SCIOMETRIX

Engage Patients, Improve Outcomes.

NEURAL SEARCH

System For University

SCIOMETRIX
Engage Patients, Improve Outcomes.

Ai Project

Name	Chiranjeev Singh
Section	D2323
Reg. no.	12300735
Roll no	B36
Submitted To	Dr. Pallavi Vyas

SCIOMETRIX
Engage Patients, Improve Outcomes.

Table of Contents

SCIOMETRIX
Engage Patients, Improve Outcomes.

What is **Neural search?**

Neural search

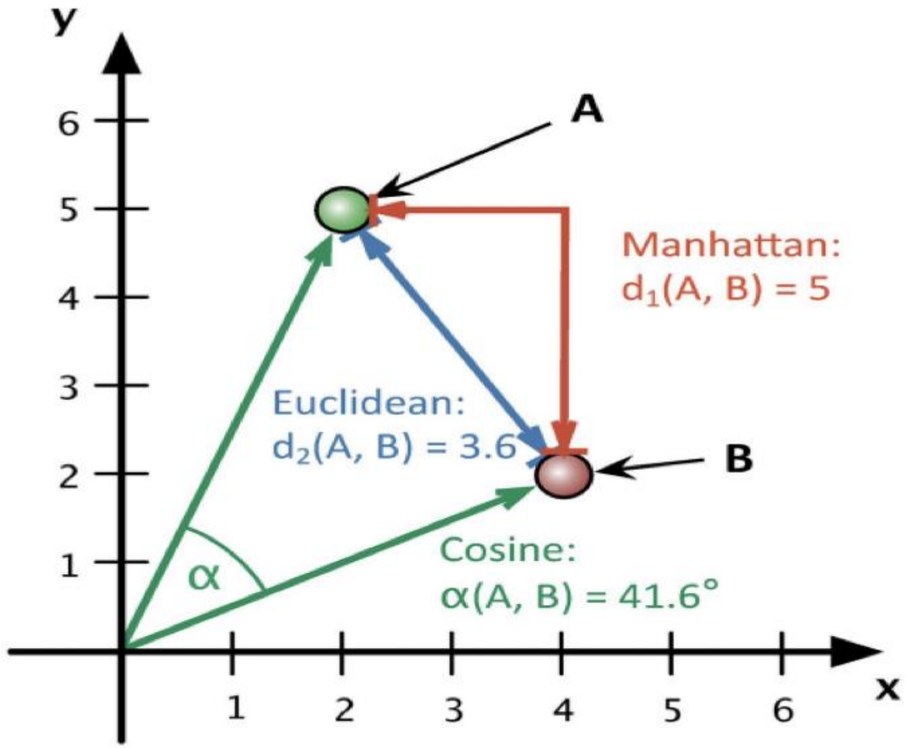
- Utilizes neural networks for search.
- Enhances accuracy by understanding context.
- Aims for personalized, relevant results.
- Leverages deep learning for data processing.

Working of Neural Search

- Neural Search works with embeddings, the section below describes embeddings
- Embeddings are generated by neural network model which will translate a text into a numerical vector
- From sentence_transformers import SentenceTransformer

SCIOMETRIX
Engage Patients, Improve Outcomes.

What is **Neural search?**



Neural search

- Utilizes neural networks for search.
- Enhances accuracy by understanding context.
- Aims for personalized, relevant results.
- Leverages deep learning for data processing.

Working of Neural Search

- Neural Search works with embeddings, the section below describes embeddings
- Embeddings are generated by neural network model which will translate a text into a numerical vector
- From sentence_transformers import SentenceTransformer

Introduction to
Neural search?

Neural search

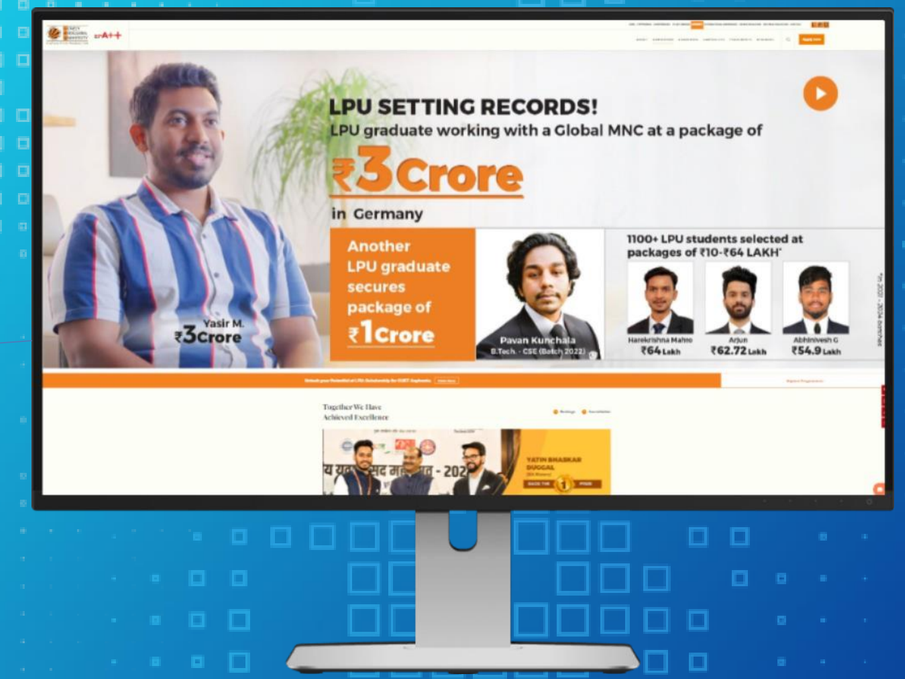
- Imagine you're looking for something specific on your university's website, like information about a particular course or event. Neural search helps you find it faster and more accurately by using smart computer programs that understand what you're looking for and where to find it.

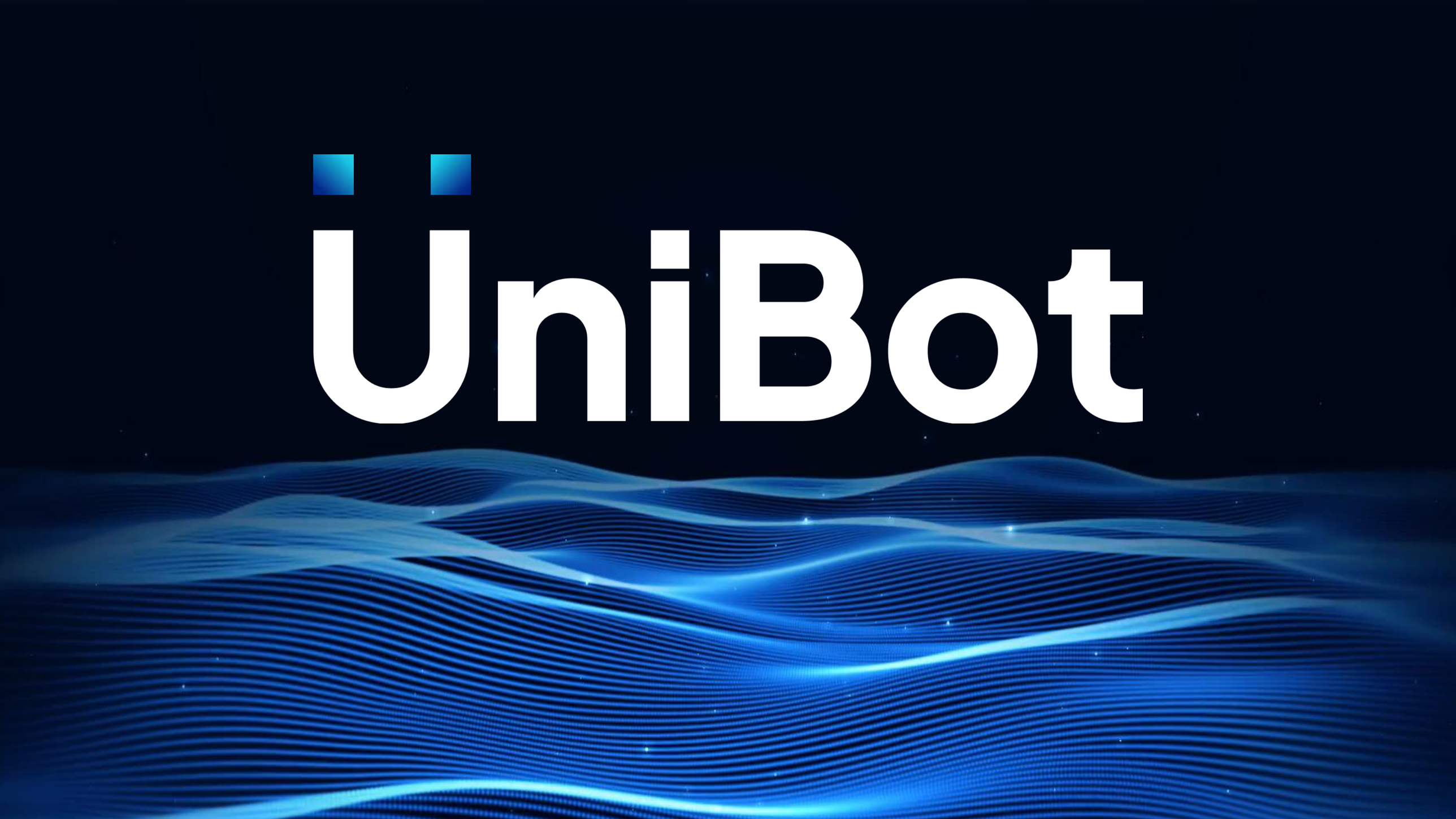
SCIOMETRIX
Engage Patients, Improve Outcomes.

Challenges in
Website Data Search

Neural search

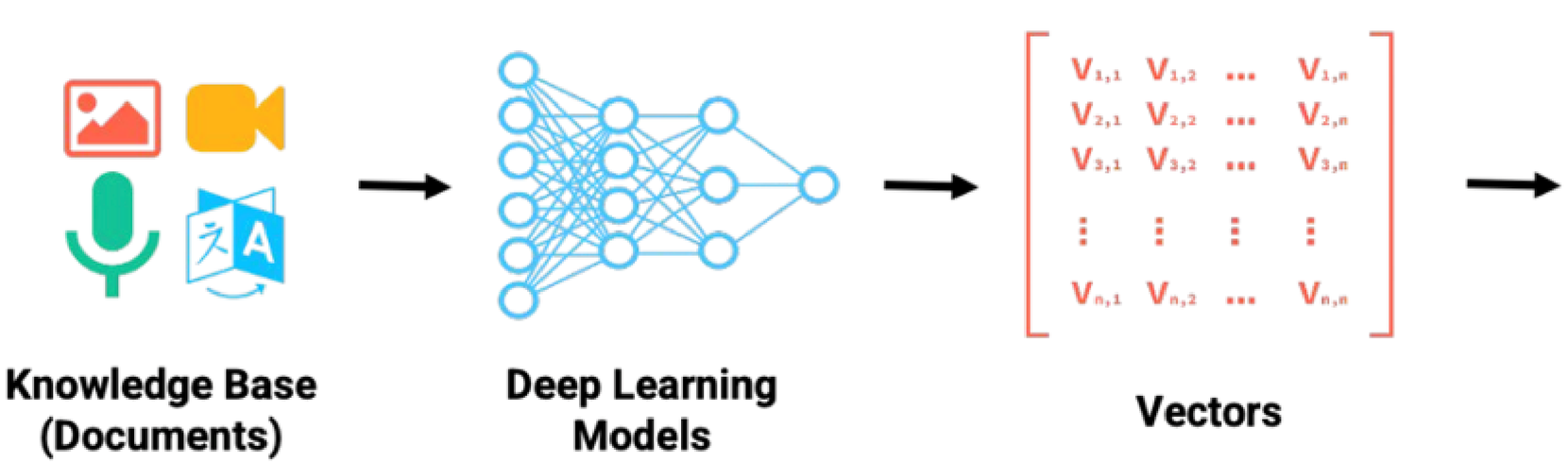
- University websites are like treasure troves full of all kinds of information, from class schedules to research papers. But sorting through all that information can be tough because people search for things in different ways, and the website might not always understand what you're asking for.



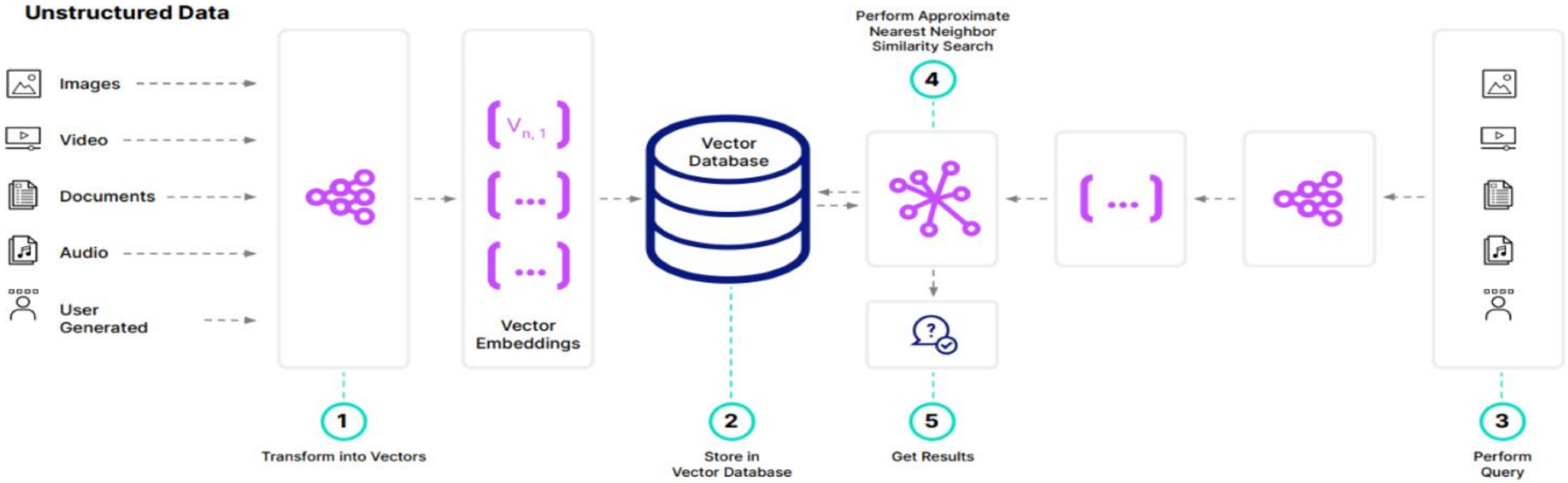

UniBot

Vector Database

Vector databases are purpose-built to store, index, and query vector embeddings from unstructured data

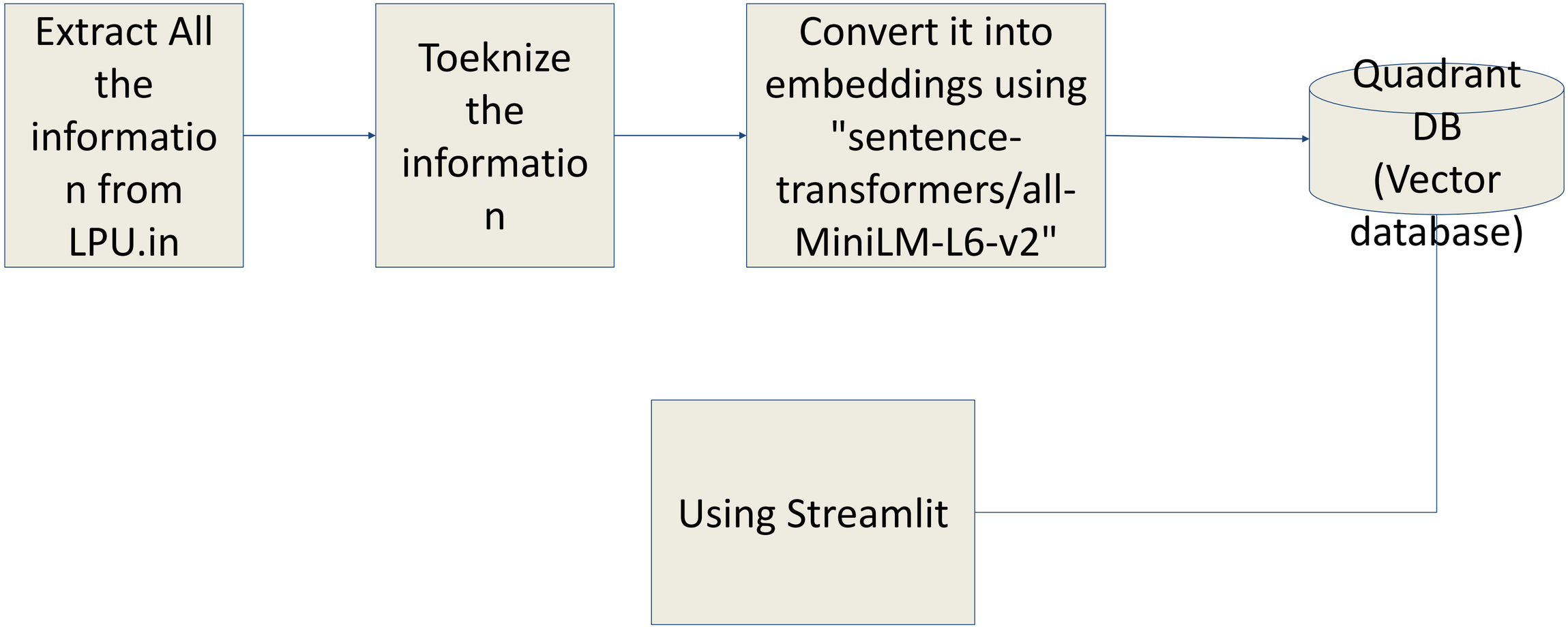


How it works?


Unstructured Data
Images
Video
Documents
Audio
User Generated
1
Transform into Vectors
Vector Embeddings
Vector Database
2
Store in Vector Database
Perform Approximate Nearest Neighbor Similarity Search
4
5
Get Results
3
Perform Query

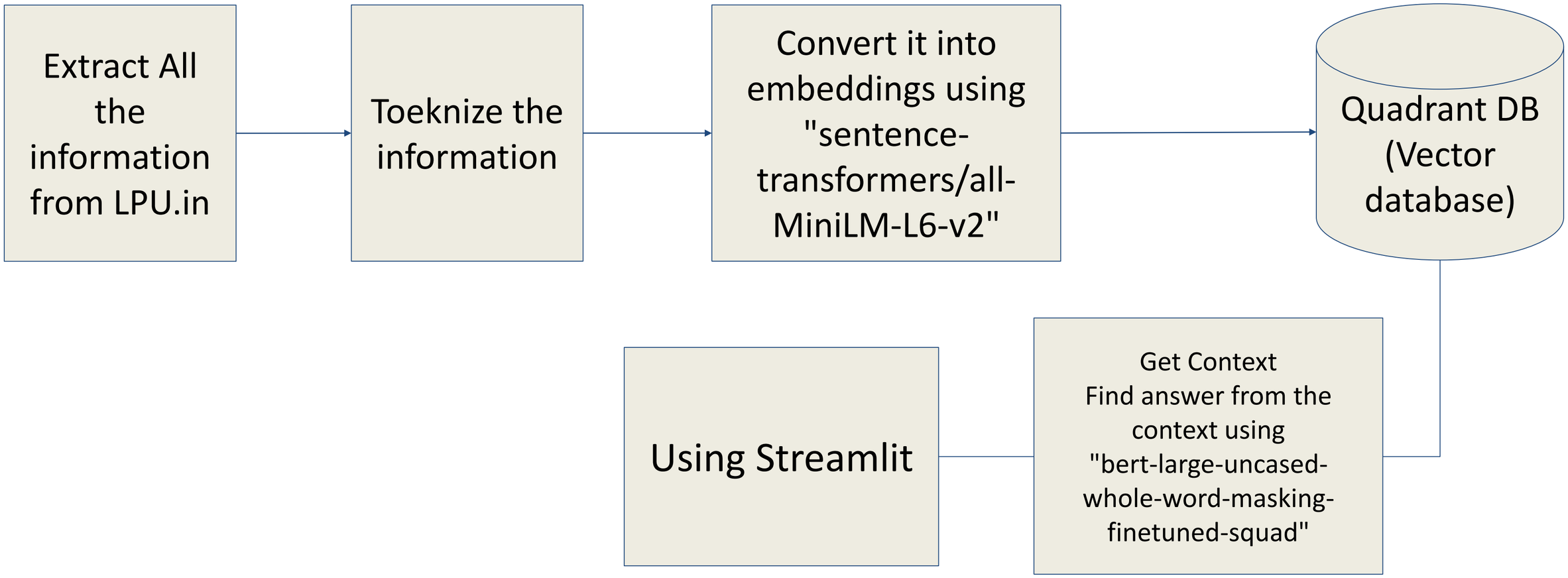
System High Level View Phase I

Backend (Developed entirely in Python)



System High Level View Phase II

Backend (Developed entirely in Python)



Sample Query

The system will provide top 3 answers with assigned probability. If probability is less than .5 then system will say, could not find the answer. Please rephrase your question.

Question: What is the admission criteria for Ph.D.?

Answer 1: **Entrance Test:** LPUNEST (Ph.D.) will provide a gateway to all Ph.D. programmes in LPU and will act as an Entrance test for admission to Ph.D. Programmes at LPU. For details, refer “Eligibility”. With probability : .80

Answer 2: Lovely Professional University is committed to provide an outstanding environment for excellence in Research and Development in all the disciplines. The gamut of LPU research activities makes it a unique academic environment that stimulates creativity and critical thinking, and provides impetus to recruiting, supporting, and developing highly-skilled motivated researchers. With probability : .65

FrontEnd - Streamlit

What is Streamlit?

- Streamlit is a free and open-source framework to rapidly build and share beautiful machine learning and data science web apps
- It is a Python-based library specifically designed for machine learning engineers.
- Streamlit allows you to create a stunning-looking application with only a few lines of code.

How to use Streamlit

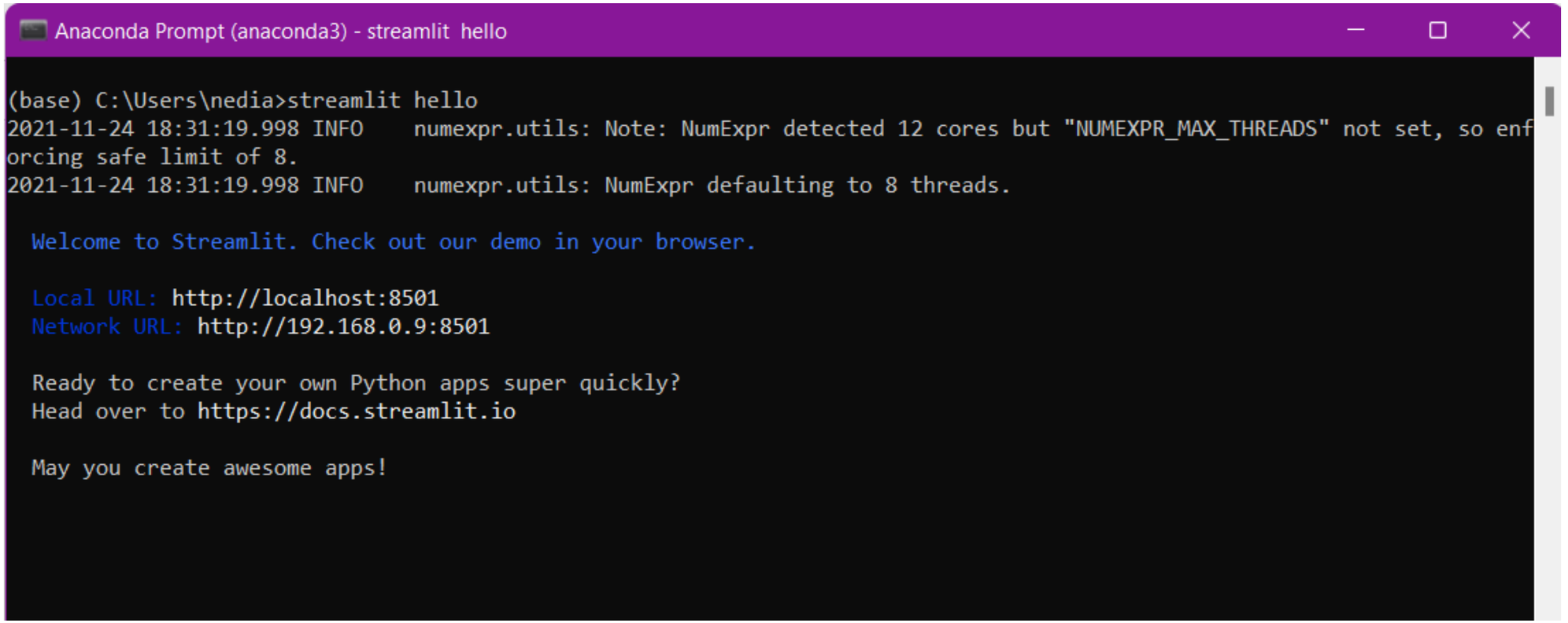
Install Streamlit

On Windows:

1. Install Anaconda and create your environment
2. Open the terminal

2. Test if the installation worked:

`streamlit hello`



How to scrap all the data from University

The sample code will extract all the link from lpu site and then we can download all the text data from these pages.

```
import requests

from bs4 import BeautifulSoup



url = 'https://www.lpu.in/'

reqs = requests.get(url)

soup = BeautifulSoup(reqs.text, 'html.parser')


urls = []

for link in soup.find_all('a'):

    print(link.get('href'))
```

Tokenize Text

We shall use Langchain:

- Tokenize Text, use overlapping text for better answers
- Convert it into embeddings
- Insert into Vector DB (Qdrant DB)

https://python.langchain.com/docs/modules/data_connection/document_transformers/split_by_token/

https://qdrant.tech/documentation/frameworks/langchain/

System Phased Approach

Phase 1: will provide semantic similarity serach on the lpu data

- Extract Text
- Tokenize it and generate embeddings
- Insert it into Vector DB

Phase 2: will provide question and answer using the context from Phase 1

- It will use (model_name = "bert-large-uncased-whole-word-masking-finetuned-squad")

SCIOMETRIX
Engage Patients, Improve Outcomes.

Thank
You