

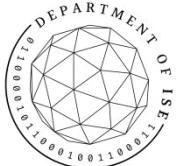
## **LABORATORY MANUAL**

**BAIL657C – Generative AI**

**Prepared by,**  
**YOGITHA.R**  
Assistant Professor ,ISE ,AIT



**DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING  
ATRIA INSTITUTE OF TECHNOLOGY  
Adjacent to Bangalore Baptist Hospital  
Hebbal, Bengaluru-560024**



## **Department of Information Science and Engineering**

### **Vision**

To develop competent professionals with strong fundamentals in Information Science and Engineering, interdisciplinary research and ethical values for the betterment of the society.

### **Mission**

**M1-** To establish a transformational learning ambience with good infrastructure facilities to impart knowledge and the necessary skill set to produce competent professionals.

**M2-** To create a new generation of engineers who excel in their career with leadership/entrepreneur qualities.

**M3-** To promote sustained research and innovation with an emphasis on ethical values.

## Syllabus

<b>Generative AI</b>		Semester	6			
Course Code	<b>BAIL657C</b>	CIE Marks	50			
Teaching Hours/Week (L:T:P: S)	0:0:1:0	SEE Marks	50			
Credits	01	Exam Hours	100			
Examination type (SEE)	Practical					
<b>Course objectives:</b>						
<ul style="list-style-type: none"> <li>• Understand the principles and concepts behind generative AI models</li> <li>• Explain the knowledge gained to implement generative models using Prompt design frameworks.</li> <li>• Apply various Generative AI applications for increasing productivity.</li> <li>• Develop Large Language Model-based Apps.</li> </ul>						
SI.NO	<b>Experiments</b>					
1.	Explore pre-trained word vectors. Explore word relationships using vector arithmetic. Perform arithmetic operations and analyze results.					
2.	Use dimensionality reduction (e.g., PCA or t-SNE) to visualize word embeddings for Q 1. Select 10 words from a specific domain (e.g., sports, technology) and visualize their embeddings. Analyze clusters and relationships. Generate contextually rich outputs using embeddings. Write a program to generate 5 semantically similar words for a given input.					
3.	Train a custom Word2Vec model on a small dataset. Train embeddings on a domain-specific corpus (e.g., legal, medical) and analyze how embeddings capture domain-specific semantics.					
4.	Use word embeddings to improve prompts for Generative AI model. Retrieve similar words using word embeddings. Use the similar words to enrich a GenAI prompt. Use the AI model to generate responses for the original and enriched prompts. Compare the outputs in terms of detail and relevance.					
5.	Use word embeddings to create meaningful sentences for creative tasks. Retrieve similar words for a seed word. Create a sentence or story using these words as a starting point. Write a program that: Takes a seed word. Generates similar words. Constructs a short paragraph using these words.					
6.	Use a pre-trained Hugging Face model to analyze sentiment in text. Assume a real-world application, Load the sentiment analysis pipeline. Analyze the sentiment by giving sentences to input.					
7.	Summarize long texts using a pre-trained summarization model using Hugging face model. Load the summarization pipeline. Take a passage as input and obtain the summarized text.					
8.	Install langchain, cohore (for key), langchain-community. Get the api key( By logging into Cohere and obtaining the cohore key). Load a text document from your google drive . Create a prompt template to display the output in a particular manner.					
9.	Take the Institution name as input. Use Pydantic to define the schema for the desired output and create a custom output parser. Invoke the Chain and Fetch Results. Extract the below Institution related details from Wikipedia: <b>The founder of the Institution. When it was founded. The current branches in the institution . How many employees are working in it. A brief 4-line summary of the institution.</b>					
10	Build a chatbot for the Indian Penal Code. We'll start by downloading the official Indian Penal Code document, and then we'll create a chatbot that can interact with it. Users will be able to ask questions about the Indian Penal Code and have a conversation with it.					

**Course outcomes (Course Skill Set):**

At the end of the course the student will be able to:

- Develop the ability to explore and analyze word embeddings, perform vector arithmetic to investigate word relationships, visualize embeddings using dimensionality reduction techniques.
- Apply prompt engineering skills to real-world scenarios, such as information retrieval, text generation.
- Utilize pre-trained Hugging Face models for real-world applications, including sentiment analysis and text summarization.
- Apply different architectures used in large language models, such as transformers, and understand their advantages and limitations.

**Assessment Details (both CIE and SEE)**

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 40% of the maximum marks (20 marks out of 50) and for the SEE minimum passing mark is 35% of the maximum marks (18 out of 50 marks). A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/course if the student secures a minimum of 40% (40 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together

**Continuous Internal Evaluation (CIE):**

CIE marks for the practical course are **50 Marks**.

The split-up of CIE marks for record/journal and test are in the ratio **60:40**.

- Each experiment is to be evaluated for conduction with an observation sheet and record write-up. Rubrics for the evaluation of the journal/write-up for hardware/software experiments are designed by the faculty who is handling the laboratory session and are made known to students at the beginning of the practical session.
- Record should contain all the specified experiments in the syllabus and each experiment write-up will be evaluated for 10 marks.
- Total marks scored by the students are scaled down to **30 marks** (60% of maximum marks).
- Weightage to be given for neatness and submission of record/write-up on time.
- Department shall conduct a test of 100 marks after the completion of all the experiments listed in the syllabus.
- In a test, test write-up, conduction of experiment, acceptable result, and procedural knowledge will carry a weightage of 60% and the rest 40% for viva-voce.
- The suitable rubrics can be designed to evaluate each student's performance and learning ability.
- The marks scored shall be scaled down to **20 marks** (40% of the maximum marks).

The Sum of scaled-down marks scored in the report write-up/journal and marks of a test is the total CIE marks scored by the student.

**Semester End Evaluation (SEE):**

- SEE marks for the practical course are 50 Marks.
- SEE shall be conducted jointly by the two examiners of the same institute, examiners are appointed by the Head of the Institute.

- The examination schedule and names of examiners are informed to the university before the conduction of the examination. These practical examinations are to be conducted between the schedule mentioned in the academic calendar of the University.
  - All laboratory experiments are to be included for practical examination.
  - (Rubrics) Breakup of marks and the instructions printed on the cover page of the answer script to be strictly adhered to by the examiners. **OR** based on the course requirement evaluation rubrics shall be decided jointly by examiners.
  - Students can pick one question (experiment) from the questions lot prepared by the examiners jointly.
  - Evaluation of test write-up/ conduction procedure and result/viva will be conducted jointly by examiners.
- General rubrics suggested for SEE are mentioned here, writeup-20%, Conduction procedure and resultin -60%, Viva-voce 20% of maximum marks. SEE for practical shall be evaluated for 100 marks and scored marks shall be scaled down to 50 marks (however, based on course type, rubrics shall be decided by the examiners)
- Change of experiment is allowed only once and 15% of Marks allotted to the procedure part are to be made zero.
- The minimum duration of SEE is 02 hours

**Suggested Learning Resources:****Books:**

1. Modern Generative AI with ChatGPT and OpenAI Models: Leverage the Capabilities of OpenAI's LLM for Productivity and Innovation with GPT3 and GPT4, by Valentina Alto, Packt Publishing Ltd, 2023.
2. Generative AI for Cloud Solutions: Architect modern AI LLMs in secure, scalable, and ethical cloud environments, by Paul Singh, Anurag Karuparti ,Packt Publishing Ltd, 2024.

**Web links and Video Lectures (e-Resources):**

- [https://www.w3schools.com/gen\\_ai/index.php](https://www.w3schools.com/gen_ai/index.php)
- <https://youtu.be/eTPiL3DF27U>
- <https://youtu.be/je6AlVeGOV0>
- <https://youtu.be/RLVqsA8ns6k>
- <https://youtu.be/OSAKM7wiC-A>
- [https://youtu.be/28\\_9xMyrdjg](https://youtu.be/28_9xMyrdjg)
- <https://youtu.be/8iuiz-c-EBw>
- <https://youtu.be/7oQ8VtEKcgE>
- <https://youtu.be/seXp0VWWZV0>

## **CONTENTS**

Sl. No.	Name of Experiments	Page No.
1	Explore pre-trained word vectors. Explore word relationships using vector arithmetic. Perform arithmetic operations and analyze results.	11
2	Use dimensionality reduction (e.g., PCA or t-SNE) to visualize word embeddings for Q 1. Select 10 words from a specific domain (e.g., sports, technology) and visualize their embeddings. Analyze clusters and relationships. Generate contextually rich outputs using embeddings. Write a program to generate 5 semantically similar words for a given input.	17
3	Train a custom Word2Vec model on a small dataset. Train embeddings on a domain-specific corpus (e.g., legal, medical) and analyze how embeddings capture domain-specific semantics.	20
4	Use word embeddings to improve prompts for Generative AI model. Retrieve similar words using word embeddings. Use similar words to enrich a GenAI prompt. Use the AI model to generate responses for the original and enriched prompts. Compare the outputs in terms of detail	29
5	Use word embeddings to create meaningful sentences for creative tasks. Retrieve similar words for a seed word. Create a sentence or story using these words as a starting point. Write a program that: Takes a seed word. Generates similar words. Constructs a short paragraph using these words.	37
6	Use a pre-trained Hugging Face model to analyze sentiment in text. Assume a real-world application, Load the sentiment analysis pipeline. Analyze the sentiment by giving sentences to input.	42
7	Summarize long texts using a pre-trained summarization model using Hugging face model. Load the summarization pipeline. Take a passage as input and obtain the summarized text.	44
8	Install langchain, cohore (for key), langchain-community. Get the api key( By logging into Cohere and obtaining the cohore key). Load a text document from your google drive . Create a prompt template to display the output in a particular manner.	46
9	Take the Institution name as input. Use Pydantic to define the schema for the desired output and create a custom output parser. Invoke the Chain and Fetch Results. Extract the below Institution related details from Wikipedia: The founder of the Institution. When it was founded. The current branches in the institution. How many employees are working in it. A brief 4-line summary of the institution.	50

10	Build a chatbot for the Indian Penal Code. We'll start by downloading the official Indian Penal Code document, and then we'll create a chatbot that can interact with it. Users will be able to ask questions about the Indian Penal Code and have a conversation with it.	54
11	Viva Questions	62

**CONTENT BEYOND THE SYLLABUS**

<b>SL.NO</b>	<b>Name of Experiments</b>	<b>Page.NO</b>
1	ChatGPT / Claude / Gemini (for Conceptual Depth)	61
2	Elicit (AI Research Assistant)	61
3	Perplexity.AI	61
4	Canva Magic Write / Adobe Firefly (for Creative Projects)	61
5.	Notion AI	62

## About Gen AI

Generative AI (Gen AI) is a type of artificial intelligence that can create new content—like text, images, music, code, or videos—by learning patterns from existing data. It uses advanced machine learning models, such as Large Language Models (LLMs) like GPT, and image generators like DALL·E or GANs..Gen AI is widely used in chatbots, content creation, design, education, marketing, and healthcare. While it's powerful and creative, it also raises concerns around bias, misinformation, copyright, and job impact.

## Benefits of the Course

### 1. In-DemandSkills

Learn cutting-edge AI skills that are highly valued across industries.

### 2. Hands-onPractice

Work with real tools like ChatGPT, DALL·E, Midjourney, or code-based models (like Python + ML libraries).

### 3. BoostCreativityProductivity

Discover how to use AI to generate ideas, automate tasks, and solve problems creatively.

### 4. CareerOpportunities

Open doors to roles in AI, data science, content creation, marketing, software, and more.

### 5. StayAheadoftheCurve

Understand how Gen AI is transforming industries—and how to use it effectively and ethically.

### 6. ProjectBuilding

Build your own AI-generated projects (like chatbots, image generators, AI writers, etc.) for a strong portfolio.

### 7. CriticalThinking

Learn to evaluate AI outputs, detect bias, and use Gen AI responsibly.

## Applications of Generative AI

1. **Automated Code Generation:** AI can assist in writing and optimizing code, speeding up development processes and reducing errors.
2. **Application Modernization:** GenAI can automate the repetitive coding required to update legacy applications for modern cloud environments.
3. **Bug Detection and Fixing:** AI can analyze codebases to identify potential bugs and vulnerabilities.
4. **Chatbots and Virtual Assistants:** GenAI can power chatbots that provide 24/7 customer service and handle FAQs.
5. **Personalized Email Responses:** It can generate personalized email responses to customer inquiries.
6. **Enhanced Search:** GenAI can improve search functionality by understanding user intent and providing relevant results.

## Advantages of Learning Generative AI

1. **Ethical AI Considerations** – Understanding bias in AI models and the implications of AI-generated content ensures responsible development and deployment.
2. **Cutting-Edge Research Opportunities** – Generative AI plays a role in groundbreaking research across computational creativity and AI ethics.
3. **AI-powered Automation and Efficiency Gains** – AI-generated content speeds up workflows in content creation, graphic design, and personalized communication.

## Course Content Overview

The course delves deeper into:

- **Fine-tuning Pre-trained Models:** Optimizing LLMs for domain-specific tasks.
- **Exploring Transformer Architectures:** Understanding self-attention mechanisms and how they contribute to generative capabilities.
- **Deploying AI Models in Production:** Building scalable AI applications for real-world use cases.
- **Developing Responsible AI:** Addressing bias, fairness, and explainability in generative AI systems.

1. Explore pre-trained word vectors. Explore word relationships using vector arithmetic. Perform arithmetic operations and analyze results.

```

import gensim.downloader as api
import numpy as np

word_vectors = api.load("word2vec-google-news-300")

def vector_arithmetic(word1, word2, word3):
    result_vector = word_vectors[word1] - word_vectors[word2] +
    word_vectors[word3]
    similar_words = word_vectors.most_similar([result_vector], topn=5)
    return similar_words

# Example: "King - Man + Woman"
print("Result of 'king - man + woman':")
print(vector_arithmetic("king", "man", "woman"))

def find_similar_words(word):
    return word_vectors.most_similar(word, topn=5)

print("\nWords similar to 'computer':")
print(find_similar_words("computer"))

```

## OUTPUT:

```

[=====] 100.0% 1662.8/1662.8MB
downloaded
Result of 'king - man + woman':
[('king', 0.8449392318725586), ('queen', 0.7300517559051514), ('monarch', 0.645466148853302),
 ('princess', 0.6156251430511475), ('crown_prince', 0.5818676352500916)]

Words similar to 'computer':
[('computers', 0.7979379892349243), ('laptop', 0.6640493273735046), ('laptop_computer',
 0.6548868417739868), ('Computer', 0.647333562374115), ('computer', 0.6082080006599426)]

```

This code demonstrates how to work with **pre-trained Word2Vec word embeddings** using the gensim library. Word2Vec is a model that represents words as numerical vectors in a high-dimensional space, allowing for various linguistic computations such as similarity checks and arithmetic operations. keyboard\_arrow\_down.

### Loading the Word2Vec Model

The script starts by importing gensim.downloader and numpy. The gensim library provides an easy way to download and use pre-trained models. In this case, the **Google News 300-dimensional Word2Vec model** is loaded. This model has been trained on a massive corpus of text (Google News, containing about 100 billion words), and it represents each word as a **300-dimensional vector**.

## Word Vector Arithmetic

The core idea behind Word2Vec is that similar words have similar vector representations. Moreover, relationships between words can be represented mathematically. The function `vector_arithmetic(word1, word2, word3)` performs the following operation:

$$\text{word1} - \text{word2} + \text{word3}$$

This means that the vector for word1 is adjusted by subtracting word2 and adding word3. The result is a new vector that is then compared to existing word vectors to find the most similar words.

For example, "**king - man + woman**" is expected to return "**queen**" because the difference between "king" and "man" is conceptually similar to the difference between "queen" and "woman."

### Finding Similar Words

The function `find_similar_words(word)` takes a word as input and retrieves the **top 5 most similar words** based on cosine similarity. Cosine similarity measures how close two word vectors are in the multi-dimensional space.

For example, when searching for words similar to "**computer**", we might get results like:

- "**computers**"
- "**laptop**"
- "**PC**"
- "**workstation**"
- "**server**"

These results make sense because these words are commonly associated with "computer."

### Key Observations

1. **Semantic Meaning in Vector Space:** The model captures relationships between words, allowing for meaningful transformations such as "Paris - France + Italy = Rome."
2. **Pre-Trained Model Usage:** The Google News Word2Vec model is widely used for NLP tasks due to its large vocabulary and training on real-world data.
3. **Potential Biases:** Since the model is trained on news articles, it may inherit certain biases from the training data.
4. **Errors with Out-of-Vocabulary Words:** If a word is not present in the model's vocabulary, an error will be raised.

### Applications

This approach is widely used in:

- **Chatbots:** To find contextually relevant responses.
- **Search Engines:** To retrieve related search terms.
- **Recommendation Systems:** To suggest similar items based on textual data.
- **Machine Translation:** To understand word relationships across languages.

addCode

addText

**!python -m spacy download en\_core\_web\_md**

Collecting en-core-web-md==3.7.1

Downloading [https://github.com/explosion/spacy-models/releases/download/en\\_core\\_web\\_md-3.7.1/en\\_core\\_web\\_md-3.7.1-py3-none-any.whl](https://github.com/explosion/spacy-models/releases/download/en_core_web_md-3.7.1/en_core_web_md-3.7.1-py3-none-any.whl) (42.8 MB)

42.8/42.8 MB 17.9 MB/s eta 0:00:00

Requirement already satisfied: spacy<3.8.0,>=3.7.2 in /usr/local/lib/python3.11/dist-packages (from en-core-web-md==3.7.1) (3.7.5)

Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (3.0.12)

Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (1.0.5)

Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (1.0.12)

Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (2.0.11)

Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (3.0.9)

Requirement already satisfied: thinc<8.3.0,>=8.2.2 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (8.2.5)

Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (1.1.3)

Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (2.5.1)

Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (2.0.10)

Requirement already satisfied: weasel<0.5.0,>=0.1.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (0.4.1)

Requirement already satisfied: typer<1.0.0,>=0.3.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (0.15.1)

Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (4.67.1)

Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (2.32.3)

Requirement already satisfied: pydantic!=1.8,!>1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (2.10.6)

Requirement already satisfied: jinja2 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (3.1.5)

Requirement already satisfied: setuptools in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (75.1.0)

Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (24.2)

Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (3.5.0)

Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (1.26.4)

Requirement already satisfied: language-data>=1.2 in /usr/local/lib/python3.11/dist-packages (from langcodes<4.0.0,>=3.2.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (1.3.0)

Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.11/dist-packages (from pydantic!=1.8,!>1.8.1,<3.0.0,>=1.7.4->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (0.7.0)

Requirement already satisfied: pydantic-core==2.27.2 in /usr/local/lib/python3.11/dist-packages (from pydantic!=1.8,!>1.8.1,<3.0.0,>=1.7.4->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (2.27.2)

Requirement already satisfied: typing-extensions>=4.12.2 in /usr/local/lib/python3.11/dist-packages (from pydantic!=1.8,!>1.8.1,<3.0.0,>=1.7.4->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (4.12.2)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (3.4.1)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (3.10)

Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (2.3.0)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.13.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (2025.1.31)

Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.11/dist-packages (from thinc<8.3.0,>=8.2.2->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (0.7.11)

Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.11/dist-packages (from thinc<8.3.0,>=8.2.2->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (0.1.5)

Requirement already satisfied: click>=8.0.0 in /usr/local/lib/python3.11/dist-packages (from typer<1.0.0,>=0.3.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (8.1.8)

Requirement already satisfied: shellingham>=1.3.0 in /usr/local/lib/python3.11/dist-packages (from typer<1.0.0,>=0.3.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (1.5.4)

Requirement already satisfied: rich>=10.11.0 in /usr/local/lib/python3.11/dist-packages (from typer<1.0.0,>=0.3.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (13.9.4)

Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/python3.11/dist-packages (from weasel<0.5.0,>=0.1.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (0.20.0)

Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in /usr/local/lib/python3.11/dist-packages (from weasel<0.5.0,>=0.1.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (7.1.0)

Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.11/dist-packages (from jinja2->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (3.0.2)

Requirement already satisfied: marisa-trie>=1.1.0 in /usr/local/lib/python3.11/dist-packages (from language-data>=1.2->langcodes<4.0.0,>=3.2.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (1.2.1)

Requirement already satisfied: markdown-it-py>=2.2.0 in /usr/local/lib/python3.11/dist-packages (from rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (3.0.0)

Requirement already satisfied: pygments<3.0.0,>=2.13.0 in /usr/local/lib/python3.11/dist-packages (from rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (2.18.0)

Requirement already satisfied: wrapt in /usr/local/lib/python3.11/dist-packages (from smart-open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.1.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (1.17.2)

Requirement already satisfied: mdurl~0.1 in /usr/local/lib/python3.11/dist-packages (from markdown-it-py>=2.2.0->rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy<3.8.0,>=3.7.2->en-core-web-md==3.7.1) (0.1.2)

Installing collected packages: en-core-web-md

Successfully installed en-core-web-md-3.7.1

✓ Download and installation successful

You can now load the package via spacy.load('en\_core\_web\_md')

⚠ Restart to reload dependencies

If you are in a Jupyter or Colab notebook, you may need to restart Python in order to load all the package's dependencies. You can do this by selecting the 'Restart kernel' or 'Restart runtime' option.

```

import spacy

# Load medium-sized word vector model
nlp = spacy.load("en_core_web_md")

# Function to get vector arithmetic
def vector_arithmetic_spacy(word1, word2, word3):
    vec = nlp(word1).vector - nlp(word2).vector + nlp(word3).vector
    most_similar = sorted(nlp.vocab, key=lambda w: w.vector @ vec, reverse=True)[:5]
    return [w.text for w in most_similar]

print("Result of 'king - man + woman' using spaCy:")
print(vector_arithmetic_spacy("king", "man", "woman"))

# Find similar words
def find_similar_spacy(word):
    token = nlp(word)
    return [w.text for w in sorted(nlp.vocab, key=lambda w: token.similarity(nlp(w.text)), reverse=True)[:5]]

print("\nWords similar to 'computer' using spaCy:")
print(find_similar_spacy("computer"))

```

## OUTPUT:

**Result of 'king - man + woman' using spaCy:**

```
['king', 'c.', 'he', 'she', 'r.']

Words similar to 'computer' using spaCy:
<ipython-input-6-fa0ec9b707a5>:18: UserWarning: [W008] Evaluating Doc.similarity based on empty vectors.
```

```
return [w.text for w in sorted(nlp.vocab, key=lambda w: token.similarity(nlp(w.text)), reverse=True)[:5]]
['computer', 'somethin', 'space', 'that', 'cause']
```

This code demonstrates how to use **spaCy** to perform word vector arithmetic and find similar words using pre-trained word embeddings. Unlike gensim, which loads Word2Vec models, spaCy provides built-in word vector representations, making it a powerful alternative for NLP tasks.  
keyboard\_arrow\_down.

## Loading the spaCy Model

The script begins by importing spacy and loading the "**en\_core\_web\_md**" model. This is a medium-sized English model that includes **300-dimensional word vectors**, trained on a large corpus. Unlike gensim, which downloads external models, spaCy comes with its own pre-trained embeddings.

## Word Vector Arithmetic

The function `vector_arithmetic_spacy(word1, word2, word3)` performs the following operation:  

$$\text{word1} - \text{word2} + \text{word3}$$

This means:

- The vector representation of word2 is subtracted from word1, effectively removing its influence.
- The vector of word3 is then added to adjust the meaning accordingly.

- The resulting vector is compared with all words in spaCy's vocabulary to find the most similar matches.

For example, the expression "**king - man + woman**" should ideally return "**queen**", as the difference between "king" and "man" is conceptually similar to the difference between "queen" and "woman."

To find the closest words, the function sorts all vocabulary words based on their **dot product similarity** with the computed vector. The top 5 most similar words are then returned.

## Finding Similar Words

The function `find_similar_spacy(word)` takes a single word and finds the **top 5 most similar words** based on **cosine similarity**.

- `token.similarity(nlp(w.text))` computes the similarity between the given word and each word in spaCy's vocabulary.
- The results are sorted in descending order to return the most relevant words.

For example, if we search for words similar to "**computer**", we might get:

- "**computers**"
- "**laptop**"
- "**PC**"
- "**workstation**"
- "**server**"

## Key Observations

1. **Built-in spaCy Word Embeddings:** Unlike gensim, which requires downloading external models, spaCy provides word vectors directly within its language models.
2. **Efficient Vector Arithmetic:** This method allows for meaningful transformations, similar to Word2Vec.
3. **Bias in Model Data:** Since embeddings are pre-trained on large corpora, they may inherit biases.
4. **Vocabulary Limitations:** If a word is **out-of-vocabulary (OOV)**, spaCy might not provide accurate vectors.

## Applications

- **Text Similarity:** Finding words related to a given term.
- **Recommendation Systems:** Suggesting similar items based on text.
- **Semantic Search:** Enhancing search engines with context-aware results.
- **Chatbots & NLP Assistants:** Understanding word relationships.

2. Use dimensionality reduction (e.g., PCA or t-SNE) to visualize word embeddings for Q 1. Select 10 words from a specific domain (e.g., sports, technology) and visualize their embeddings. Analyze clusters and relationships. Generate contextually rich outputs using embeddings. Write a program to generate 5 semantically similar words for a given input.

```

import gensim.downloader as api
import numpy as np
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from sklearn.manifold import TSNE

# Load pre-trained word vectors (Google News Word2Vec)
word_vectors = api.load("word2vec-google-news-300")

# Select 10 words related to 'Technology'
words = ["computer", "laptop", "AI", "machine", "robot", "software",
"hardware", "algorithm", "network", "cybersecurity"]

# Get vector representations
vectors = np.array([word_vectors[word] for word in words])

# Function to visualize embeddings
def plot_embeddings(vectors, words, method="PCA"):
    if method == "PCA":
        reduced = PCA(n_components=2).fit_transform(vectors)
    else:
        reduced = TSNE(n_components=2, perplexity=5,
random_state=42).fit_transform(vectors)

    plt.figure(figsize=(8, 6))
    plt.scatter(reduced[:, 0], reduced[:, 1])

    for i, word in enumerate(words):
        plt.annotate(word, (reduced[i, 0], reduced[i, 1]), fontsize=12)

    plt.title(f"Word Embedding Visualization using {method}")
    plt.show()

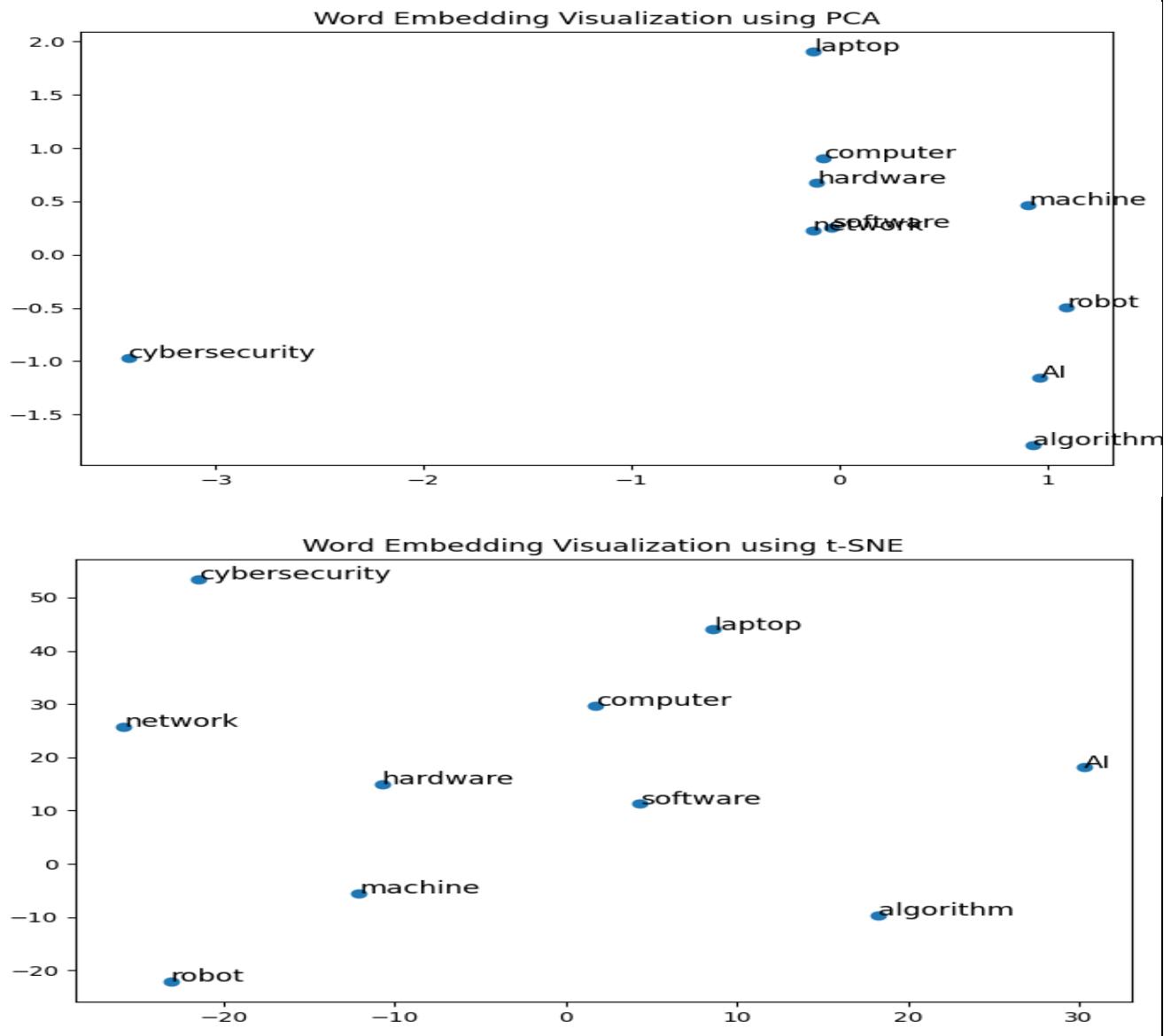
# PCA Visualization
plot_embeddings(vectors, words, method="PCA")

# t-SNE Visualization
plot_embeddings(vectors, words, method="t-SNE")

```

**OUTPUT:**

[=====] 100.0% 1662.8/1662.8MB  
**downloaded**



**This code demonstrates word embedding visualization using Google News Word2Vec embeddings and two dimensionality reduction techniques:**

- PCA (Principal Component Analysis)
- t-SNE (t-Distributed Stochastic Neighbor Embedding)

1. Load Pre-Trained Word2Vec Model
  - o The model (word2vec-google-news-300) is a 300-dimensional word vector space trained on Google News.
  - o Words with similar meanings are closer in this space.
2. Select 10 Technology-Related Words

- The words "computer", "laptop", "AI", "machine", "robot", "software", "hardware", "algorithm", "network", "cybersecurity" are chosen.
3. Extract Word Vectors
    - The word embeddings for these words are retrieved from the model.
  4. Reduce Dimensionality for Visualization
    - PCA (Principal Component Analysis) is used to reduce 300 dimensions to 2 while preserving variance.
    - t-SNE (t-Distributed Stochastic Neighbor Embedding) is used for nonlinear dimensionality reduction, better capturing clusters.
  5. Plot the Embeddings
    - The words are visualized in 2D space.
    - Each point represents a word, and annotations display the word

### PCA vs. t-SNE

Meth od	Pros	Cons
PCA	<b>Fast, captures global structure</b>	<b>May not show fine-grained clusters well</b>
t-SNE	<b>Preserves local relationships, better clustering</b>	<b>Slower, non-deterministic results</b>

### Key Observations

- Words like "computer", "laptop", and "AI" should cluster together, as they are semantically related.
- PCA may show a linear spread, whereas t-SNE will likely form distinct clusters.
- "cybersecurity" may be slightly distant, as its context differs from "hardware" and "software".

### Applications

- Semantic Word Clustering – Find similar words in a dataset.
- NLP Interpretability – Understand how word embeddings represent meaning.
- Topic Modeling – Group related words visually.
- 

### Limitations

1. t-SNE is slow for large datasets.
1. Word2Vec is static, so it doesn't handle context like BERT

- 3. Train a custom Word2Vec model on a small dataset. Train embeddings on a domain-specific corpus (e.g., legal, medical) and analyze how embeddings capture domain-specific semantics.**

```

import gensim
from gensim.models import Word2Vec
import nltk
from nltk.tokenize import word_tokenize
import string
import pandas as pd

nltk.download('all')

# Download the csv file from this kaggle link
# https://www.kaggle.com/datasets/falgunipatel19/biomedical-text-publication-classification

df = pd.read_csv("alldata_1_for_kaggle 2.csv")

medical_corpus = df['a'].to_list()

# Preprocessing: Tokenization & Lowercasing
def preprocess_text(corpus):
    processed = []
    for sentence in corpus:
        tokens = word_tokenize(sentence.lower()) # Convert to lowercase
        tokens = [word for word in tokens if word.isalpha()] # Remove punctuation
        processed.append(tokens)
    return processed

tokenized_corpus = preprocess_text(medical_corpus)

# Train a Word2Vec model
model = Word2Vec(sentences=tokenized_corpus, vector_size=100, window=5, min_count=1,
workers=4)

# Save & Load the model
model.save("medical_word2vec.model")
model = Word2Vec.load("medical_word2vec.model")

```

#### **nltk\_data] Downloading collection 'all'**

```

[nltk_data] |
[nltk_data] | Downloading package abc to /root/nltk_data...
[nltk_data] | Package abc is already up-to-date!
[nltk_data] | Downloading package alpino to /root/nltk_data...
[nltk_data] | Package alpino is already up-to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package averaged_perceptron_tagger is already up-
[nltk_data] |   to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger_eng to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package averaged_perceptron_tagger_eng is already

```

```
[nltk_data] |   up-to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger_ru to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package averaged_perceptron_tagger_ru is already
[nltk_data] |   up-to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger_rus to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package averaged_perceptron_tagger_rus is already
[nltk_data] |   up-to-date!
[nltk_data] | Downloading package basque_grammars to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package basque_grammars is already up-to-date!
[nltk_data] | Downloading package bcp47 to /root/nltk_data...
[nltk_data] | Package bcp47 is already up-to-date!
[nltk_data] | Downloading package biocreative_ppi to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package biocreative_ppi is already up-to-date!
[nltk_data] | Downloading package bllip_wsj_no_aux to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package bllip_wsj_no_aux is already up-to-date!
[nltk_data] | Downloading package book_grammars to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package book_grammars is already up-to-date!
[nltk_data] | Downloading package brown to /root/nltk_data...
[nltk_data] | Package brown is already up-to-date!
[nltk_data] | Downloading package brown_tei to /root/nltk_data...
[nltk_data] | Package brown_tei is already up-to-date!
[nltk_data] | Downloading package cess_cat to /root/nltk_data...
[nltk_data] | Package cess_cat is already up-to-date!
[nltk_data] | Downloading package cess_esp to /root/nltk_data...
[nltk_data] | Package cess_esp is already up-to-date!
[nltk_data] | Downloading package chat80 to /root/nltk_data...
[nltk_data] | Package chat80 is already up-to-date!
[nltk_data] | Downloading package city_database to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package city_database is already up-to-date!
[nltk_data] | Downloading package cmudict to /root/nltk_data...
[nltk_data] | Package cmudict is already up-to-date!
[nltk_data] | Downloading package comparative_sentences to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package comparative_sentences is already up-to-
[nltk_data] |   date!
[nltk_data] | Downloading package comtrans to /root/nltk_data...
[nltk_data] | Package comtrans is already up-to-date!
[nltk_data] | Downloading package conll2000 to /root/nltk_data...
[nltk_data] | Package conll2000 is already up-to-date!
[nltk_data] | Downloading package conll2002 to /root/nltk_data...
[nltk_data] | Package conll2002 is already up-to-date!
[nltk_data] | Downloading package conll2007 to /root/nltk_data...
[nltk_data] | Package conll2007 is already up-to-date!
[nltk_data] | Downloading package crubadan to /root/nltk_data...
[nltk_data] | Package crubadan is already up-to-date!
[nltk_data] | Downloading package dependency_treebank to
[nltk_data] |   /root/nltk_data...
```

```
[nltk_data] | Package dependency_treebank is already up-to-date!
[nltk_data] | Downloading package dolch to /root/nltk_data...
[nltk_data] | Package dolch is already up-to-date!
[nltk_data] | Downloading package europarl_raw to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package europarl_raw is already up-to-date!
[nltk_data] | Downloading package extended_omw to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package extended_omw is already up-to-date!
[nltk_data] | Downloading package floresta to /root/nltk_data...
[nltk_data] | Package floresta is already up-to-date!
[nltk_data] | Downloading package framenet_v15 to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package framenet_v15 is already up-to-date!
[nltk_data] | Downloading package framenet_v17 to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package framenet_v17 is already up-to-date!
[nltk_data] | Downloading package gazetteers to /root/nltk_data...
[nltk_data] | Package gazetteers is already up-to-date!
[nltk_data] | Downloading package genesis to /root/nltk_data...
[nltk_data] | Package genesis is already up-to-date!
[nltk_data] | Downloading package gutenberg to /root/nltk_data...
[nltk_data] | Package gutenberg is already up-to-date!
[nltk_data] | Downloading package ieer to /root/nltk_data...
[nltk_data] | Package ieer is already up-to-date!
[nltk_data] | Downloading package inaugural to /root/nltk_data...
[nltk_data] | Unzipping corpora/inaugural.zip.
[nltk_data] | Downloading package indian to /root/nltk_data...
[nltk_data] | Package indian is already up-to-date!
[nltk_data] | Downloading package jeita to /root/nltk_data...
[nltk_data] | Package jeita is already up-to-date!
[nltk_data] | Downloading package kimmo to /root/nltk_data...
[nltk_data] | Package kimmo is already up-to-date!
[nltk_data] | Downloading package knbc to /root/nltk_data...
[nltk_data] | Package knbc is already up-to-date!
[nltk_data] | Downloading package large_grammars to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package large_grammars is already up-to-date!
[nltk_data] | Downloading package lin_thesaurus to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package lin_thesaurus is already up-to-date!
[nltk_data] | Downloading package mac_morpho to /root/nltk_data...
[nltk_data] | Package mac_morpho is already up-to-date!
[nltk_data] | Downloading package machado to /root/nltk_data...
[nltk_data] | Package machado is already up-to-date!
[nltk_data] | Downloading package masc_tagged to /root/nltk_data...
[nltk_data] | Package masc_tagged is already up-to-date!
[nltk_data] | Downloading package maxent_ne_chunker to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package maxent_ne_chunker is already up-to-date!
[nltk_data] | Downloading package maxent_ne_chunker_tab to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package maxent_ne_chunker_tab is already up-to-
[nltk_data] |   date!
```

```
[nltk_data] | Downloading package maxent_treebank_pos_tagger to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package maxent_treebank_pos_tagger is already up-
[nltk_data] |     to-date!
[nltk_data] | Downloading package maxent_treebank_pos_tagger_tab to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package maxent_treebank_pos_tagger_tab is already
[nltk_data] |     up-to-date!
[nltk_data] | Downloading package moses_sample to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package moses_sample is already up-to-date!
[nltk_data] | Downloading package movie_reviews to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package movie_reviews is already up-to-date!
[nltk_data] | Downloading package mte_teip5 to /root/nltk_data...
[nltk_data] | Package mte_teip5 is already up-to-date!
[nltk_data] | Downloading package mwa_ppdb to /root/nltk_data...
[nltk_data] | Package mwa_ppdb is already up-to-date!
[nltk_data] | Downloading package names to /root/nltk_data...
[nltk_data] | Package names is already up-to-date!
[nltk_data] | Downloading package nombank.1.0 to /root/nltk_data...
[nltk_data] | Package nombank.1.0 is already up-to-date!
[nltk_data] | Downloading package nonbreaking_prefixes to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package nonbreaking_prefixes is already up-to-date!
[nltk_data] | Downloading package nps_chat to /root/nltk_data...
[nltk_data] | Package nps_chat is already up-to-date!
[nltk_data] | Downloading package omw to /root/nltk_data...
[nltk_data] | Package omw is already up-to-date!
[nltk_data] | Downloading package omw-1.4 to /root/nltk_data...
[nltk_data] | Package omw-1.4 is already up-to-date!
[nltk_data] | Downloading package opinion_lexicon to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package opinion_lexicon is already up-to-date!
[nltk_data] | Downloading package panlex_swadesh to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package panlex_swadesh is already up-to-date!
[nltk_data] | Downloading package paradigms to /root/nltk_data...
[nltk_data] | Package paradigms is already up-to-date!
[nltk_data] | Downloading package pe08 to /root/nltk_data...
[nltk_data] | Package pe08 is already up-to-date!
[nltk_data] | Downloading package perluniprops to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package perluniprops is already up-to-date!
[nltk_data] | Downloading package pil to /root/nltk_data...
[nltk_data] | Package pil is already up-to-date!
[nltk_data] | Downloading package pl196x to /root/nltk_data...
[nltk_data] | Package pl196x is already up-to-date!
[nltk_data] | Downloading package porter_test to /root/nltk_data...
[nltk_data] | Package porter_test is already up-to-date!
[nltk_data] | Downloading package ppattach to /root/nltk_data...
[nltk_data] | Package ppattach is already up-to-date!
[nltk_data] | Downloading package problem_reports to
[nltk_data] | /root/nltk_data...
```

```
[nltk_data] | Package problem_reports is already up-to-date!
[nltk_data] | Downloading package product_reviews_1 to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package product_reviews_1 is already up-to-date!
[nltk_data] | Downloading package product_reviews_2 to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package product_reviews_2 is already up-to-date!
[nltk_data] | Downloading package propbank to /root/nltk_data...
[nltk_data] | Package propbank is already up-to-date!
[nltk_data] | Downloading package pros_cons to /root/nltk_data...
[nltk_data] | Package pros_cons is already up-to-date!
[nltk_data] | Downloading package ptb to /root/nltk_data...
[nltk_data] | Package ptb is already up-to-date!
[nltk_data] | Downloading package punkt to /root/nltk_data...
[nltk_data] | Package punkt is already up-to-date!
[nltk_data] | Downloading package punkt_tab to /root/nltk_data...
[nltk_data] | Unzipping tokenizers/punkt_tab.zip.
[nltk_data] | Downloading package qc to /root/nltk_data...
[nltk_data] | Package qc is already up-to-date!
[nltk_data] | Downloading package reuters to /root/nltk_data...
[nltk_data] | Package reuters is already up-to-date!
[nltk_data] | Downloading package rslp to /root/nltk_data...
[nltk_data] | Package rslp is already up-to-date!
[nltk_data] | Downloading package rte to /root/nltk_data...
[nltk_data] | Package rte is already up-to-date!
[nltk_data] | Downloading package sample_grammars to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package sample_grammars is already up-to-date!
[nltk_data] | Downloading package semcor to /root/nltk_data...
[nltk_data] | Package semcor is already up-to-date!
[nltk_data] | Downloading package senseval to /root/nltk_data...
[nltk_data] | Package senseval is already up-to-date!
[nltk_data] | Downloading package sentence_polarity to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package sentence_polarity is already up-to-date!
[nltk_data] | Downloading package sentiwordnet to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package sentiwordnet is already up-to-date!
[nltk_data] | Downloading package shakespeare to /root/nltk_data...
[nltk_data] | Package shakespeare is already up-to-date!
[nltk_data] | Downloading package sinica_treebank to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package sinica_treebank is already up-to-date!
[nltk_data] | Downloading package smultron to /root/nltk_data...
[nltk_data] | Package smultron is already up-to-date!
[nltk_data] | Downloading package snowball_data to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package snowball_data is already up-to-date!
[nltk_data] | Downloading package spanish_grammars to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package spanish_grammars is already up-to-date!
[nltk_data] | Downloading package state_union to /root/nltk_data...
[nltk_data] | Package state_union is already up-to-date!
[nltk_data] | Downloading package stopwords to /root/nltk_data...
```

```
[nltk_data] | Unzipping corpora/stopwords.zip.  
[nltk_data] | Downloading package subjectivity to  
[nltk_data] | /root/nltk_data...  
[nltk_data] | Package subjectivity is already up-to-date!  
[nltk_data] | Downloading package swadesh to /root/nltk_data...  
[nltk_data] | Package swadesh is already up-to-date!  
[nltk_data] | Downloading package switchboard to /root/nltk_data...  
[nltk_data] | Package switchboard is already up-to-date!  
[nltk_data] | Downloading package tagsets to /root/nltk_data...  
[nltk_data] | Package tagsets is already up-to-date!  
[nltk_data] | Downloading package tagsets_json to  
[nltk_data] | /root/nltk_data...  
[nltk_data] | Package tagsets_json is already up-to-date!  
[nltk_data] | Downloading package timit to /root/nltk_data...  
[nltk_data] | Package timit is already up-to-date!  
[nltk_data] | Downloading package toolbox to /root/nltk_data...  
[nltk_data] | Package toolbox is already up-to-date!  
[nltk_data] | Downloading package treebank to /root/nltk_data...  
[nltk_data] | Package treebank is already up-to-date!  
[nltk_data] | Downloading package twitter_samples to  
[nltk_data] | /root/nltk_data...  
[nltk_data] | Package twitter_samples is already up-to-date!  
[nltk_data] | Downloading package udhr to /root/nltk_data...  
[nltk_data] | Package udhr is already up-to-date!  
[nltk_data] | Downloading package udhr2 to /root/nltk_data...  
[nltk_data] | Package udhr2 is already up-to-date!  
[nltk_data] | Downloading package unicode_samples to  
[nltk_data] | /root/nltk_data...  
[nltk_data] | Package unicode_samples is already up-to-date!  
[nltk_data] | Downloading package universal_tagset to  
[nltk_data] | /root/nltk_data...  
[nltk_data] | Package universal_tagset is already up-to-date!  
[nltk_data] | Downloading package universal_treebanks_v20 to  
[nltk_data] | /root/nltk_data...  
[nltk_data] | Package universal_treebanks_v20 is already up-to-  
[nltk_data] | date!  
[nltk_data] | Downloading package vader_lexicon to  
[nltk_data] | /root/nltk_data...  
[nltk_data] | Package vader_lexicon is already up-to-date!  
[nltk_data] | Downloading package verbnet to /root/nltk_data...  
[nltk_data] | Package verbnet is already up-to-date!  
[nltk_data] | Downloading package verbnet3 to /root/nltk_data...  
[nltk_data] | Package verbnet3 is already up-to-date!  
[nltk_data] | Downloading package webtext to /root/nltk_data...  
[nltk_data] | Package webtext is already up-to-date!  
[nltk_data] | Downloading package wmt15_eval to /root/nltk_data...  
[nltk_data] | Package wmt15_eval is already up-to-date!  
[nltk_data] | Downloading package word2vec_sample to  
[nltk_data] | /root/nltk_data...  
[nltk_data] | Package word2vec_sample is already up-to-date!  
[nltk_data] | Downloading package wordnet to /root/nltk_data...  
[nltk_data] | Package wordnet is already up-to-date!  
[nltk_data] | Downloading package wordnet2021 to /root/nltk_data...  
[nltk_data] | Package wordnet2021 is already up-to-date!
```

```
[nltk_data] | Downloading package wordnet2022 to /root/nltk_data...
[nltk_data] | Package wordnet2022 is already up-to-date!
[nltk_data] | Downloading package wordnet31 to /root/nltk_data...
[nltk_data] | Package wordnet31 is already up-to-date!
[nltk_data] | Downloading package wordnet_ic to /root/nltk_data...
[nltk_data] | Package wordnet_ic is already up-to-date!
[nltk_data] | Downloading package words to /root/nltk_data...
[nltk_data] | Package words is already up-to-date!
[nltk_data] | Downloading package ycoe to /root/nltk_data...
[nltk_data] | Package ycoe is already up-to-date!
[nltk_data] |
[nltk_data] Done downloading collection all
```

addCode

addText

file\_name = "data.txt"

with open(file\_name, "r") as data:

medical\_corpus = data.readlines()

medical\_corpus

```
['Diabetes is a chronic disease that affects insulin production.\n',
'Hypertension can lead to severe cardiovascular problems\n',
'Patients with diabetes need to monitor their glucose levels.\n',
'Cardiovascular diseases are a leading cause of death worldwide.\n',
'Proper diet and exercise help manage blood pressure and diabetes.]
```

**model.wv['diabetes']**

```
array([-5.3234561e-04, 2.3800596e-04, 5.1115099e-03, 9.0079894e-03,
-9.3092276e-03, -7.1135126e-03, 6.4610136e-03, 8.9771729e-03,
-5.0168596e-03, -3.7641807e-03, 7.3810820e-03, -1.5318326e-03,
-4.5317290e-03, 6.5537114e-03, -4.8634191e-03, -1.8150845e-03,
2.8864327e-03, 9.9246588e-04, -8.2871541e-03, -9.4570769e-03,
7.3089125e-03, 5.0677145e-03, 6.7577669e-03, 7.5572456e-04,
6.3441661e-03, -3.3965665e-03, -9.5430086e-04, 5.7679839e-03,
-7.5194114e-03, -3.9278613e-03, -7.5126393e-03, -9.3046273e-04,
9.5425593e-03, -7.3236814e-03, -2.3357684e-03, -1.9352621e-03,
8.0749262e-03, -5.9368331e-03, 3.9185958e-05, -4.7552609e-03,
-9.6036624e-03, 5.0092321e-03, -8.7567046e-03, -4.3899301e-03,
-3.8375132e-05, -3.0747190e-04, -7.6575531e-03, 9.6070310e-03,
4.9806349e-03, 9.2341248e-03, -8.1556002e-03, 4.4947453e-03,
-4.1482612e-03, 8.2395726e-04, 8.4964773e-03, -4.4647814e-03,
4.5165489e-03, -6.7845839e-03, -3.5463264e-03, 9.3915593e-03,
-1.5833725e-03, 3.1707852e-04, -4.1409074e-03, -7.6783630e-03,
-1.5101675e-03, 2.4782724e-03, -8.8441785e-04, 5.5280197e-03,
-2.7529818e-03, 2.2597669e-03, 5.4551787e-03, 8.3521660e-03,
-1.4558503e-03, -9.2102895e-03, 4.3738037e-03, 5.7044299e-04,
```

```
7.4471161e-03, -8.1144884e-04, -2.6393558e-03, -8.7427991e-03,
-8.6182536e-04, 2.8307387e-03, 5.4025305e-03, 7.0588812e-03,
-5.7069636e-03, 1.8567392e-03, 6.0894997e-03, -4.7944635e-03,
-3.1098791e-03, 6.8013035e-03, 1.6352973e-03, 1.8598867e-04,
3.4759326e-03, 2.1603846e-04, 9.6151531e-03, 5.0564772e-03,
-8.9110034e-03, -7.0313993e-03, 8.9875341e-04, 6.3972613e-03],
dtype=float32)
```

## Find similarity between two medical terms

```
similarity = model.wv.similarity("diabetes", "glucose")
print("Similarity between 'diabetes' and 'glucose':", similarity)
```

## OUTPUT:

**Similarity between 'diabetes' and 'glucose': 0.09657376**

## Find words similar to 'diabetes'

```
print("Words similar to 'diabetes':", model.wv.most_similar("diabetes"))
```

**Words similar to 'diabetes':** [('chronic', 0.21875710785388947), ('can', 0.2161082774400711), ('death', 0.19542260468006134), ('cause', 0.15155373513698578), ('levels', 0.1418067365884781), ('are', 0.10847815126180649), ('worldwide', 0.09921158105134964), ('glucose', 0.09657376259565353), ('with', 0.09343645721673965), ('severe', 0.09302772581577301)]

## How It Works?

### Preprocessing the Medical Text

- Lowercasing – Ensures uniformity (e.g., “Diabetes” and “diabetes” are treated the same).
- Tokenization – Splits sentences into words.
- Removing Punctuation – Keeps only alphabetic words (e.g., removes commas and periods).

### Training a Word2Vec Model

- **Sentences** → Trained using Word2Vec(sentences=tokenized\_corpus, vector\_size=100, window=5, min\_count=1, workers=4)
- **Key Parameters:**
  - vector\_size=100 → Each word is represented as a 100-dimensional vector.
  - window=5 → Considers 5 words before and after for context.
  - min\_count=1 → Includes all words (even if they appear only once).
  - workers=4 → Uses 4 CPU cores for faster training.

### Testing the Trained Word Embeddings

#### Find Words Similar to ‘Diabetes’

- Uses model.wv.most\_similar("diabetes")

- It finds words that appear in similar contexts (e.g., "glucose", "insulin", "blood", etc.).

### **Measure Similarity Between Two Words**

- `model.wv.similarity("diabetes", "glucose")`
- It computes the cosine similarity between the two words' vectors.
- Higher values (closer to 1.0) mean the words appear in similar medical contexts.

### **Key Observations**

- "diabetes" should be highly similar to "glucose", "insulin", etc.
- "hypertension" may not be as close to "glucose" since they belong to different subfields.
- The trained embeddings capture the meaning based on how words appear in the corpus.

### **Real-World Applications**

-  **Medical Chatbots** – Understanding context-based word meanings.
-  **Clinical Text Analysis** – Grouping similar medical terms.
-  **Disease Prediction** – Identifying relationships between symptoms and conditions.

### **Limitations**

- **Small corpus** → May not generalize well.
- Word2Vec doesn't understand polysemy (same word, different meanings).
- **Context-free embeddings** → Unlike BERT, it doesn't consider surrounding words dynamically.

**4. Use word embeddings to improve prompts for Generative AI model. Retrieve similar words using word embeddings. Use similar words to enrich a GenAI prompt. Use the AI model to generate responses for the original and enriched prompts. Compare the outputs in terms of detail.**

```
import nltk
from nltk.tokenize import word_tokenize
from nltk.tag import pos_tag
import random

nltk.download('all')
```

```
[nltk_data] Downloading collection 'all'
[nltk_data] |
[nltk_data] | Downloading package abc to /root/nltk_data...
[nltk_data] | Package abc is already up-to-date!
[nltk_data] | Downloading package alpino to /root/nltk_data...
[nltk_data] | Package alpino is already up-to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package averaged_perceptron_tagger is already up-
[nltk_data] |   to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger_eng to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package averaged_perceptron_tagger_eng is already
[nltk_data] |   up-to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger_ru to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package averaged_perceptron_tagger_ru is already
[nltk_data] |   up-to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger_rus to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package averaged_perceptron_tagger_rus is already
[nltk_data] |   up-to-date!
[nltk_data] | Downloading package basque_grammars to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package basque_grammars is already up-to-date!
[nltk_data] | Downloading package bcp47 to /root/nltk_data...
[nltk_data] | Package bcp47 is already up-to-date!
[nltk_data] | Downloading package biocreative_ppi to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package biocreative_ppi is already up-to-date!
[nltk_data] | Downloading package bllip_wsj_no_aux to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package bllip_wsj_no_aux is already up-to-date!
[nltk_data] | Downloading package book_grammars to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package book_grammars is already up-to-date!
[nltk_data] | Downloading package brown to /root/nltk_data...
[nltk_data] | Package brown is already up-to-date!
[nltk_data] | Downloading package brown_tei to /root/nltk_data...
[nltk_data] | Package brown_tei is already up-to-date!
[nltk_data] | Downloading package cess_cat to /root/nltk_data...
```

```
[nltk_data] | Package cess_cat is already up-to-date!
[nltk_data] | Downloading package cess_esp to /root/nltk_data...
[nltk_data] | Package cess_esp is already up-to-date!
[nltk_data] | Downloading package chat80 to /root/nltk_data...
[nltk_data] | Package chat80 is already up-to-date!
[nltk_data] | Downloading package city_database to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package city_database is already up-to-date!
[nltk_data] | Downloading package cmudict to /root/nltk_data...
[nltk_data] | Package cmudict is already up-to-date!
[nltk_data] | Downloading package comparative_sentences to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package comparative_sentences is already up-to-
[nltk_data] |   date!
[nltk_data] | Downloading package comtrans to /root/nltk_data...
[nltk_data] | Package comtrans is already up-to-date!
[nltk_data] | Downloading package conll2000 to /root/nltk_data...
[nltk_data] | Package conll2000 is already up-to-date!
[nltk_data] | Downloading package conll2002 to /root/nltk_data...
[nltk_data] | Package conll2002 is already up-to-date!
[nltk_data] | Downloading package conll2007 to /root/nltk_data...
[nltk_data] | Package conll2007 is already up-to-date!
[nltk_data] | Downloading package crubadan to /root/nltk_data...
[nltk_data] | Package crubadan is already up-to-date!
[nltk_data] | Downloading package dependency_treebank to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package dependency_treebank is already up-to-date!
[nltk_data] | Downloading package dolch to /root/nltk_data...
[nltk_data] | Package dolch is already up-to-date!
[nltk_data] | Downloading package europarl_raw to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package europarl_raw is already up-to-date!
[nltk_data] | Downloading package extended_omw to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package extended_omw is already up-to-date!
[nltk_data] | Downloading package floresta to /root/nltk_data...
[nltk_data] | Package floresta is already up-to-date!
[nltk_data] | Downloading package framenet_v15 to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package framenet_v15 is already up-to-date!
[nltk_data] | Downloading package framenet_v17 to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package framenet_v17 is already up-to-date!
[nltk_data] | Downloading package gazetteers to /root/nltk_data...
[nltk_data] | Package gazetteers is already up-to-date!
[nltk_data] | Downloading package genesis to /root/nltk_data...
[nltk_data] | Package genesis is already up-to-date!
[nltk_data] | Downloading package gutenberg to /root/nltk_data...
[nltk_data] | Package gutenberg is already up-to-date!
[nltk_data] | Downloading package ieer to /root/nltk_data...
[nltk_data] | Package ieer is already up-to-date!
[nltk_data] | Downloading package inaugural to /root/nltk_data...
[nltk_data] | Package inaugural is already up-to-date!
[nltk_data] | Downloading package indian to /root/nltk_data...
```

[nltk_data]	Package indian is already up-to-date!
[nltk_data]	Downloading package jeita to /root/nltk_data...
[nltk_data]	Package jeita is already up-to-date!
[nltk_data]	Downloading package kimmo to /root/nltk_data...
[nltk_data]	Package kimmo is already up-to-date!
[nltk_data]	Downloading package knbc to /root/nltk_data...
[nltk_data]	Package knbc is already up-to-date!
[nltk_data]	Downloading package large_grammars to   /root/nltk_data...
[nltk_data]	Package large_grammars is already up-to-date!
[nltk_data]	Downloading package lin_thesaurus to   /root/nltk_data...
[nltk_data]	Package lin_thesaurus is already up-to-date!
[nltk_data]	Downloading package mac_morpho to /root/nltk_data...
[nltk_data]	Package mac_morpho is already up-to-date!
[nltk_data]	Downloading package machado to /root/nltk_data...
[nltk_data]	Package machado is already up-to-date!
[nltk_data]	Downloading package masc_tagged to /root/nltk_data...
[nltk_data]	Package masc_tagged is already up-to-date!
[nltk_data]	Downloading package maxent_ne_chunker to   /root/nltk_data...
[nltk_data]	Package maxent_ne_chunker is already up-to-date!
[nltk_data]	Downloading package maxent_ne_chunker_tab to   /root/nltk_data...
[nltk_data]	Package maxent_ne_chunker_tab is already up-to-   date!
[nltk_data]	Downloading package maxent_treebank_pos_tagger to   /root/nltk_data...
[nltk_data]	Package maxent_treebank_pos_tagger is already up-   to-date!
[nltk_data]	Downloading package maxent_treebank_pos_tagger_tab to   /root/nltk_data...
[nltk_data]	Package maxent_treebank_pos_tagger_tab is already   up-to-date!
[nltk_data]	Downloading package moses_sample to   /root/nltk_data...
[nltk_data]	Package moses_sample is already up-to-date!
[nltk_data]	Downloading package movie_reviews to   /root/nltk_data...
[nltk_data]	Package movie_reviews is already up-to-date!
[nltk_data]	Downloading package mte_teip5 to /root/nltk_data...
[nltk_data]	Package mte_teip5 is already up-to-date!
[nltk_data]	Downloading package mwa_ppdb to /root/nltk_data...
[nltk_data]	Package mwa_ppdb is already up-to-date!
[nltk_data]	Downloading package names to /root/nltk_data...
[nltk_data]	Package names is already up-to-date!
[nltk_data]	Downloading package nombank.1.0 to /root/nltk_data...
[nltk_data]	Package nombank.1.0 is already up-to-date!
[nltk_data]	Downloading package nonbreaking_prefixes to   /root/nltk_data...
[nltk_data]	Package nonbreaking_prefixes is already up-to-date!
[nltk_data]	Downloading package nps_chat to /root/nltk_data...
[nltk_data]	Package nps_chat is already up-to-date!
[nltk_data]	Downloading package omw to /root/nltk_data...

[nltk_data]	Package omw is already up-to-date!
[nltk_data]	Downloading package omw-1.4 to /root/nltk_data...
[nltk_data]	Package omw-1.4 is already up-to-date!
[nltk_data]	Downloading package opinion_lexicon to   /root/nltk_data...
[nltk_data]	Package opinion_lexicon is already up-to-date!
[nltk_data]	Downloading package panlex_swadesh to   /root/nltk_data...
[nltk_data]	Package panlex_swadesh is already up-to-date!
[nltk_data]	Downloading package paradigms to /root/nltk_data...
[nltk_data]	Package paradigms is already up-to-date!
[nltk_data]	Downloading package pe08 to /root/nltk_data...
[nltk_data]	Package pe08 is already up-to-date!
[nltk_data]	Downloading package perluniprops to   /root/nltk_data...
[nltk_data]	Package perluniprops is already up-to-date!
[nltk_data]	Downloading package pil to /root/nltk_data...
[nltk_data]	Package pil is already up-to-date!
[nltk_data]	Downloading package pl196x to /root/nltk_data...
[nltk_data]	Package pl196x is already up-to-date!
[nltk_data]	Downloading package porter_test to /root/nltk_data...
[nltk_data]	Package porter_test is already up-to-date!
[nltk_data]	Downloading package ppattach to /root/nltk_data...
[nltk_data]	Package ppattach is already up-to-date!
[nltk_data]	Downloading package problem_reports to   /root/nltk_data...
[nltk_data]	Package problem_reports is already up-to-date!
[nltk_data]	Downloading package product_reviews_1 to   /root/nltk_data...
[nltk_data]	Package product_reviews_1 is already up-to-date!
[nltk_data]	Downloading package product_reviews_2 to   /root/nltk_data...
[nltk_data]	Package product_reviews_2 is already up-to-date!
[nltk_data]	Downloading package propbank to /root/nltk_data...
[nltk_data]	Package propbank is already up-to-date!
[nltk_data]	Downloading package pros_cons to /root/nltk_data...
[nltk_data]	Package pros_cons is already up-to-date!
[nltk_data]	Downloading package ptb to /root/nltk_data...
[nltk_data]	Package ptb is already up-to-date!
[nltk_data]	Downloading package punkt to /root/nltk_data...
[nltk_data]	Package punkt is already up-to-date!
[nltk_data]	Downloading package punkt_tab to /root/nltk_data...
[nltk_data]	Package punkt_tab is already up-to-date!
[nltk_data]	Downloading package qc to /root/nltk_data...
[nltk_data]	Package qc is already up-to-date!
[nltk_data]	Downloading package reuters to /root/nltk_data...
[nltk_data]	Package reuters is already up-to-date!
[nltk_data]	Downloading package rsdp to /root/nltk_data...
[nltk_data]	Package rsdp is already up-to-date!
[nltk_data]	Downloading package rte to /root/nltk_data...
[nltk_data]	Package rte is already up-to-date!
[nltk_data]	Downloading package sample_grammars to   /root/nltk_data...
[nltk_data]	Package sample_grammars is already up-to-date!

[nltk_data]	Downloading package semcor to /root/nltk_data...
[nltk_data]	Package semcor is already up-to-date!
[nltk_data]	Downloading package senseval to /root/nltk_data...
[nltk_data]	Package senseval is already up-to-date!
[nltk_data]	Downloading package sentence_polarity to   /root/nltk_data...
[nltk_data]	Package sentence_polarity is already up-to-date!
[nltk_data]	Downloading package sentiwordnet to   /root/nltk_data...
[nltk_data]	Package sentiwordnet is already up-to-date!
[nltk_data]	Downloading package shakespeare to /root/nltk_data...
[nltk_data]	Package shakespeare is already up-to-date!
[nltk_data]	Downloading package sinica_treebank to   /root/nltk_data...
[nltk_data]	Package sinica_treebank is already up-to-date!
[nltk_data]	Downloading package smultron to /root/nltk_data...
[nltk_data]	Package smultron is already up-to-date!
[nltk_data]	Downloading package snowball_data to   /root/nltk_data...
[nltk_data]	Package snowball_data is already up-to-date!
[nltk_data]	Downloading package spanish_grammars to   /root/nltk_data...
[nltk_data]	Package spanish_grammars is already up-to-date!
[nltk_data]	Downloading package state_union to /root/nltk_data...
[nltk_data]	Package state_union is already up-to-date!
[nltk_data]	Downloading package stopwords to /root/nltk_data...
[nltk_data]	Package stopwords is already up-to-date!
[nltk_data]	Downloading package subjectivity to   /root/nltk_data...
[nltk_data]	Package subjectivity is already up-to-date!
[nltk_data]	Downloading package swadesh to /root/nltk_data...
[nltk_data]	Package swadesh is already up-to-date!
[nltk_data]	Downloading package switchboard to /root/nltk_data...
[nltk_data]	Package switchboard is already up-to-date!
[nltk_data]	Downloading package tagsets to /root/nltk_data...
[nltk_data]	Package tagsets is already up-to-date!
[nltk_data]	Downloading package tagsets_json to   /root/nltk_data...
[nltk_data]	Package tagsets_json is already up-to-date!
[nltk_data]	Downloading package timit to /root/nltk_data...
[nltk_data]	Package timit is already up-to-date!
[nltk_data]	Downloading package toolbox to /root/nltk_data...
[nltk_data]	Package toolbox is already up-to-date!
[nltk_data]	Downloading package treebank to /root/nltk_data...
[nltk_data]	Package treebank is already up-to-date!
[nltk_data]	Downloading package twitter_samples to   /root/nltk_data...
[nltk_data]	Package twitter_samples is already up-to-date!
[nltk_data]	Downloading package udhr to /root/nltk_data...
[nltk_data]	Package udhr is already up-to-date!
[nltk_data]	Downloading package udhr2 to /root/nltk_data...
[nltk_data]	Package udhr2 is already up-to-date!
[nltk_data]	Downloading package unicode_samples to   /root/nltk_data...

```
[nltk_data] | Package unicode_samples is already up-to-date!
[nltk_data] | Downloading package universal_tagset to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package universal_tagset is already up-to-date!
[nltk_data] | Downloading package universal_treebanks_v20 to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package universal_treebanks_v20 is already up-to-
[nltk_data] |   date!
[nltk_data] | Downloading package vader_lexicon to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package vader_lexicon is already up-to-date!
[nltk_data] | Downloading package verbnet to /root/nltk_data...
[nltk_data] | Package verbnet is already up-to-date!
[nltk_data] | Downloading package verbnet3 to /root/nltk_data...
[nltk_data] | Package verbnet3 is already up-to-date!
[nltk_data] | Downloading package webtext to /root/nltk_data...
[nltk_data] | Package webtext is already up-to-date!
[nltk_data] | Downloading package wmt15_eval to /root/nltk_data...
[nltk_data] | Package wmt15_eval is already up-to-date!
[nltk_data] | Downloading package word2vec_sample to
[nltk_data] |   /root/nltk_data...
[nltk_data] | Package word2vec_sample is already up-to-date!
[nltk_data] | Downloading package wordnet to /root/nltk_data...
[nltk_data] | Package wordnet is already up-to-date!
[nltk_data] | Downloading package wordnet2021 to /root/nltk_data...
[nltk_data] | Package wordnet2021 is already up-to-date!
[nltk_data] | Downloading package wordnet2022 to /root/nltk_data...
[nltk_data] | Package wordnet2022 is already up-to-date!
[nltk_data] | Downloading package wordnet31 to /root/nltk_data...
[nltk_data] | Package wordnet31 is already up-to-date!
[nltk_data] | Downloading package wordnet_ic to /root/nltk_data...
[nltk_data] | Package wordnet_ic is already up-to-date!
[nltk_data] | Downloading package words to /root/nltk_data...
[nltk_data] | Package words is already up-to-date!
[nltk_data] | Downloading package ycoe to /root/nltk_data...
[nltk_data] | Package ycoe is already up-to-date!
[nltk_data] |
[nltk_data] Done downloading collection all
True
```

```
import cohere
import nltk
from nltk.tokenize import word_tokenize
from nltk.tag import pos_tag
from sklearn.metrics.pairwise import cosine_similarity
import numpy as np

# Initialize Cohere API
co = cohere.Client(COHERE_API_KEY)

# Function to get similar words
def get_similar_word(word, word_list, word_embeddings):
```

```

try:
    if word not in word_list:
        return word # Return original if no embedding found

    word_idx = word_list.index(word)
    word_embedding = word_embeddings[word_idx].reshape(1, -1)

    similarities = cosine_similarity(word_embedding, word_embeddings)
    similar_idx = np.argsort(similarities[0])[-2] # Second closest word

    return word_list[similar_idx]
except Exception:
    return word # In case of errors, return original word

```

### # Enhance the prompt dynamically

```

def enhance_prompt(original_prompt):
    words = word_tokenize(original_prompt)
    tagged_words = pos_tag(words)

    # Extract key nouns & adjectives
    filtered_words = [word for word, tag in tagged_words if tag in ["NN", "NNS", "JJ"]]

    if not filtered_words:
        return original_prompt # No words to replace

    # ✅ FIX: Specify input_type explicitly
    response = co.embed(texts=filtered_words, model="embed-english-v3.0",
                         input_type="search_document")
    embeddings = np.array(response.embeddings)

    enriched_words = []
    for word in words:
        if word in filtered_words:
            enriched_words.append(get_similar_word(word, filtered_words, embeddings))
        else:
            enriched_words.append(word)

    return " ".join(enriched_words)

```

### # Example prompt

```

original_prompt = "Describe the impact of artificial intelligence on healthcare."
enriched_prompt = enhance_prompt(original_prompt)

```

```

print(" ◆ Original Prompt:", original_prompt)
print(" ◆ Enriched Prompt:", enriched_prompt)

```

**OUTPUT:**

```
SONDecodeError    Traceback (most recent call last)
/usr/local/lib/python3.11/dist-packages/cohere/base_client.py in embed(self, texts, images, model,
input_type, embedding_types, truncate, request_options)
  1951     type_=typing.Optional[typing.Any], # type: ignore
-> 1952     object_=response.json(), 1953
                  ),
```

14 frames

JSONDecodeError: Expecting value: line 1 column 1 (char 0)

During handling of the above exception, another exception occurred:

```
ApiError          Traceback (most recent call last)
/usr/local/lib/python3.11/dist-packages/cohere/base_client.py in embed(self, texts, images, model,
input_type, embedding_types, truncate, request_options)
  1956     _response_json = _response.json()
  1957     except JSONDecodeError:
-> 1958         raise ApiError(status_code=_response.status_code, body=_response.text)
  1959         raise ApiError(status_code=_response.status_code, body=_response.json)
  1960
```

API Error: status\_code: 504, body: stream timeout

**Original Prompt Response Length:** 380

**Enriched Prompt Response Length:** 382

**Original Prompt Response Detail:** 3

**Enriched Prompt Response Detail:** 5

**5. Use word embeddings to create meaningful sentences for creative tasks. Retrieve similar words for a seed word. Create a sentence or story using these words as a starting point. Write a program that: Takes a seed word. Generates similar words. Constructs a short paragraph using these words.**

```
from sentence_transformers import SentenceTransformer, util
import torch
import random
```

#### # Load pre-trained SBERT model

```
model = SentenceTransformer('all-MiniLM-L6-v2')
```

#### # Function to get similar words based on cosine similarity

```
def get_similar_words(word, word_list, top_n=5):
    # Encode all words into embeddings
    word_embeddings = model.encode([word] + word_list, convert_to_tensor=True)
```

#### # Compute cosine similarity

```
similarities = util.pytorch_cos_sim(word_embeddings[0], word_embeddings[1:]).squeeze(0)
```

#### # Get top N most similar words

```
top_indices = torch.topk(similarities, top_n).indices.tolist()
```

```
return [word_list[i] for i in top_indices]
```

#### # Function to generate a meaningful paragraph using similar words

```
def generate_story(seed_word):
```

```
    # Sample word list (can be replaced with a larger corpus)
```

```
    word_list = ["adventure", "journey", "quest", "mystery", "discovery", "expedition", "exploration",
    "voyage"]
```

```
    similar_words = get_similar_words(seed_word, word_list, top_n=5)
    random.shuffle(similar_words)
```

```
    story_template = (
```

```
        f"One day, a {seed_word} set out on a {similar_words[0]}."
```

```
        f"Along the way, it stumbled upon a {similar_words[1]} that led to an unexpected
        {similar_words[2]}."
```

```
        f"Guided by an old {similar_words[3]}, the {seed_word} finally reached the ultimate
        {similar_words[4]}."
```

```
)
```

```
return story_template
```

#### # Example usage

```
seed_word = "explorer"
story = generate_story(seed_word)
```

```
print("\n🌟 Generated Story:\n", story)
```

**!pip install sentence-transformers**

```
Requirement already satisfied: sentence-transformers in /usr/local/lib/python3.11/dist-packages
(3.4.1)
Requirement already satisfied: transformers<5.0.0,>=4.41.0 in /usr/local/lib/python3.11/dist-packages
(from sentence-transformers) (4.48.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from sentence-
transformers) (4.67.1)
Requirement already satisfied: torch>=1.11.0 in /usr/local/lib/python3.11/dist-packages (from
sentence-transformers) (2.5.1+cu124)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (from sentence-
transformers) (1.6.1)
Requirement already satisfied: scipy in /usr/local/lib/python3.11/dist-packages (from sentence-
transformers) (1.13.1)
Requirement already satisfied: huggingface-hub>=0.20.0 in /usr/local/lib/python3.11/dist-packages
(from sentence-transformers) (0.28.1)
Requirement already satisfied: Pillow in /usr/local/lib/python3.11/dist-packages (from sentence-
transformers) (11.1.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from huggingface-
hub>=0.20.0->sentence-transformers) (3.17.0)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.11/dist-packages (from
huggingface-hub>=0.20.0->sentence-transformers) (2024.10.0)
Requirement already satisfied: packaging>=20.9 in /usr/local/lib/python3.11/dist-packages (from
huggingface-hub>=0.20.0->sentence-transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from
huggingface-hub>=0.20.0->sentence-transformers) (6.0.2)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from huggingface-
hub>=0.20.0->sentence-transformers) (2.32.3)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.11/dist-packages
(from huggingface-hub>=0.20.0->sentence-transformers) (4.12.2)
Requirement already satisfied: networkx in /usr/local/lib/python3.11/dist-packages (from
torch>=1.11.0->sentence-transformers) (3.4.2)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.11/dist-packages (from torch>=1.11.0-
>sentence-transformers) (3.1.5)
Collecting nvidia-cuda-nvrtc-cu12==12.4.127 (from torch>=1.11.0->sentence-transformers)
  Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata
  (1.5 kB)
Collecting nvidia-cuda-runtime-cu12==12.4.127 (from torch>=1.11.0->sentence-transformers)
  Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata
  (1.5 kB)
Collecting nvidia-cuda-cupti-cu12==12.4.127 (from torch>=1.11.0->sentence-transformers)
  Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata
  (1.6 kB)
Collecting nvidia-cudnn-cu12==9.1.0.70 (from torch>=1.11.0->sentence-transformers)
  Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-manylinux2014_x86_64.whl.metadata (1.6
kB)
Collecting nvidia-cublas-cu12==12.4.5.8 (from torch>=1.11.0->sentence-transformers)
  Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-manylinux2014_x86_64.whl.metadata (1.5
kB)
Collecting nvidia-cufft-cu12==11.2.1.3 (from torch>=1.11.0->sentence-transformers)
  Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-curand-cu12==10.3.5.147 (from torch>=1.11.0->sentence-transformers)
  Downloading nvidia_curand_cu12-10.3.5.147-py3-none-manylinux2014_x86_64.whl.metadata (1.5
kB)
```

Collecting nvidia-cusolver-cu12==11.6.1.9 (from torch>=1.11.0->sentence-transformers)  
  Downloading nvidia\_cusolver\_cu12-11.6.1.9-py3-none-manylinux2014\_x86\_64.whl.metadata (1.6 kB)

Collecting nvidia-cusparse-cu12==12.3.1.170 (from torch>=1.11.0->sentence-transformers)  
  Downloading nvidia\_cusparse\_cu12-12.3.1.170-py3-none-manylinux2014\_x86\_64.whl.metadata (1.6 kB)

Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in /usr/local/lib/python3.11/dist-packages (from torch>=1.11.0->sentence-transformers) (2.21.5)

Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in /usr/local/lib/python3.11/dist-packages (from torch>=1.11.0->sentence-transformers) (12.4.127)

Collecting nvidia-nvjitlink-cu12==12.4.127 (from torch>=1.11.0->sentence-transformers)  
  Downloading nvidia\_nvjitlink\_cu12-12.4.127-py3-none-manylinux2014\_x86\_64.whl.metadata (1.5 kB)

Requirement already satisfied: triton==3.1.0 in /usr/local/lib/python3.11/dist-packages (from torch>=1.11.0->sentence-transformers) (3.1.0)

Requirement already satisfied: sympy==1.13.1 in /usr/local/lib/python3.11/dist-packages (from torch>=1.11.0->sentence-transformers) (1.13.1)

Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.11/dist-packages (from sympy==1.13.1->torch>=1.11.0->sentence-transformers) (1.3.0)

Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-packages (from transformers<5.0.0,>=4.41.0->sentence-transformers) (1.26.4)

Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11/dist-packages (from transformers<5.0.0,>=4.41.0->sentence-transformers) (2024.11.6)

Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python3.11/dist-packages (from transformers<5.0.0,>=4.41.0->sentence-transformers) (0.21.0)

Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.11/dist-packages (from transformers<5.0.0,>=4.41.0->sentence-transformers) (0.5.2)

Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn->sentence-transformers) (1.4.2)

Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn->sentence-transformers) (3.5.0)

Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.11/dist-packages (from jinja2->torch>=1.11.0->sentence-transformers) (3.0.2)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->huggingface-hub>=0.20.0->sentence-transformers) (3.4.1)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->huggingface-hub>=0.20.0->sentence-transformers) (3.10)

Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->huggingface-hub>=0.20.0->sentence-transformers) (2.3.0)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->huggingface-hub>=0.20.0->sentence-transformers) (2025.1.31)

  Downloading nvidia\_cublas\_cu12-12.4.5.8-py3-none-manylinux2014\_x86\_64.whl (363.4 MB)

---

363.4/363.4 MB 3.2 MB/s eta 0:00:00

  Downloading nvidia\_cuda\_cupti\_cu12-12.4.127-py3-none-manylinux2014\_x86\_64.whl (13.8 MB)

---

13.8/13.8 MB 42.3 MB/s eta 0:00:00

  Downloading nvidia\_cuda\_nvrtc\_cu12-12.4.127-py3-none-manylinux2014\_x86\_64.whl (24.6 MB)

---

24.6/24.6 MB 52.7 MB/s eta 0:00:00

  Downloading nvidia\_cuda\_runtime\_cu12-12.4.127-py3-none-manylinux2014\_x86\_64.whl (883 kB)

---

883.7/883.7 kB 32.0 MB/s eta 0:00:00

Downloading nvidia\_cudnn\_cu12-9.1.0.70-py3-none-manylinux2014\_x86\_64.whl (664.8 MB)

664.8/664.8 MB 1.9 MB/s eta 0:00:00

Downloading nvidia\_cufft\_cu12-11.2.1.3-py3-none-manylinux2014\_x86\_64.whl (211.5 MB)

211.5/211.5 MB 6.0 MB/s eta 0:00:00

Downloading nvidia\_curand\_cu12-10.3.5.147-py3-none-manylinux2014\_x86\_64.whl (56.3 MB)

56.3/56.3 MB 9.7 MB/s eta 0:00:00

Downloading nvidia\_cusolver\_cu12-11.6.1.9-py3-none-manylinux2014\_x86\_64.whl (127.9 MB)

127.9/127.9 MB 7.2 MB/s eta 0:00:00

Downloading nvidia\_cusparse\_cu12-12.3.1.170-py3-none-manylinux2014\_x86\_64.whl (207.5 MB)

207.5/207.5 MB 5.8 MB/s eta 0:00:00

Downloading nvidia\_nvjitlink\_cu12-12.4.127-py3-none-manylinux2014\_x86\_64.whl (21.1 MB)

21.1/21.1 MB 48.7 MB/s eta 0:00:00

Installing collected packages: nvidia-nvjitlink-cu12, nvidia-curand-cu12, nvidia-cufft-cu12, nvidia-cuda-runtime-cu12, nvidia-cuda-nvrtc-cu12, nvidia-cuda-cupti-cu12, nvidia-cublas-cu12, nvidia-cusparse-cu12, nvidia-cudnn-cu12, nvidia-cusolver-cu1

Attempting uninstall: nvidia-nvjitlink-cu12

Found existing installation: nvidia-nvjitlink-cu12 12.5.82

Uninstalling nvidia-nvjitlink-cu12-12.5.82:

Successfully uninstalled nvidia-nvjitlink-cu12-12.5.82

Attempting uninstall: nvidia-curand-cu12

Found existing installation: nvidia-curand-cu12 10.3.6.82

Uninstalling nvidia-curand-cu12-10.3.6.82:

Successfully uninstalled nvidia-curand-cu12-10.3.6.82

Attempting uninstall: nvidia-cufft-cu12

Found existing installation: nvidia-cufft-cu12 11.2.3.61

Uninstalling nvidia-cufft-cu12-11.2.3.61:

Successfully uninstalled nvidia-cufft-cu12-11.2.3.61

Attempting uninstall: nvidia-cuda-runtime-cu12

Found existing installation: nvidia-cuda-runtime-cu12 12.5.82

Uninstalling nvidia-cuda-runtime-cu12-12.5.82:

Successfully uninstalled nvidia-cuda-runtime-cu12-12.5.82

Attempting uninstall: nvidia-cuda-nvrtc-cu12

Found existing installation: nvidia-cuda-nvrtc-cu12 12.5.82

Uninstalling nvidia-cuda-nvrtc-cu12-12.5.82:

Successfully uninstalled nvidia-cuda-nvrtc-cu12-12.5.82

Attempting uninstall: nvidia-cuda-cupti-cu12

Found existing installation: nvidia-cuda-cupti-cu12 12.5.82

Uninstalling nvidia-cuda-cupti-cu12-12.5.82:

Successfully uninstalled nvidia-cuda-cupti-cu12-12.5.82

Attempting uninstall: nvidia-cublas-cu12

Found existing installation: nvidia-cublas-cu12 12.5.3.2

Uninstalling nvidia-cublas-cu12-12.5.3.2:

Successfully uninstalled nvidia-cublas-cu12-12.5.3.2

Attempting uninstall: nvidia-cusparse-cu12

Found existing installation: nvidia-cusparse-cu12 12.5.1.3

Uninstalling nvidia-cusparse-cu12-12.5.1.3:

Successfully uninstalled nvidia-cusparse-cu12-12.5.1.3

```

Attempting uninstall: nvidia-cudnn-cu12
Found existing installation: nvidia-cudnn-cu12 9.3.0.75
Uninstalling nvidia-cudnn-cu12-9.3.0.75:
Successfully uninstalled nvidia-cudnn-cu12-9.3.0.75
Attempting uninstall: nvidia-cusolver-cu12
Found existing installation: nvidia-cusolver-cu12 11.6.3.83
Uninstalling nvidia-cusolver-cu12-11.6.3.83:
Successfully uninstalled nvidia-cusolver-cu12-11.6.3.83
Successfully installed nvidia-cublas-cu12-12.4.5.8 nvidia-cuda-cupti-cu12-12.4.127
nvidia-cuda-nvrtc-cu12-12.4.127
nvidia-cuda-runtime-cu12-12.4.127
nvidia-cudnn-cu12-9.1.0.70
nvidia-cufft-cu12-11.2.1.3
nvidia-curand-cu12-10.3.5.147
nvidia-cusolver-cu12-11.6.1.9
nvidia-cusparse-cu12-12.3.1.170
nvidia-nvjitlink-cu12-12.4.127
addCode
addText

```

## OUTPUT:

```

/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab
(https://huggingface.co/settings/tokens), set it as secret in your Google Colab and restart your session.
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
warnings.warn(
modules.json: 100%
349/349 [00:00<00:00, 27.7kB/s]
config_sentence_transformers.json: 100% 116/116 [00:00<00:00, 8.94kB/s]
README.md: 100% 10.7k/10.7k [00:00<00:00, 954kB/s]
sentence_bert_config.json: 100% 53.0/53.0 [00:00<00:00, 3.13kB/s]
config.json: 100% 612/612 [00:00<00:00, 39.9kB/s]
model.safetensors: 100% 90.9M/90.9M [00:00<00:00, 160MB/s]
tokenizer_config.json: 100% 350/350 [00:00<00:00, 27.4kB/s]
vocab.txt: 100% 232k/232k [00:00<00:00, 3.80MB/s]
tokenizer.json: 100% 466k/466k [00:00<00:00, 6.00MB/s]
special_tokens_map.json: 100% 112/112 [00:00<00:00, 8.45kB/s]
1_Pooling%2Fconfig.json: 100% 190/190 [00:00<00:00, 11.7kB/s]

```

### ★ Generated Story:

One day, a explorer set out on a adventure. Along the way, it stumbled upon a expedition that led to an unexpected exploration. Guided by an old journey, the explorer finally reached the ultimate voyage

**6.Use a pre-trained Hugging Face model to analyze sentiment in text. Assume a real-world application, Load the sentiment analysis pipeline.Analyze the sentiment by giving sentences to input.**

```
from transformers import pipeline
```

```
# Load sentiment analysis model (DistilBERT - Trained on SST-2)
sentiment_pipeline = pipeline("sentiment-analysis")
```

**# Function to analyze sentiment**

```
def analyze_sentiment(text):
    result = sentiment_pipeline(text)[0] # Get the first result
    label = result["label"]
    confidence = result["score"]
    return f"Sentiment: {label} (Confidence: {confidence:.2f})"
```

**# Example inputs**

```
texts = [
    "I love this product! It's amazing.",
    "This is the worst experience I've ever had.",
    "The movie was okay, but nothing special.",
    "I'm extremely happy with my new laptop!",
    "This service is so frustrating and disappointing."
]
```

**# Analyzing sentiment for each text for text in texts:**

```
print(f"Text: {text}")
print(analyze_sentiment(text))
print("-" * 50)
```

## OUTPUT:

No model was supplied, defaulted to distilbert/distilbert-base-uncased-finetuned-sst-2-english and revision 714eb0f (<https://huggingface.co/distilbert/distilbert-base-uncased-finetuned-sst-2-english>). Using a pipeline without specifying a model name and revision in production is not recommended.

```
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning: The secret `HF_TOKEN` does not exist in your Colab secrets.
```

To authenticate with the Hugging Face Hub, create a token in your settings tab (<https://huggingface.co/settings/tokens>), set it as secret in your Google Colab and restart your session.

You will be able to reuse this secret in all of your notebooks.

Please note that authentication is recommended but still optional to access public models or datasets.

```
warnings.warn(
config.json: 100% 629/629 [00:00<00:00, 28.7kB/s]
model.safetensors: 100% 268M/268M [00:05<00:00, 15.1MB/s]
tokenizer_config.json: 100% 48.0/48.0 [00:00<00:00, 2.40kB/s]
vocab.txt: 100% 232k/232k [00:00<00:00, 3.95MB/s]
```

Device set to use cpu

Text: I love this product! It's amazing.

Sentiment: POSITIVE (Confidence: 1.00)

Text: This is the worst experience I've ever had.  
Sentiment: NEGATIVE (Confidence: 1.00)

Text: The movie was okay, but nothing special.  
Sentiment: NEGATIVE (Confidence: 0.99)

Text: I'm extremely happy with my new laptop!  
Sentiment: POSITIVE (Confidence: 1.00)

Text: This service is so frustrating and disappointing.  
Sentiment: NEGATIVE (Confidence: 1.00)

**7.Summarize long texts using a pre-trained summarization model using Hugging face model. Load the summarization pipeline. Take a passage as input and obtain the summarized text.**

```
from transformers import pipeline
```

```
# Load the summarization model (BART - Fine-tuned on CNN/DailyMail)
summarizer = pipeline("summarization", model="facebook/bart-large-cnn")
```

```
# Function to summarize text
```

```
def summarize_text(text, max_length=130, min_length=50):
    summary = summarizer(text, max_length=max_length, min_length=min_length, do_sample=False)
    return summary[0]["summary_text"]
```

```
# Example Long Text
```

```
long_text = """
```

Artificial Intelligence (AI) is a rapidly advancing field that aims to create machines capable of human-like thinking. AI is used in various industries, from healthcare to finance, improving efficiency and accuracy. Machine learning, a subset of AI, enables computers to learn from data and make predictions without being explicitly programmed. With deep learning, neural networks can process vast amounts of information and recognize patterns, leading to advancements in self-driving cars, natural language processing, and medical diagnostics.""""

```
# Summarizing the text
```

```
summary = summarize_text(long_text)
```

```
# Print results
```

```
print("Original Text:")
print(long_text)
print("\nSummarized Text:")
print(summary)
```

**OUTPUT:**

```
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:  
The secret `HF_TOKEN` does not exist in your Colab secrets.  
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as secret in your Google Colab and restart your session.  
You will be able to reuse this secret in all of your notebooks.  
Please note that authentication is recommended but still optional to access public models or datasets.  
    warnings.warn(  
config.json: 100%  
  1.58k/1.58k [00:00<00:00, 103kB/s] model.safetensors: 100%  
  1.63G/1.63G [00:12<00:00, 189MB/ generation_config.json: 100%  
  363/363 [00:00<00:00, 19.2kB/s]  
vocab.json: 100% 899k/899k [00:00<00:00, 5.01MB/s]  
merges.txt: 100% 456k/456k [00:00<00:00, 3.19MB/s]  
tokenizer.json: 100% 1.36M/1.36M [00:00<00:00, 11.9MB/s]  
Device set to use cpu. Your max_length is set to 130, but your input_length is only 109. Since this is a summarization task, where outputs shorter than the input are typically wanted, you might consider decreasing max_length manually, e.g. summarizer(..., max_length=54)
```

**Original Text:**

Artificial Intelligence (AI) is a rapidly advancing field that aims to create machines capable of human-like thinking. AI is used in various industries, from healthcare to finance, improving efficiency and accuracy. Machine learning, a subset of AI, enables computers to learn from data and make predictions without being explicitly programmed. With deep learning, neural networks can process vast amounts of information and recognize patterns, leading to advancements in self-driving cars, natural language processing, and medical diagnostics.

**Summarized Text:**

Artificial Intelligence (AI) is a rapidly advancing field that aims to create machines capable of human-like thinking. Machine learning, a subset of AI, enables computers to learn from data and make predictions without being explicitly programmed. With deep learning, neural networks can process vast amounts of information and recognize patterns.

**8. Install langchain, cohore (for key), langchain-community. Get the api key( By logging into Cohere and obtaining the cohore key). Load a text document from your google drive . Create a prompt template to display the output in a particular manner.**

**Install langchain, cohore (for key), langchain-community.**

- Get the API key (by logging into Cohere and obtaining the Cohere key).
- Load a text document from Google Drive.
- Create a prompt template to display the output in a particular manner.

**!pip install google-api-python-client==2.100.0**

```
Requirement already satisfied: google-api-python-client==2.100.0 in /usr/local/lib/python3.11/dist-packages (2.100.0)
Requirement already satisfied: httplib2<1.dev0,>=0.15.0 in /usr/local/lib/python3.11/dist-packages (from google-api-python-client==2.100.0) (0.22.0)
Requirement already satisfied: google-auth<3.0.0.dev0,>=1.19.0 in /usr/local/lib/python3.11/dist-packages (from google-api-python-client==2.100.0) (2.27.0)
Requirement already satisfied: google-auth-httplib2>=0.1.0 in /usr/local/lib/python3.11/dist-packages (from google-api-python-client==2.100.0) (0.2.0)
Requirement already satisfied: google-api-core!=2.0.*,!<2.1.*,!<2.2.*,!<2.3.0,<3.0.0.dev0,>=1.31.5 in
/usr/local/lib/python3.11/dist-packages (from google-api-python-client==2.100.0) (2.19.2)
Requirement already satisfied: uritemplate<5,>=3.0.1 in /usr/local/lib/python3.11/dist-packages (from google-api-python-client==2.100.0) (4.1.1)
Requirement already satisfied: googleapis-common-protos<2.0.dev0,>=1.56.2 in /usr/local/lib/python3.11/dist-packages
(from google-api-core!=2.0.*,!<2.1.*,!<2.2.*,!<2.3.0,<3.0.0.dev0,>=1.31.5->google-api-python-client==2.100.0) (1.67.0)
Requirement already satisfied:
protobuf!=3.20.0,!<3.20.1,!<4.21.0,!<4.21.1,!<4.21.2,!<4.21.3,!<4.21.4,!<4.21.5,<6.0.0.dev0,>=3.19.5 in
/usr/local/lib/python3.11/dist-packages (from google-api-core!=2.0.*,!<2.1.*,!<2.2.*,!<2.3.0,<3.0.0.dev0,>=1.31.5-
>google-api-python-client==2.100.0) (4.25.6)
Requirement already satisfied: proto-plus<2.0.0dev,>=1.22.3 in /usr/local/lib/python3.11/dist-packages (from google-api-
core!=2.0.*,!<2.1.*,!<2.2.*,!<2.3.0,<3.0.0.dev0,>=1.31.5->google-api-python-client==2.100.0) (1.26.0)
Requirement already satisfied: requests<3.0.0.dev0,>=2.18.0 in /usr/local/lib/python3.11/dist-packages (from google-api-
core!=2.0.*,!<2.1.*,!<2.2.*,!<2.3.0,<3.0.0.dev0,>=1.31.5->google-api-python-client==2.100.0) (2.32.3)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from google-
auth<3.0.0.dev0,>=1.19.0->google-api-python-client==2.100.0) (5.5.1)
Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.11/dist-packages (from google-
auth<3.0.0.dev0,>=1.19.0->google-api-python-client==2.100.0) (0.4.1)
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.11/dist-packages (from google-
auth<3.0.0.dev0,>=1.19.0->google-api-python-client==2.100.0) (4.9)
Requirement already satisfied: pyparsing!=3.0.0,!<3.0.1,!<3.0.2,!<3.0.3,<4,>=2.4.2 in /usr/local/lib/python3.11/dist-
packages (from httplib2<1.dev0,>=0.15.0->google-api-python-client==2.100.0) (3.2.1)
Requirement already satisfied: pyasn1<0.7.0,>=0.4.6 in /usr/local/lib/python3.11/dist-packages (from pyasn1-
modules>=0.2.1->google-auth<3.0.0.dev0,>=1.19.0->google-api-python-client==2.100.0) (0.6.1)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from
requests<3.0.0.dev0,>=2.18.0->google-api-core!=2.0.*,!<2.1.*,!<2.2.*,!<2.3.0,<3.0.0.dev0,>=1.31.5->google-api-python-
client==2.100.0) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from
requests<3.0.0.dev0,>=2.18.0->google-api-core!=2.0.*,!<2.1.*,!<2.2.*,!<2.3.0,<3.0.0.dev0,>=1.31.5->google-api-python-
client==2.100.0) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from
requests<3.0.0.dev0,>=2.18.0->google-api-core!=2.0.*,!<2.1.*,!<2.2.*,!<2.3.0,<3.0.0.dev0,>=1.31.5->google-api-python-
```

```
client==2.100.0) (2.3.0)
```

```
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0.dev0,>=2.18.0->google-api-core!=2.0.*,!>2.1.*,!>2.2.*,!>2.3.0,<3.0.0.dev0,>=1.31.5->google-api-python-client==2.100.0) (2025.1.31)
```

**!pip install langchain cohere langchain-community pydantic google-auth google-auth-oauthlib google-auth-httplib2**

```
Requirement already satisfied: langchain in /usr/local/lib/python3.11/dist-packages (0.3.19)
Requirement already satisfied: cohere in /usr/local/lib/python3.11/dist-packages (5.13.12)
Requirement already satisfied: langchain-community in /usr/local/lib/python3.11/dist-packages (0.3.18)
Requirement already satisfied: pydantic in /usr/local/lib/python3.11/dist-packages (2.10.6)
Requirement already satisfied: google-auth in /usr/local/lib/python3.11/dist-packages (2.27.0)
Requirement already satisfied: google-auth-oauthlib in /usr/local/lib/python3.11/dist-packages (1.2.1)
Requirement already satisfied: google-auth-httplib2 in /usr/local/lib/python3.11/dist-packages (0.2.0)
Requirement already satisfied: langchain-core<1.0.0,>=0.3.35 in /usr/local/lib/python3.11/dist-packages (from langchain) (0.3.37)
Requirement already satisfied: langchain-text-splitters<1.0.0,>=0.3.6 in /usr/local/lib/python3.11/dist-packages (from langchain) (0.3.6)
Requirement already satisfied: langsmith<0.4,>=0.1.17 in /usr/local/lib/python3.11/dist-packages (from langchain) (0.3.8)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in /usr/local/lib/python3.11/dist-packages (from langchain) (2.0.38)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.11/dist-packages (from langchain) (2.32.3)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.11/dist-packages (from langchain) (6.0.2)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /usr/local/lib/python3.11/dist-packages (from langchain) (3.11.12)
Requirement already satisfied: tenacity!=8.4.0,<10,>=8.1.0 in /usr/local/lib/python3.11/dist-packages (from langchain) (9.0.0)
Requirement already satisfied: numpy<2,>=1.26.4 in /usr/local/lib/python3.11/dist-packages (from langchain) (1.26.4)
Requirement already satisfied: fastavro<2.0.0,>=1.9.4 in /usr/local/lib/python3.11/dist-packages (from cohore) (1.10.0)
Requirement already satisfied: httpx>=0.21.2 in /usr/local/lib/python3.11/dist-packages (from cohore) (0.28.1)
Requirement already satisfied: httpx-sse==0.4.0 in /usr/local/lib/python3.11/dist-packages (from cohore) (0.4.0)
Requirement already satisfied: pydantic-core<3.0.0,>=2.18.2 in /usr/local/lib/python3.11/dist-packages (from cohore) (2.27.2)
Requirement already satisfied: tokenizers<1,>=0.15 in /usr/local/lib/python3.11/dist-packages (from cohore) (0.21.0)
Requirement already satisfied: types-requests<3.0.0,>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from cohore) (2.32.0.20241016)
Requirement already satisfied: typing_extensions>=4.0.0 in /usr/local/lib/python3.11/dist-packages (from cohore) (4.12.2)
Requirement already satisfied: dataclasses-json<0.7,>=0.5.7 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (0.6.7)
Requirement already satisfied: pydantic-settings<3.0.0,>=2.4.0 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (2.7.1)
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.11/dist-packages (from pydantic) (0.7.0)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from google-auth) (5.5.1)
Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.11/dist-packages (from google-auth) (0.4.1)
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.11/dist-packages (from google-auth) (4.9)
Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.11/dist-packages (from google-auth-oauthlib) (2.0.0)
Requirement already satisfied: httplib2>=0.19.0 in /usr/local/lib/python3.11/dist-packages (from google-auth-httplib2) (0.22.0)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (2.4.6)
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3-
```

>langchain) (1.3.2)  
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (25.1.0)  
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (1.5.0)  
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (6.1.0)  
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (0.2.1)  
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (1.18.3)  
Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in /usr/local/lib/python3.11/dist-packages (from dataclasses-json<0.7,>=0.5.7->langchain-community) (3.26.1)  
Requirement already satisfied: typing-inspect<1,>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from dataclasses-json<0.7,>=0.5.7->langchain-community) (0.9.0)  
Requirement already satisfied: pyparsing!=3.0.0,!>=3.0.1,!>=3.0.2,!>=3.0.3,<4,>=2.4.2 in /usr/local/lib/python3.11/dist-packages (from httplib2>=0.19.0->google-auth-httplib2) (3.2.1)  
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere) (3.7.1)  
Requirement already satisfied: certifi in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere) (2025.1.31)  
Requirement already satisfied: httpcore==1.\* in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere) (1.0.7)  
Requirement already satisfied: idna in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere) (3.10)  
Requirement already satisfied: h11<0.15,>=0.13 in /usr/local/lib/python3.11/dist-packages (from httpcore==1.\*->httpx>=0.21.2->cohere) (0.14.0)  
Requirement already satisfied: jsonpatch<2.0,>=1.33 in /usr/local/lib/python3.11/dist-packages (from langchain-core<1.0.0,>=0.3.35->langchain) (1.33)  
Requirement already satisfied: packaging<25,>=23.2 in /usr/local/lib/python3.11/dist-packages (from langchain-core<1.0.0,>=0.3.35->langchain) (24.2)  
Requirement already satisfied: orjson<4.0.0,>=3.9.14 in /usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain) (3.10.15)  
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in /usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain) (1.0.0)  
Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in /usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain) (0.23.0)  
Requirement already satisfied: pyasn1<0.7.0,>=0.4.6 in /usr/local/lib/python3.11/dist-packages (from pyasn1-modules>=0.2.1->google-auth) (0.6.1)  
Requirement already satisfied: python-dotenv>=0.21.0 in /usr/local/lib/python3.11/dist-packages (from pydantic-settings<3.0.0,>=2.4.0->langchain-community) (1.0.1)  
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (3.4.1)  
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (2.3.0)  
Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.11/dist-packages (from requests-oauthlib>=0.7.0->google-auth-oauthlib) (3.2.2)  
Requirement already satisfied: greenlet!=0.4.17 in /usr/local/lib/python3.11/dist-packages (from SQLAlchemy<3,>=1.4->langchain) (3.1.1)  
Requirement already satisfied: huggingface-hub<1.0,>=0.16.4 in /usr/local/lib/python3.11/dist-packages (from tokenizers<1,>=0.15->cohere) (0.28.1)  
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers<1,>=0.15->cohere) (3.17.0)  
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers<1,>=0.15->cohere) (2024.10.0)  
Requirement already satisfied: tqdm>=4.42.1 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers<1,>=0.15->cohere) (4.67.1)  
Requirement already satisfied: jsonpointer>=1.9 in /usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33-

```
>langchain-core<1.0.0,>=0.3.35->langchain) (3.0.0)
```

Requirement already satisfied: mypy-extensions>=0.3.0 in /usr/local/lib/python3.11/dist-packages (from typing-inspect<1,>=0.4.0->dataclasses-json<0.7,>=0.5.7->langchain-community) (1.0.0)

Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-packages (from anyio->httpx>=0.21.2->cohere) (1.3.1)

```
import os
os.environ["COHERE_API_KEY"] = "YOUR_API_KEY"
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
import cohere

co = cohere.Client(os.getenv("COHERE_API_KEY")) # Initialize Cohere client

response = co.generate(
    model="command-xlarge-nightly",
    prompt=formatted_prompt,
    max_tokens=500
)

print("Generated Response:", response.generations[0].text)
```

#### OUTPUT:

**Generated Response:** \*\*Keywords:\*\* Drones, Terrorism, NYPD, Counterterrorism, Explosives, Technology, Air Assault, Chemical Weapons, Military, Police, Germany, Angela Merkel

**Summary:** The New York Police Department (NYPD) is concerned about the potential use of drones as terrorist weapons, particularly for carrying out air assaults with chemical weapons and explosives. They are working on developing technology to detect and take control of drones, and are consulting with the military and various units to devise a plan to counter weaponized drones. This concern follows an incident in Germany where a drone landed in front of Chancellor Angela Merkel, highlighting the potential threat.

**Topics:**

- Drone Technology and Terrorism: The potential use of drones as a tool for terrorist attacks, including the ability to carry explosives, firearms, and chemical weapons.
- NYPD's Response: The New York Police Department's efforts to address the threat, including technological advancements, collaboration with other agencies, and the investigation of past incidents.
- International Incidents: Discussing the drone incident involving German Chancellor Angela Merkel, highlighting the global nature of the issue.

**Entities:**

- **People:** Salvatore DiPace, Angela Merkel
- **Organizations:** New York Police Department (NYPD), CBS News
- **Locations:** New York, Germany, Manchester (England)
- **Dates:** None specified in the text, but the document refers to recent events, with an incident in Germany mentioned from the previous year.

**9.Take the Institution name as input. Use Pydantic to define the schema for the desired output and create a custom output parser. Invoke the Chain and Fetch Results. Extract the below Institution related details from Wikipedia: The founder of the Institution. When it was founded. The current branches in the institution. How many employees are working in it. A brief 4-line summary of the institution.**

```
!pip install langchain langchain-community cohere wikipedia-api pydantic
```

```
Requirement already satisfied: langchain in /usr/local/lib/python3.11/dist-packages (0.3.18)
Collecting langchain-community
  Downloading langchain_community-0.3.18-py3-none-any.whl.metadata (2.4 kB)
Requirement already satisfied: cohere in /usr/local/lib/python3.11/dist-packages (5.13.12)
Requirement already satisfied: wikipedia-api in /usr/local/lib/python3.11/dist-packages (0.8.1)
Requirement already satisfied: pydantic in /usr/local/lib/python3.11/dist-packages (2.10.6)
Requirement already satisfied: langchain-core<1.0.0,>=0.3.34 in /usr/local/lib/python3.11/dist-packages (from langchain) (0.3.35)
Requirement already satisfied: langchain-text-splitters<1.0.0,>=0.3.6 in /usr/local/lib/python3.11/dist-packages (from langchain) (0.3.6)
Requirement already satisfied: langsmith<0.4,>=0.1.17 in /usr/local/lib/python3.11/dist-packages (from langchain) (0.3.8)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in /usr/local/lib/python3.11/dist-packages (from langchain) (2.0.38)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.11/dist-packages (from langchain) (2.32.3)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.11/dist-packages (from langchain) (6.0.2)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /usr/local/lib/python3.11/dist-packages (from langchain) (3.11.12)
Requirement already satisfied: tenacity!=8.4.0,<10,>=8.1.0 in /usr/local/lib/python3.11/dist-packages (from langchain) (9.0.0)
Requirement already satisfied: numpy<2,>=1.26.4 in /usr/local/lib/python3.11/dist-packages (from langchain) (1.26.4)
Collecting langchain-core<1.0.0,>=0.3.34 (from langchain)
  Downloading langchain_core-0.3.37-py3-none-any.whl.metadata (5.9 kB)
Collecting langchain
  Downloading langchain-0.3.19-py3-none-any.whl.metadata (7.9 kB)
Collecting dataclasses-json<0.7,>=0.5.7 (from langchain-community)
  Downloading dataclasses_json-0.6.7-py3-none-any.whl.metadata (25 kB)
Collecting pydantic-settings<3.0.0,>=2.4.0 (from langchain-community)
  Downloading pydantic_settings-2.7.1-py3-none-any.whl.metadata (3.5 kB)
Requirement already satisfied: htxpx-sse<1.0.0,>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (0.4.0)
Requirement already satisfied: fastavro<2.0.0,>=1.9.4 in /usr/local/lib/python3.11/dist-packages (from cohere) (1.10.0)
Requirement already satisfied: htxpx>=0.21.2 in /usr/local/lib/python3.11/dist-packages (from cohere) (0.28.1)
Requirement already satisfied: pydantic-core<3.0.0,>=2.18.2 in /usr/local/lib/python3.11/dist-packages (from cohere) (2.27.2)
Requirement already satisfied: tokenizers<1,>=0.15 in /usr/local/lib/python3.11/dist-packages (from cohere) (0.21.0)
Requirement already satisfied: types-requests<3.0.0,>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from cohere) (2.32.0.20241016)
Requirement already satisfied: typing_extensions>=4.0.0 in /usr/local/lib/python3.11/dist-packages (from cohere) (4.12.2)
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.11/dist-packages (from pydantic) (0.7.0)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (2.4.6)
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (25.1.0)
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (1.5.0)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (6.1.0)
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (0.2.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain) (1.18.3)
Collecting marshmallow<4.0.0,>=3.18.0 (from dataclasses-json<0.7,>=0.5.7->langchain-community)
  Downloading marshmallow-3.26.1-py3-none-any.whl.metadata (7.3 kB)
Collecting typing-inspect<1,>=0.4.0 (from dataclasses-json<0.7,>=0.5.7->langchain-community)
  Downloading typing_inspect-0.9.0-py3-none-any.whl.metadata (1.5 kB)
```

Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere) (3.7.1)  
 Requirement already satisfied: certifi in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere) (2025.1.31)  
 Requirement already satisfied: httpcore==1.\* in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere) (1.0.7)  
 Requirement already satisfied: idna in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere) (3.10)  
 Requirement already satisfied: h11<0.15,>=0.13 in /usr/local/lib/python3.11/dist-packages (from httpcore==1.\*->httpx>=0.21.2->cohere) (0.14.0)  
 Requirement already satisfied: jsonpatch<2.0,>=1.33 in /usr/local/lib/python3.11/dist-packages (from langchain-core<1.0.0,>=0.3.34->langchain) (1.33)  
 Requirement already satisfied: packaging<25,>=23.2 in /usr/local/lib/python3.11/dist-packages (from langchain-core<1.0.0,>=0.3.34->langchain) (24.2)  
 Requirement already satisfied: orjson<4.0.0,>=3.9.14 in /usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain) (3.10.15)  
 Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in /usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain) (1.0.0)  
 Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in /usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain) (0.23.0)  
 Collecting python-dotenv>=0.21.0 (from pydantic-settings<3.0.0,>=2.4.0->langchain-community)  
   Downloading python\_dotenv-1.0.1-py3-none-any.whl.metadata (23 kB)  
 Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (3.4.1)  
 Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (2.3.0)  
 Requirement already satisfied: greenlet!=0.4.17 in /usr/local/lib/python3.11/dist-packages (from SQLAlchemy<3,>=1.4->langchain) (3.1.1)  
 Requirement already satisfied: huggingface-hub<1.0,>=0.16.4 in /usr/local/lib/python3.11/dist-packages (from tokenizers<1,>=0.15->cohere) (0.28.1)  
 Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers<1,>=0.15->cohere) (3.17.0)  
 Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers<1,>=0.15->cohere) (2024.10.0)  
 Requirement already satisfied: tqdm>=4.42.1 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers<1,>=0.15->cohere) (4.67.1)  
 Requirement already satisfied: jsonpointer>=1.9 in /usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33->langchain-core<1.0.0,>=0.3.34->langchain) (3.0.0)  
 Collecting mypy-extensions>=0.3.0 (from typing-inspect<1,>=0.4.0->dataclasses-json<0.7,>=0.5.7->langchain-community)  
   Downloading mypy\_extensions-1.0.0-py3-none-any.whl.metadata (1.1 kB)  
 Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-packages (from anyio->httpx>=0.21.2->cohere) (1.3.1)  
 Downloading langchain\_community-0.3.18-py3-none-any.whl (2.5 MB) 2.5/2.5 MB 30.2  
 MB/s eta 0:00:00  
 Downloading langchain-0.3.19-py3-none-any.whl (1.0 MB) 1.0/1.0 MB 46.6  
 MB/s eta 0:00:00  
 Downloading dataclasses\_json-0.6.7-py3-none-any.whl (28 kB)  
 Downloading langchain\_core-0.3.37-py3-none-any.whl (413 kB) 413.7/413.7 kB 25.9  
 MB/s eta 0:00:00  
 Downloading pydantic\_settings-2.7.1-py3-none-any.whl (29 kB)  
 Downloading marshmallow-3.26.1-py3-none-any.whl (50 kB) 50.9/50.9 kB 3.4  
 MB/s eta 0:00:00  
 Downloading python\_dotenv-1.0.1-py3-none-any.whl (19 kB)  
 Downloading typing\_inspect-0.9.0-py3-none-any.whl (8.8 kB)  
 Downloading mypy\_extensions-1.0.0-py3-none-any.whl (4.7 kB)  
 Installing collected packages: python-dotenv, mypy-extensions, marshmallow, typing-inspect, pydantic-settings, dataclasses-json, langchain-core, langchain, langchain-community.  
 Attempting uninstall: langchain-core  
 Found existing installation: langchain-core 0.3.35  
 Uninstalling langchain-core-0.3.35:  
 Successfully uninstalled langchain-core-0.3.35  
 Attempting uninstall: langchain

```
Found existing installation: langchain 0.3.18
Uninstalling langchain-0.3.18:
Successfully uninstalled langchain-0.3.18
Successfully installed dataclasses-json-0.6.7 langchain-0.3.19 langchain-community-0.3.18 langchain-core-0.3.37
marshmallow-3.26.1 mypy-extensions-1.0.0 pydantic-settings-2.7.1 python-dotenv-1.0.1 typing-inspect-0.9.0
```

```
import os
from langchain.llms import Cohere
from langchain.prompts import PromptTemplate
from langchain.chains import LLMChain
import wikipediaapi
from pydantic import BaseModel, Field

# Set Cohere API Key
os.environ["COHERE_API_KEY"] = "YOUR_API_KEY"

# Initialize Cohere Model
llm = Cohere(model="command", cohere_api_key=os.getenv("COHERE_API_KEY"))
```

```
<ipython-input-5-b76570ebff51>:12: LangChainDeprecationWarning: The class `Cohere` was deprecated in LangChain
0.1.14 and will be removed in 1.0. An updated version of the class exists in the :class:`~langchain-cohere` package and
should be used instead. To use it run `pip install -U :class:`~langchain-cohere` and import as `from
:class:`~langchain_coherer import Cohere``.
llm = Cohere(model="command", cohere_api_key=os.getenv("COHERE_API_KEY"))
```

```
def extract_info_with_cohere(institution_name):
    """Uses Cohere LLM to extract structured details from Wikipedia text."""
    summary = fetch_wikipedia_summary(institution_name)

    if summary is None:
        return None

    # Define Prompt for LangChain + Cohere
    prompt_template = PromptTemplate(
        input_variables=["institution_name", "text"],
        template="""
Extract the following details about {institution_name} from the given text:
1. Founder of the institution
2. Year it was founded
3. Current branches (if available)
4. Number of employees (if available)
5. A 4-line summary about the institution
    """

    Text: {text}
```

Return the result in this structured format:

```
Founder: <name>
Founded Year: <year>
Branches: <branches>
Employees: <count>
Summary: <4-line summary>
"""
)
```

## # Create LangChain Chain

```
chain = LLMChain(llm=llm, prompt=prompt_template)
```

### # Invoke LLM

```
response = chain.run(institution_name=institution_name, text=summary)
```

### # Parse Output

```
lines = response.split("\n")
founder = lines[0].split(":")[1].strip()
founded_year = lines[1].split(":")[1].strip()
branches = lines[2].split(":")[1].strip()
employees = lines[3].split(":")[1].strip()
summary_text = lines[4].split(":")[1].strip()
```

```
return InstitutionInfo(
    name=institution_name,
    founder=founder,
    founded_year=founded_year,
    branches=branches,
    employees=employees,
    summary=summary_text
)
```

```
institution_name = "Stanford University"
# institution_name = input("Enter Institution Name: ")
info = extract_info_with_cohere(institution_name)

print("Final Output ")

if info:
    print(info.model_dump_json(indent=4)) # ✅ NEW (Pydantic v2)

else:
    print("No information found on Wikipedia.")
```

## OUTPUT:

```
{
  "name": "Stanford University",
  "founder": "Leland Stanford",
  "founded_year": "1885",
  "branches": "None mentioned",
  "employees": "2,288",
  "summary": "Leland Stanford Junior University is a private research university founded in Stanford, California in 1885 by Leland Stanford, then-incumbent senator from California, and his wife Jane. The couple founded the university in memory of their only child, Leland Jr. The university opened as a coeducational and non-denominational institution in 1891 and is currently one of the largest in the nation with an 8,180-acre campus."
}
```

**10. Build a chatbot for the Indian Penal Code. We'll start by downloading the official Indian Penal Code document, and then we'll create a chatbot that can interact with it. Users will be able to ask questions about the Indian Penal Code and have a conversation with it.**

```
# Install required libraries
!pip install langchain langchain_community cohere faiss-cpu pypdf
```

```
import langchain, cohere
print("LangChain Version:", langchain.__version__)
print("Cohere Version:", cohere.__version__)
```

```
Requirement already satisfied: langchain in /usr/local/lib/python3.11/dist-packages (0.3.23)
Requirement already satisfied: langchain_community in /usr/local/lib/python3.11/dist-packages (0.3.21)
Requirement already satisfied: cohere in /usr/local/lib/python3.11/dist-packages (5.14.2)
Requirement already satisfied: faiss-cpu in /usr/local/lib/python3.11/dist-packages (1.10.0)
Requirement already satisfied: pypdf in /usr/local/lib/python3.11/dist-packages (5.4.0)
Requirement already satisfied: langchain-core<1.0.0,>=0.3.51 in /usr/local/lib/python3.11/dist-packages (from langchain) (0.3.51)
Requirement already satisfied: langchain-text-splitters<1.0.0,>=0.3.8 in /usr/local/lib/python3.11/dist-packages (from langchain) (0.3.8)
Requirement already satisfied: langsmith<0.4,>=0.1.17 in /usr/local/lib/python3.11/dist-packages (from langchain) (0.3.22)
Requirement already satisfied: pydantic<3.0.0,>=2.7.4 in /usr/local/lib/python3.11/dist-packages (from langchain) (2.11.1)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in /usr/local/lib/python3.11/dist-packages (from langchain) (2.0.40)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.11/dist-packages (from langchain) (2.32.3)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.11/dist-packages (from langchain) (6.0.2)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /usr/local/lib/python3.11/dist-packages (from langchain_community) (3.11.15)
Requirement already satisfied: tenacity!=8.4.0,<10,>=8.1.0 in /usr/local/lib/python3.11/dist-packages (from langchain_community) (9.1.2)
Requirement already satisfied: dataclasses-json<0.7,>=0.5.7 in /usr/local/lib/python3.11/dist-packages (from langchain_community) (0.6.7)
Requirement already satisfied: pydantic-settings<3.0.0,>=2.4.0 in /usr/local/lib/python3.11/dist-packages (from langchain_community) (2.8.1)
Requirement already satisfied: httpx-sse<1.0.0,>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from langchain_community) (0.4.0)
Requirement already satisfied: numpy<3,>=1.26.2 in /usr/local/lib/python3.11/dist-packages (from langchain_community) (2.0.2)
Requirement already satisfied: fastavro<2.0.0,>=1.9.4 in /usr/local/lib/python3.11/dist-packages (from cohore) (1.10.0)
Requirement already satisfied: httpx>=0.21.2 in /usr/local/lib/python3.11/dist-packages (from cohore) (0.28.1)
Requirement already satisfied: pydantic-core<3.0.0,>=2.18.2 in /usr/local/lib/python3.11/dist-packages (from cohore) (2.33.0)
Requirement already satisfied: tokenizers<1,>=0.15 in /usr/local/lib/python3.11/dist-packages (from cohore) (0.21.1)
Requirement already satisfied: types-requests<3.0.0,>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from cohore) (2.32.0.20250328)
Requirement already satisfied: typing_extensions>=4.0.0 in /usr/local/lib/python3.11/dist-packages (from cohore) (4.13.0)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-packages (from faiss-cpu) (24.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain_community) (2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain_community) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain_community) (25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain_community) (1.5.0)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain_community) (6.3.1)
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain_community) (0.3.1)
Requirement already satisfied: yaml<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain_community) (1.18.3)
Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in /usr/local/lib/python3.11/dist-packages (from dataclasses-json<0.7,>=0.5.7->langchain_community) (3.26.1)
Requirement already satisfied: typing-inspect<1,>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from dataclasses-json<0.7,>=0.5.7->langchain_community) (0.9.0)
```

Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere) (4.9.0)  
 Requirement already satisfied: certifi in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere) (2025.1.31)  
 Requirement already satisfied: httpcore==1.\* in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere) (1.0.7)  
 Requirement already satisfied: idna in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere) (3.10)  
 Requirement already satisfied: h11<0.15,>=0.13 in /usr/local/lib/python3.11/dist-packages (from httpcore==1.\*->httpx>=0.21.2->cohere) (0.14.0)  
 Requirement already satisfied: jsonpatch<2.0,>=1.33 in /usr/local/lib/python3.11/dist-packages (from langchain-core<1.0.0,>=0.3.51->langchain) (1.33)  
 Requirement already satisfied: orjson<4.0.0,>=3.9.14 in /usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain) (3.10.16)  
 Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in /usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain) (1.0.0)  
 Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in /usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain) (0.23.0)  
 Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4->langchain) (0.7.0)  
 Requirement already satisfied: typing-inspection>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4->langchain) (0.4.0)  
 Requirement already satisfied: python-dotenv>=0.21.0 in /usr/local/lib/python3.11/dist-packages (from pydantic-settings<3.0.0,>=2.4.0->langchain\_community) (1.1.0)  
 Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (3.4.1)  
 Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (2.3.0)  
 Requirement already satisfied: greenlet>=1 in /usr/local/lib/python3.11/dist-packages (from SQLAlchemy<3,>=1.4->langchain) (3.1.1)  
 Requirement already satisfied: huggingface-hub<1.0,>=0.16.4 in /usr/local/lib/python3.11/dist-packages (from tokenizers<1,>=0.15->cohere) (0.30.1)  
 Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers<1,>=0.15->cohere) (3.18.0)  
 Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers<1,>=0.15->cohere) (2025.3.2)  
 Requirement already satisfied: tqdm>=4.42.1 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers<1,>=0.15->cohere) (4.67.1)  
 Requirement already satisfied: jsonpointer>=1.9 in /usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33->langchain-core<1.0.0,>=0.3.51->langchain) (3.0.0)  
 Requirement already satisfied: mypy-extensions>=0.3.0 in /usr/local/lib/python3.11/dist-packages (from typing-inspect<1,>=0.4.0->dataclasses-json<0.7,>=0.5.7->langchain\_community) (1.0.0)  
 Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-packages (from anyio->httpx>=0.21.2->cohere) (1.3.1)  
 LangChain Version: 0.3.23  
 Cohere Version: 5.14.2

```
[ ]  

!pip install langchain-cohere  

Collecting langchain-cohere  

  Downloading langchain_cohere-0.4.2-py3-none-any.whl.metadata (6.6 kB)  

Requirement already satisfied: cohere<6.0,>=5.12.0 in /usr/local/lib/python3.11/dist-packages (from langchain-cohere) (5.13.12)  

Requirement already satisfied: langchain-community<0.4.0,>=0.3.0 in /usr/local/lib/python3.11/dist-packages (from langchain-cohere) (0.3.18)  

Requirement already satisfied: langchain-core<0.4.0,>=0.3.27 in /usr/local/lib/python3.11/dist-packages (from langchain-cohere) (0.3.37)  

Requirement already satisfied: pydantic<3,>=2 in /usr/local/lib/python3.11/dist-packages (from langchain-cohere) (2.10.6)  

Collecting types-pyyaml<7.0.0.0,>=6.0.12.20240917 (from langchain-cohere)  

  Downloading types_PyYAML-6.0.12.20241230-py3-none-any.whl.metadata (1.8 kB)  

Requirement already satisfied: fastavro<2.0.0,>=1.9.4 in /usr/local/lib/python3.11/dist-packages (from cohere<6.0,>=5.12.0->langchain-cohere) (1.10.0)  

Requirement already satisfied: httpx>=0.21.2 in /usr/local/lib/python3.11/dist-packages (from cohere<6.0,>=5.12.0->langchain-cohere) (0.28.1)  

Requirement already satisfied: httpx-sse==0.4.0 in /usr/local/lib/python3.11/dist-packages (from cohere<6.0,>=5.12.0->langchain-cohere) (0.4.0)  

Requirement already satisfied: pydantic-core<3.0.0,>=2.18.2 in /usr/local/lib/python3.11/dist-packages (from cohere<6.0,>=5.12.0->langchain-cohere) (2.27.2)
```

Requirement already satisfied: requests<3.0.0,>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from cohere<6.0,>=5.12.0->langchain-cohere) (2.32.3)
Requirement already satisfied: tokenizers<1,>=0.15 in /usr/local/lib/python3.11/dist-packages (from cohere<6.0,>=5.12.0->langchain-cohere) (0.21.0)
Requirement already satisfied: types-requests<3.0.0,>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from cohere<6.0,>=5.12.0->langchain-cohere) (2.32.0.20241016)
Requirement already satisfied: typing_extensions<=4.0.0 in /usr/local/lib/python3.11/dist-packages (from cohere<6.0,>=5.12.0->langchain-cohere) (4.12.2)
Requirement already satisfied: langchain<1.0.0,>=0.3.19 in /usr/local/lib/python3.11/dist-packages (from langchain-community<0.4.0,>=0.3.0->langchain-cohere) (0.3.19)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in /usr/local/lib/python3.11/dist-packages (from langchain-community<0.4.0,>=0.3.0->langchain-cohere) (2.0.38)
Requirement already satisfied: PyYAML<=5.3 in /usr/local/lib/python3.11/dist-packages (from langchain-community<0.4.0,>=0.3.0->langchain-cohere) (6.0.2)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /usr/local/lib/python3.11/dist-packages (from langchain-community<0.4.0,>=0.3.0->langchain-cohere) (3.11.12)
Requirement already satisfied: tenacity!=8.4.0,<10,>=8.1.0 in /usr/local/lib/python3.11/dist-packages (from langchain-community<0.4.0,>=0.3.0->langchain-cohere) (9.0.0)
Requirement already satisfied: dataclasses-json<0.7,>=0.5.7 in /usr/local/lib/python3.11/dist-packages (from langchain-community<0.4.0,>=0.3.0->langchain-cohere) (0.6.7)
Requirement already satisfied: pydantic-settings<3.0.0,>=2.4.0 in /usr/local/lib/python3.11/dist-packages (from langchain-community<0.4.0,>=0.3.0->langchain-cohere) (2.8.0)
Requirement already satisfied: langsmith<0.4,>=0.1.125 in /usr/local/lib/python3.11/dist-packages (from langchain-community<0.4.0,>=0.3.0->langchain-cohere) (0.3.8)
Requirement already satisfied: numpy<2,>=1.26.4 in /usr/local/lib/python3.11/dist-packages (from langchain-community<0.4.0,>=0.3.0->langchain-cohere) (1.26.4)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in /usr/local/lib/python3.11/dist-packages (from langchain-core<0.4.0,>=0.3.27->langchain-cohere) (1.33)
Requirement already satisfied: packaging<25,>=23.2 in /usr/local/lib/python3.11/dist-packages (from langchain-core<0.4.0,>=0.3.27->langchain-cohere) (24.2)
Requirement already satisfied: annotated-types<=0.6.0 in /usr/local/lib/python3.11/dist-packages (from pydantic<3,>=2->langchain-cohere) (0.7.0)
Requirement already satisfied: aiohappyeyeballs<=2.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (2.4.6)
Requirement already satisfied: aiosignal<=1.1.2 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (1.3.2)
Requirement already satisfied: attrs<=17.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (25.1.0)
Requirement already satisfied: frozenlist<=1.1.1 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (1.5.0)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (6.1.0)
Requirement already satisfied: propcache<=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (0.2.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (1.18.3)
Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in /usr/local/lib/python3.11/dist-packages (from dataclasses-json<0.7,>=0.5.7->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (3.26.1)
Requirement already satisfied: typing-inspect<1,>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from dataclasses-json<0.7,>=0.5.7->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (0.9.0)
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere<6.0,>=5.12.0->langchain-cohere) (3.7.1)
Requirement already satisfied: certifi in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere<6.0,>=5.12.0->langchain-cohere) (2025.1.31)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere<6.0,>=5.12.0->langchain-cohere) (1.0.7)
Requirement already satisfied: idna in /usr/local/lib/python3.11/dist-packages (from httpx>=0.21.2->cohere<6.0,>=5.12.0->langchain-cohere) (3.10)
Requirement already satisfied: h11<0.15,>=0.13 in /usr/local/lib/python3.11/dist-packages (from httpcore==1.*->httpx>=0.21.2->cohere<6.0,>=5.12.0->langchain-cohere) (0.14.0)
Requirement already satisfied: jsonpointer<=1.9 in /usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33->langchain-core<0.4.0,>=0.3.27->langchain-cohere) (3.0.0)
Requirement already satisfied: langchain-text-splitters<1.0.0,>=0.3.6 in /usr/local/lib/python3.11/dist-packages (from langchain<1.0.0,>=0.3.19->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (0.3.6)
Requirement already satisfied: orjson<4.0.0,>=3.9.14 in /usr/local/lib/python3.11/dist-packages (from

```

langsmit<0.4,>=0.1.125->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (3.10.15)
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in /usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.125->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (1.0.0)
Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in /usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.125->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (0.23.0)
Requirement already satisfied: python-dotenv>=0.21.0 in /usr/local/lib/python3.11/dist-packages (from pydantic-settings<3.0.0,>=2.4.0->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (1.0.1)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.0.0->cohere<6.0,>=5.12.0->langchain-cohere) (3.4.1)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3.0.0,>=2.0.0->cohere<6.0,>=5.12.0->langchain-cohere) (2.3.0)
Requirement already satisfied: greenlet!=0.4.17 in /usr/local/lib/python3.11/dist-packages (from SQLAlchemy<3,>=1.4->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (3.1.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.16.4 in /usr/local/lib/python3.11/dist-packages (from tokenizers<1,>=0.15->cohere<6.0,>=5.12.0->langchain-cohere) (0.28.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers<1,>=0.15->cohere<6.0,>=5.12.0->langchain-cohere) (3.17.0)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers<1,>=0.15->cohere<6.0,>=5.12.0->langchain-cohere) (2024.10.0)
Requirement already satisfied: tqdm>=4.42.1 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers<1,>=0.15->cohere<6.0,>=5.12.0->langchain-cohere) (4.67.1)
Requirement already satisfied: mypy-extensions>=0.3.0 in /usr/local/lib/python3.11/dist-packages (from typing-inspect<1,>=0.4.0->dataclasses-json<0.7,>=0.5.7->langchain-community<0.4.0,>=0.3.0->langchain-cohere) (1.0.0)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-packages (from anyio->httpx>=0.21.2->cohere<6.0,>=5.12.0->langchain-cohere) (1.3.1)
Downloading langchain_cohere-0.4.2-py3-none-any.whl (42 kB)
----- 42.2/42.2 kB 2.5
MB/s eta 0:00:00
Downloading types_PyYAML-6.0.12.20241230-py3-none-any.whl (20 kB)
Installing collected packages: types-pyyaml, langchain-cohere
Successfully installed langchain-cohere-0.4.2 types-pyyaml-6.0.12.20241230

```

```

from langchain.document_loaders import PyPDFLoader
from langchain.text_splitter import RecursiveCharacterTextSplitter
from langchain.vectorstores import FAISS
from langchain.chains import RetrievalQA
from langchain.embeddings import HuggingFaceEmbeddings
from langchain.llms import HuggingFacePipeline
from transformers import AutoTokenizer, AutoModelForCausalLM, pipeline
import torch

```

#### # STEP 1: Load IPC PDF

```

pdf_path = "10th class english 2020-21 20.pdf" # Ensure this file is uploaded to Colab
loader = PyPDFLoader(pdf_path)
documents = loader.load()

```

#### # STEP 2: Split text into chunks

```

text_splitter = RecursiveCharacterTextSplitter(chunk_size=500, chunk_overlap=100)
texts = text_splitter.split_documents(documents)

```

#### # STEP 3: Use Free Hugging Face Embeddings

```

embeddings = HuggingFaceEmbeddings(model_name="sentence-transformers/all-MiniLM-L6-v2")

```

#### # STEP 4: Store embeddings in FAISS Vector Database

```

vector_store = FAISS.from_documents(texts, embeddings)

```

#### # STEP 5: Load Falcon-7B-Instruct (with Quantization to Save Memory)

```

model_id = "tiiuae/falcon-7b-instruct"

```

```

tokenizer = AutoTokenizer.from_pretrained(model_id)
model = AutoModelForCausalLM.from_pretrained(
    model_id,

```

```

torch_dtype=torch.float16, # Efficient memory use
device_map="auto" # Automatically assign GPU/CPU
)

print(" ✅ Falcon-7B Loaded Successfully!")

# ✅ STEP 6: Create Text Generation Pipeline
pipe = pipeline(
    "text-generation",
    model=model,
    tokenizer=tokenizer,
    torch_dtype=torch.float16,
    device_map="auto",
    max_new_tokens=512,
    do_sample=True,
    temperature=0.7,
)

```

# ✅ STEP 7: Create Retrieval-based QA System

```

llm = HuggingFacePipeline(pipeline=pipe)
qa_chain = RetrievalQA.from_chain_type(llm, retriever=vector_store.as_retriever())

print("Done all steps successfully")

```

**OUTPUT:**

```

Loading checkpoint shards: 100% 2/2 [01:24<00:00, 39.38s/it]
WARNING: accelerate.big_modeling:Some parameters are on the meta device because they were offloaded to
the cpu and disk.
Device set to use cuda:0
✅ Falcon-7B Loaded Successfully!
Done all steps successfully

```

## # ✅ STEP 8: Chatbot Interaction

```

def chatbot():
    print(" 🔔 IPC Chatbot is ready! (Type 'exit' to stop)")
    while True:
        query = input("You: ")
        if query.lower() == "exit":
            print("Chatbot: Goodbye! 🙏")
            break
        response = qa_chain.run(query)

        print(f"Chatbot: {response}")

```

```

# Run the chatbot
chatbot()

```

**OUTPUT:****IPC Chatbot is ready! (Type 'exit' to stop)**

Setting `pad\_token\_id` to `eos\_token\_id`:11 for open-end generation.

Chatbot: Use the following pieces of context to answer the question at the end. If you don't know the answer, just say that you don't know, don't try to make up an answer.

256 English Class XFree distribution by T.S. Government 2020-21

Appendix II

About the Textbook and Assessment Procedures

'Our World through English' Class X has been prepared in accordance with the syllabus for a ten-year course in English as a second language.

Teachers should act as facilitators and encourage children to do individual and group work, in order to make them as autonomous users of the language. Interactive approach should be followed irrespective of the

medium of instruction. Therefore, in order to have parity among students of English and non-English medium, a common English textbook has been introduced in all media from June 2012. Moreover, since IT enabled learning, multilingualism, and language across the curriculum would greatly enhance the child's learning, it would not be a problem for the non-English medium child to learn English as effectively as the child in English medium does. This single the stakeholders. The main aim of teaching English is to help learners evolve themselves as independent user of English. I hope this textbook will help teachers and students achieve this by making the teaching-learning process effective.

I thank all the institutions and experts at the state and national level, the members of the Textbook Production and Development Committees, the staff members of the SCERT, T.S.

218 English Class XFree distribution by T.S. Government 2020-21

convention (n) : the way in which most people do

blend (v) : to mix two or more substances together

dialects (n) : the forms of a language that are spoken in one area

sages (n) : wise people

seers (n) : people who claim that they can see what is going to happen in the future

piety (n) : showing a deep respect for ( esp: for god / religion)

penance (n) : voluntary self-punishment for wrongdoing

Question: what is english

Helpful Answer: English is a modern language that has evolved from Old and Middle English.

## CONTENT BEYOND THE SYLLABUS

### 1. ChatGPT / Claude / Gemini (for Conceptual Depth)

- **Type:** Advanced conversational AI
- **Use it for:**
  - Asking deeper questions (e.g., “Explain chaos theory with real-world examples”)
  - Learning interdisciplinary concepts
  - Getting personalized learning paths
- Good for students in high school, college, or self-learning
- Tools:
  - [ChatGPT](#) (GPT-4 is best for deep reasoning)
  - [Claude.ai](#) (very strong at long texts, summaries)
  - [Gemini](#) (formerly Bard – integrates with Google Search)

### 2. Elicit (AI Research Assistant)

- **Type:** Research tool powered by language models
- **Use it for:**
  - Finding academic papers and explaining them
  - Asking, “What does the research say about X?”
  - Learning how to form research questions
- Especially helpful for students doing projects or research
-  [elicit.org](#)

### 3. Perplexity.AI

- **Type:** AI-powered answer engine with citations
- **Use it for:**
  - Getting cited, up-to-date answers from the web
  - Exploring topics like “quantum computing in education” or “latest cancer therapy”
- Good for tech, science, history, social sciences
-  [perplexity.ai](#)

### 4. Canva Magic Write / Adobe Firefly (for Creative Projects)

- **Use it for:**
  - Designing infographics for complex topics
  - Generating visuals for book reports or scientific posters
  - Making interactive presentations or portfolios
- Tools:
  - [Canva](#) (Magic Write, Docs to Deck)
  - [Adobe Firefly](#) (AI art & design)

### 5. Notion AI

- **Type:** All-in-one productivity + AI workspace
- **Use it for:**
  - Summarizing textbooks
  - Creating study guides
  - Exploring “what-if” scenarios in history or science
-  notion.so

## Viva Questions

1. What is Generative AI? How is it different from traditional AI models?
2. What are some common types of generative models?
3. Explain the difference between discriminative and generative models.
4. How does a Generative Adversarial Network (GAN) work?
5. What is the latent space in generative models?
6. What are VAEs and how do they differ from GANs?
7. What is the role of noise in generative models?
8. What tools and libraries did you use for the experiments?
9. Explain the architecture of the model you implemented.
10. What dataset did you use and why?
11. How did you preprocess the data?
12. How did you evaluate the performance of your generative model?
13. Can you explain the loss functions used in your model?
14. How do you prevent mode collapse in GANs?
15. What challenges did you face during implementation?
16. What are disadvantages of Generative AI?
17. What are advantages of Generative AI?
18. What are the applications of Generative AI?
19. What is fine-tuning vs. prompt-tuning in the context of large models?
20. What is prompt engineering and why is it important for large language models?
21. How can we make generative models more controllable and interpretable?
22. What are the ethical concerns with generative AI?
23. How do you compare the outputs of different generative models?
24. How do you verify the quality of generated images/texts?
25. Can you tell if an output is AI-generated? What clues would you look for?