer Olympic Games.\n\n', 'source': 'https://en.wikipedia.org/wiki/Artistic_gymnastics'}]
* Beginning in the 1990s, hundreds of gymnasts-primarily minors-were sexually abused over two decades in the United States, which is considered the largest sexual abuse scandal in sports history. More than 500 athletes alleged that they were sexually assaulted by gym owners, coaches, and staff working for gymnastics programs across the country, including USA Gymnastics (USAG) and Michigan State University (MSU). Hundreds of them sued USAG, MSU, and the United States Olympic Committee (USOC, later USOPC), which settled the suits in 2018 and 2021 for a total of nearly $900 million.
The breadth of the abuses was first revealed by The Indianapolis Star, which reported in September 2016 that "predatory coaches were allowed to move from gym to gym, undetected by a lax system of oversight, or dangerously passed on by USA Gymnastics-certified gyms". Coaches and officials perpetrated, facilitated, or worked to conceal abuse in Michigan, Pennsylvania, California, Rhode Island, Indiana and elsewhere. FBI agents declined to investigate early allegations of abuse, then lied about it, according to a U.S. Justice Department report. Dozens of officials at USAG, MSU, and the United States Olympic Committee (USOC, later USOPC) ultimately resigned under pressure or had their contracts terminated. Several coaches and officials faced criminal charges, though few were convicted. A central figure was Larry Nassar, a longtime national-team doctor for USAG and osteopathic physician in MSU's athletic department. More than 265 women said Nassar had sexually abused them under the pretense of providing medical treatment, including former USAG national team members Jessica Howard, Jamie Dantzscher, Morgan White, Jeanette Antolin, McKayla Maroney, Aly Raisman, Maggie Nichols, Gabby Douglas, Simone Biles, Jordyn Wieber, Sabrina Vega, Ashton Locklear, Kyla Ross, Madison Kocian, Amanda Jetter, Tasha Schwikert, Mattie Larson, Bailie Key, Kennedy Baker, Alyssa Baumann, and Terin Humphrey. In 2017 and 2018, Nassar pleaded guilty to federal charges of child pornography and state charges of first-degree sexual assault; he received sentences of 60 years in prison plus another 80 to 300 years. The scandal led to the Protecting Young Victims from Sexual Abuse and Safe Sport Authorization Act of 2017, which directed the creation of the U.S. Center for SafeSport.

== Origins ==
Since 1990, USA Gymnastics has kept a list of people permanently banned from coaching for sexual abuse and other reasons. The list includes Robert Dean Head, a USAG coach in Kentucky who in 1992 pled guilty to raping a 12-year-old, and Don Peters, the national coach for the 1984 Olympic team, who was banned in 2011 after two former gymnasts accused him of sexual abuse. In 2007, USAG began requiring background checks for coaches.
Yet USAG leaders also routinely dismissed warnings about coaches. For example, USAG received complaints about coach Mark Schiefelbein long before he was

convicted in 2003 of molesting a 10-year-old girl. Similarly, USAG received complaints about coach James Bell at least five years before he was jailed in 2003 for molesting three young gymnasts. In a 2013 lawsuit, USAG officials admitted under oath that allegations of sexual abuse were routinely dismissed as hearsay unless they came directly from a victim or victim's parent.
Even when USAG leaders believed the accusers, they sometimes allowed coaches to continue coaching for years. For example, USAG leaders waited four years before telling the police that they had received credible allegations of sexual assault by Marvin Sharp, who became a USAG coach in 2010. Sharp was charged in 2015 with three counts of child molestation and four counts of sexual misconduct with a minor; he died by suicide in prison.
USAG received at least four complaints against Georgia coach William McCabe, but did not report the allegations to the police. One gym owner had warned that McCabe "should be locked in a cage before someone is raped". McCab [{'title': 'USA Gymnastics sex abuse scandal', 'summary': 'Beginning in the 1990s, hundreds of gymnasts-primarily minors-were sexually abused over two decades in the United States, which is considered the largest sexual abuse scandal in sports history. More than 500 athletes alleged that they were sexually assaulted by gym owners, coaches, and staff working for gymnastics programs across the country, including USA Gymnastics (USAG) and Michigan State University (MSU). Hundreds of them sued USAG, MSU, and the United States Olympic Committee (USOC, later USOPC), which settled the suits in 2018 and 2021 for a total of nearly $900 million.\nThe breadth of the abuses was first revealed by The Indianapolis Star, which reported in September 2016 that "predatory coaches were allowed to move from gym to gym, undetected by a lax system of oversight, or dangerously passed on by USA Gymnastics-certified gyms". Coaches and officials perpetrated, facilitated, or worked to conceal abuse in Michigan, Pennsylvania, California, Rhode Island, Indiana and elsewhere. FBI agents declined to investigate early allegations of abuse, then lied about it, according to a U.S. Justice Department report. Dozens of officials at USAG, MSU, and the United States Olympic Committee (USOC, later USOPC) ultimately resigned under pressure or had their contracts terminated. Several coaches and officials faced criminal charges, though few were convicted.\nA central figure was Larry Nassar, a longtime national-team doctor for USAG and osteopathic physician in MSU\'s athletic department. More than 265 women said Nassar had sexually abused them under the pretense of providing medical treatment, including former USAG national team members Jessica Howard, Jamie Dantzscher, Morgan White, Jeanette Antolin, McKayla Maroney, Aly Raisman, Maggie Nichols, Gabby Douglas, Simone Biles, Jordyn Wieber, Sabrina Vega, Ashton Locklear, Kyla Ross, Madison Kocian, Amanda Jetter, Tasha Schwikert, Mattie Larson, Bailie Key, Kennedy Baker, Alyssa Baumann, and Terin Humphrey. In 2017 and 2018, Nassar pleaded guilty to federal charges of child pornography and state charges of first-degree sexual assault; he received sentences of 60 years in prison plus another 80 to 300 years. The scandal led to the Protecting Young Victims from Sexual Abuse and Safe Sport Authorization Act of 2017, which directed the creation of the U.S. Center for SafeSport.', 'source': 'https://en.wikipedia.org/wiki/USA_Gymnastics_sex_abuse_scandal'}]
* This is the complete list of women's Olympic medalists in gymnastics.

```
== Artistic gymnastics ==


=== Current program ===


==== All-around, individual ====


==== All-around, team ====

Note: The International Gymnastics Federation recommended to the IOC that the
medals of the Chinese team be stripped, and awarded to the fourth-placed United
States team, as it was revealed that Dong Fangxiao was underage (14, with age
limit >16) at the time. The IOC upheld the FIG decision in April 2010.


==== Balance beam ====


==== Floor exercise ====


==== Uneven bars ====


==== Vault ====


=== Discontinued event ===


==== Portable apparatus, team ====


== Rhythmic gymnastics ==


=== All-around, individual ===


=== All-around, group ===


== Trampoline ==
```

```
=== Individual ===


== See also ==
List of top Olympic gymnastics medalists
Artistic gymnastics
Rhythmic gymnastics
Trampoline


== References ==

International Olympic Committee results database [{'title': 'List of Olympic
medalists in gymnastics (women)', 'summary': "This is the complete list of
women's Olympic medalists in gymnastics.", 'source': 'https://en.wikipedia.org/w
iki/List_of_Olympic_medalists_in_gymnastics_(women)'}]

 The most popular sport in the US is American Football, specifically the
National Football League (NFL). The Super Bowl, the championship game of the
NFL, is consistently one of the most-watched television events in the US, with
over 100 million viewers tuning in each year.

According to a survey conducted by the Sports & Fitness Industry Association
(SFIA) in 2020, the top 5 most popular sports in the US are:

1. **American Football (NFL)**: 43% of respondents named American football as
their favorite sport.
2. **Basketball (NBA)**: 24% of respondents chose basketball as their favorite
sport.
3. **Baseball (MLB)**: 18% of respondents named baseball as their favorite
sport.
4. **Soccer (MLS)**: 12% of respondents chose soccer as their favorite sport.
5. **Hockey (NHL)**: 4% of respondents named hockey as their favorite sport.

It's worth noting that baseball was traditionally the most popular sport in the
US, but its popularity has declined in recent years, with American football and
the NFL taking its place.
```

0.4.1 a) GenAI Application Development

REQUIRED: This section is where input your custom logic to create and run your agentic workflow. Feel free to add as many codes cells that are needed for this assignment

```
[ ]: ##########################################################################################
     ###### FULL SELF-REFLECTIVE AGENT WORKFLOW                                  ␣
      ↪                  ######
     ##########################################################################################
```

```python
from langgraph.graph import StateGraph, END
from typing import TypedDict, List, Dict, Any
import json
import mlflow
import time
import re


################################################################################################
# 1  Agent State
################################################################################################

class AgentState(TypedDict):
    question: str
    docs: List[Dict[str, Any]]
    claims: List[Dict[str, Any]]
    answer: str
    retry_count: int
    confidence: Dict[str, Any]
    answer_mode: str
    reflection_count: int
    critique: Dict[str, Any]


################################################################################################
# 2  Retrieval Node
################################################################################################

def retrieve_node(state: AgentState):
    retrieval_k = k if state["retry_count"] == 0 else k * 2

    results = vector_store.similarity_search(state["question"], k=retrieval_k)

    docs = []
    for i, r in enumerate(results):
        docs.append({
            "doc_index": i,
            "content": r.page_content,
            "metadata": r.metadata
        })

    return {"docs": docs}


################################################################################################
# 3  Query Rewrite Node
################################################################################################
```

```
def rewrite_query_node(state: AgentState):

    prompt = f"""
Rewrite the question to improve semantic retrieval.

Original:
{state["question"]}

Return only the improved question.
"""

    improved_query = llm.invoke(prompt).content.strip()

    print("\n  Rewritten Query:", improved_query)

    return {
        "question": improved_query,
        "retry_count": state["retry_count"] + 1
    }

def extract_json(response: str) -> str:
    """Strips markdown fences and preamble text before JSON parsing."""
    response = response.strip()
    match = re.search(r'```(?:json)?\s*(.*?)\s*```', response, re.DOTALL)
    if match:
        return match.group(1).strip()
    return response

##########################################################################################
# 4  Claim Extraction Node
##########################################################################################

def claims_node(state: AgentState):

    prompt = f"""
Extract 3-6 factual claims answering the question using ONLY the documents.

Question:
{state["question"]}

Documents:
{json.dumps(state["docs"], ensure_ascii=False)}

Return JSON:
{{
  "claims": [
```

```
    {{"claim": "...", "doc_index": 0}}
  ]
}}
"""

    response = extract_json(llm.invoke(prompt).content)
    extract_json(response)


    try:
        parsed = json.loads(response)
        claims = parsed.get("claims", [])

        cleaned = []
        for c in claims:
            if "claim" in c and "doc_index" in c:
                cleaned.append({
                    "claim": str(c["claim"]),
                    "doc_index": int(c["doc_index"])
                })

        return {"claims": cleaned}

    except:
        return {"claims": []}


################################################################################################
# 5 Evidence Validation (Hallucination Filter)
################################################################################################

def validate_claims_node(state: AgentState):

    validated_claims = []

    for claim in state["claims"]:
        idx = claim["doc_index"]
        if idx >= len(state["docs"]):
            continue

        doc_text = state["docs"][idx]["content"]

        prompt = f"""
Verify if the claim is directly supported by the document.

Claim:
{claim["claim"]}
```

```
Document:
{doc_text}

Return JSON:
{{ "supported": true/false }}
"""

        response = extract_json(llm.invoke(prompt).content)
        extract_json(response)

        try:
            parsed = json.loads(response)
            if parsed.get("supported") == True:
                validated_claims.append(claim)
        except:
            pass

    print(f"\n  Validated {len(validated_claims)} / {len(state['claims'])}␣
 ↪claims")

    return {"claims": validated_claims}


################################################################################################
# 6  Retry Decision
################################################################################################

def retry_decision(state: AgentState):
    if len(state["claims"]) < 2 and state["retry_count"] == 0:
        return "rewrite"
    return "continue"


################################################################################################
# 7  Answer Node (With Style Control)
################################################################################################

def answer_node(state: AgentState):

    if not state["claims"]:
        return {"answer": "I don't have enough evidence in the retrieved␣
 ↪documents to answer confidently."}

    mode_instruction = {
        "concise": "Write a short answer in 3-5 sentences.",
        "detailed": "Write a detailed explanation.",
```

```
        "bullet": "Write as bullet points.",
        "executive_summary": "Write a high-level executive summary."
    }.get(state["answer_mode"], "Write a concise answer.")

    prompt = f"""
You are a factual assistant.

Question:
{state["question"]}

Claims:
{json.dumps(state["claims"], ensure_ascii=False)}

Instructions:
{mode_instruction}

Rules:
- Only use provided claims.
- Cite document index like (Doc 0).
- Do not hallucinate.
"""

    response = extract_json(llm.invoke(prompt).content)

    return {"answer": response}


################################################################################################
# 8 Self-Reflection Critique Node
################################################################################################

def critique_node(state: AgentState):

    prompt = f"""
Evaluate the answer quality.

Question:
{state["question"]}

Answer:
{state["answer"]}

Claims:
{json.dumps(state["claims"], ensure_ascii=False)}

Return JSON:
{{
```

```
  "needs_improvement": true/false,
  "issues": ["..."],
  "suggestions": "..."
}}
"""

    response = extract_json(llm.invoke(prompt).content)
    extract_json(response)

    try:
        parsed = json.loads(response)
        return {"critique": parsed}
    except:
        return {"critique": {"needs_improvement": False}}


################################################################################################
# 9 Regeneration Node
################################################################################################

def improve_answer_node(state: AgentState):

    prompt = f"""
Improve the answer based on critique.

Original Answer:
{state["answer"]}

Claims:
{json.dumps(state["claims"], ensure_ascii=False)}

Critique:
{json.dumps(state["critique"], ensure_ascii=False)}

Rules:
- Only use claims.
- Improve clarity and completeness.
- Do not hallucinate.
"""

    improved = llm.invoke(prompt).content.strip()

    return {
        "answer": improved,
        "reflection_count": state["reflection_count"] + 1
    }
```

```python
##############################################################################################
#  Reflection Decision
##############################################################################################

def reflection_decision(state: AgentState):
    if (
        state["critique"].get("needs_improvement", False)
        and state["reflection_count"] < 1
    ):
        return "improve"
    return "continue"


##############################################################################################
# 11  Confidence Node
##############################################################################################

def confidence_node(state: AgentState):

    prompt = f"""
Rate confidence from 1-5.

Answer:
{state["answer"]}

Claims:
{json.dumps(state["claims"], ensure_ascii=False)}

Return JSON:
{{ "score": 1-5, "reasoning": "..." }}
"""

    response = extract_json(llm.invoke(prompt).content)
    extract_json(response)

    try:
        parsed = json.loads(response)
        return {"confidence": parsed}
    except:
        return {"confidence": {"score": 2}}


##############################################################################################
# 12  Evaluation Metrics
##############################################################################################
```

```python
def evaluation_node(state: AgentState):

    unique_docs = len(set(c["doc_index"] for c in state["claims"])) if␣
 ↪state["claims"] else 0

    metrics = {
        "num_claims": len(state["claims"]),
        "unique_docs_used": unique_docs,
        "retry_count": state["retry_count"],
        "reflection_count": state["reflection_count"]
    }

    print("\n Metrics:", metrics)
    mlflow.log_metrics(metrics)

    return {}


################################################################################################
# 13 Build Graph
################################################################################################

graph = StateGraph(AgentState)

graph.add_node("retrieve", retrieve_node)
graph.add_node("claims", claims_node)
graph.add_node("rewrite", rewrite_query_node)
graph.add_node("validate", validate_claims_node)
graph.add_node("answer", answer_node)
graph.add_node("critique", critique_node)
graph.add_node("improve", improve_answer_node)
graph.add_node("confidence", confidence_node)
graph.add_node("evaluate", evaluation_node)

graph.set_entry_point("retrieve")

graph.add_edge("retrieve", "claims")

graph.add_conditional_edges(
    "claims",
    retry_decision,
    {"rewrite": "rewrite", "continue": "validate"}
)

graph.add_edge("rewrite", "retrieve")
graph.add_edge("validate", "answer")
graph.add_edge("answer", "critique")
```

```
graph.add_conditional_edges(
    "critique",
    reflection_decision,
    {"improve": "improve", "continue": "confidence"}
)

graph.add_edge("improve", "critique")
graph.add_edge("confidence", "evaluate")
graph.add_edge("evaluate", END)

agent = graph.compile()


##############################################################################################
# 14 Run Agent
##############################################################################################

with mlflow.start_run():

    start = time.time()

    result = agent.invoke({
        "question": example_question,
        "docs": [],
        "claims": [],
        "answer": "",
        "retry_count": 0,
        "confidence": {},
        "answer_mode": "detailed",
        "reflection_count": 0,
        "critique": {}
    })

    latency = time.time() - start
    mlflow.log_metric("latency_seconds", latency)

    print("\n================================")
    print("QUESTION:", result["question"])
    print("================================\n")

    print("ANSWER:\n", result["answer"])
    print("\nCLAIMS:\n", json.dumps(result["claims"], indent=2))
    print("\nCONFIDENCE:\n", result["confidence"])
```

```
Rewritten Query: What are the fundamental characteristics and rules of soccer?
```

```
 Validated 0 / 0 claims

 Metrics: {'num_claims': 0, 'unique_docs_used': 0, 'retry_count': 1,
'reflection_count': 0}

==============================
QUESTION: What are the fundamental characteristics and rules of soccer?
==============================

ANSWER:
 I don't have enough evidence in the retrieved documents to answer confidently.

CLAIMS:
 []

CONFIDENCE:
 {'score': 3, 'reasoning': 'Confidence is being rated as 3 due to the absence of
concrete evidence in the retrieved documents, which indicates a moderate level
of uncertainty. A score of 3 suggests that the available information is limited
and further investigation or clarification may be necessary to answer more
confidently.'}
```

0.4.2 b) Reflection

1. If I had more time, what improvements would I make and why? If given additional time, I would focus on strengthening retrieval quality, reliability guarantees, and evaluation rigor.

a) Add Cross-Encoder Re-ranking Currently, retrieval uses bi-encoder embeddings (BGE + FAISS). While efficient, bi-encoders optimize for semantic similarity but may miss fine-grained relevance.
I would add a cross-encoder re-ranking stage after initial retrieval to score the top-k documents more precisely.

Why:
Cross-encoders significantly improve ranking accuracy and reduce irrelevant context.
Value: Better grounding → fewer hallucinations → stronger answers.

---

b) Implement Multi-Hop Retrieval Planning The current workflow performs single-hop retrieval. For complex questions (e.g., involving comparisons or causal reasoning), multi-hop reasoning is required.

I would: - Add a query decomposition node - Generate sub-questions - Retrieve evidence per sub-question - Merge claims across hops

Why:
Many real-world questions require combining information across multiple documents.
Value: Improves reasoning depth and correctness for complex queries.

---

c) Add Semantic Similarity Threshold Gating Instead of always accepting retrieved documents, I would introduce similarity score thresholds.

If top similarity score is below a threshold: - Return "insufficient evidence" - Or escalate to fallback mechanism

Why:
Low similarity often correlates with hallucination risk.
Value: Improves system reliability and prevents overconfident answers.

---

d) Add Offline Evaluation Benchmarking I would create a labeled evaluation dataset and compute:

- Faithfulness score
- Answer relevancy
- Citation precision
- Retrieval recall

Why:
Prototype systems often lack rigorous measurement.
Value: Enables systematic performance tracking and regression testing.

---

e) Add Caching and Cost Optimization - Cache embeddings - Cache LLM responses - Batch retrieval - Use smaller models for validation steps

Why:
Prototype systems do not optimize cost or latency.
Value: Reduces inference cost and improves scalability.

---

2. What steps are required to move this from prototype to production? Moving from prototype to production requires changes across architecture, infrastructure, evaluation, and governance.

0.4.3 A. Infrastructure Hardening

1. Deploy model endpoints behind autoscaling infrastructure.
2. Use managed vector databases (e.g., Databricks Vector Search or Pinecone).
3. Add retry mechanisms and rate limiting.
4. Implement structured logging and tracing (e.g., OpenTelemetry).

Goal: Ensure reliability, scalability, and observability.

0.4.4 B. Monitoring & Observability

1. Track:

- Latency
- Token usage
- Retrieval score distributions
- Claim validation rate
- Reflection frequency

2. Add alerting on:
   - High hallucination rate
   - Low confidence scores
   - Retrieval failures

Goal: Detect degradation early.

0.4.5 C. Security & Governance

1. Add input sanitization and prompt injection protection.
2. Enforce strict output schemas.
3. Log prompts and outputs for auditability.
4. Apply access control and authentication.
5. Implement PII detection if needed.

Goal: Enterprise-grade safety and compliance.

0.4.6 D. Robust Evaluation Pipeline

1. Create a labeled benchmark dataset.
2. Automate nightly evaluation runs.
3. Track:
   - Faithfulness
   - Citation correctness
   - Confidence calibration
4. Perform A/B testing for new model versions.

Goal: Prevent silent performance regressions.

0.4.7 E. Performance Optimization

1. Reduce multi-pass LLM calls.
2. Use smaller models for critique/validation.
3. Add asynchronous orchestration.
4. Introduce token budgeting and truncation strategies.

Goal: Lower latency and cost.

0.4.8 F. Productization Layer

1. Wrap the agent in an API (FastAPI or Databricks Serving).
2. Add streaming support.
3. Implement UI confidence indicators.
4. Provide structured JSON output mode.

Goal: Make system usable in real-world applications.

0.4.9 Final Reflection

This prototype demonstrates a self-reflective, grounded RAG architecture with structured reasoning and guardrails. However, productionizing such a system requires:

- Rigorous evaluation
- Infrastructure scalability
- Monitoring
- Security controls
- Cost optimization
- Continuous improvement pipelines

The architectural foundation is strong, but enterprise deployment requires systematic engineering beyond model prompting.

1 Advanced Features Implemented

1.1 1 Query Rewriting

If initial retrieval is weak, the system rewrites the user query to improve semantic search.

Why: Improves recall and robustness against vague queries.
Value: Higher retrieval quality without blindly increasing context size.

1.2 2 Conditional Retrieval Expansion

If evidence is insufficient, the system increases retrieval size (`k`) dynamically.

Why: Small retrieval windows may miss relevant documents.
Value: Cost-efficient scaling instead of always retrieving large contexts.

1.3 3 Structured Claim Extraction

The system extracts 3–6 factual claims tied to specific document indices before generating the final answer.

Why: Direct LLM answering risks hallucination.
Value: Forces intermediate reasoning and grounding.

1.4 4 Hallucination Detection (Evidence Validation Layer)

Each claim is validated against its source document before being used.

Why: LLMs may fabricate claims.
Value: Filters unsupported claims and increases factual reliability.

1.5 5 Controlled Answer Modes

Supports configurable output styles: - concise - detailed - bullet - executive_summary

Why: Different consumers require different formats.
Value: Improves product flexibility and usability.

1.6 6 Self-Reflection Loop (Self-Critique + Regeneration)

After generating an answer: - The system critiques its own response - Identifies weaknesses - Regenerates improved output (limited to 1 iteration)

Why: Initial answers may be shallow or incomplete.
Value: Improves clarity, citation usage, and completeness.

This mimics advanced reasoning systems such as Self-Refine and Reflexion architectures.

1.7 7 Confidence Scoring

The system assigns a confidence score (1–5) based on: - Claim strength - Evidence coverage - Answer completeness

Why: Production systems require reliability signals.
Value: Enables threshold gating and monitoring.

1.8 8 Evaluation & Observability

Metrics logged via MLflow:

- Number of extracted claims
- Unique documents used
- Retry count
- Reflection count
- Latency

Why: AI systems require instrumentation and monitoring.
Value: Enables performance tracking and iterative improvement.

2 Design Principles

- Strict grounding
- Iterative reasoning
- Conditional logic (not linear prompting)
- Modular nodes
- Observable system behavior
- Controlled generation
- Guardrail-first design