

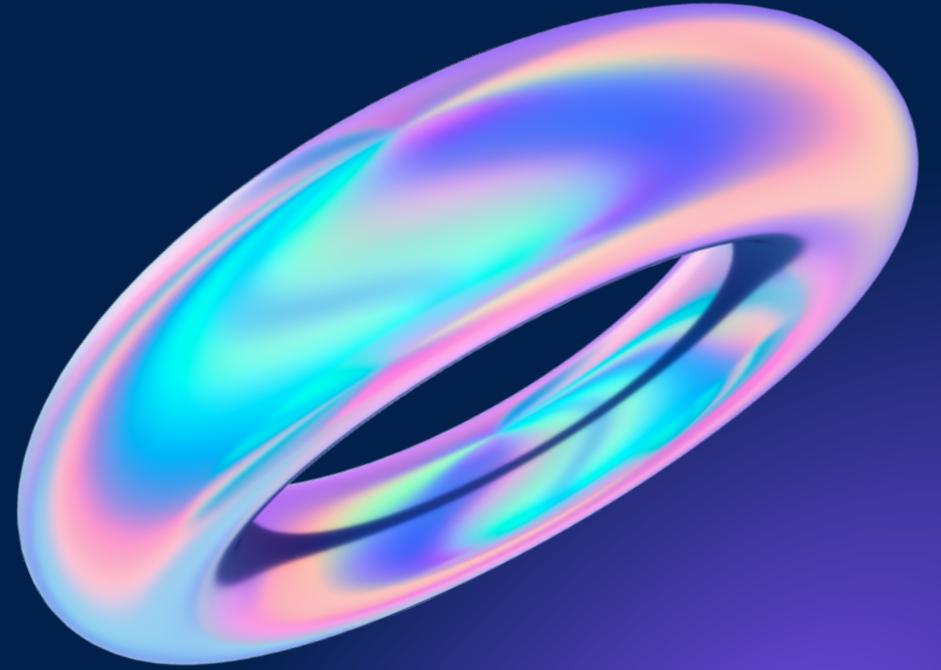


Mentored By:
Dr. Jatinder Pal Singh



DocHelper

Abdul Rehman (GCS 2040305)
Chandramauli Gupta (GCS 2040162)
Shashwat (GCS 2040395)



About DocHelper

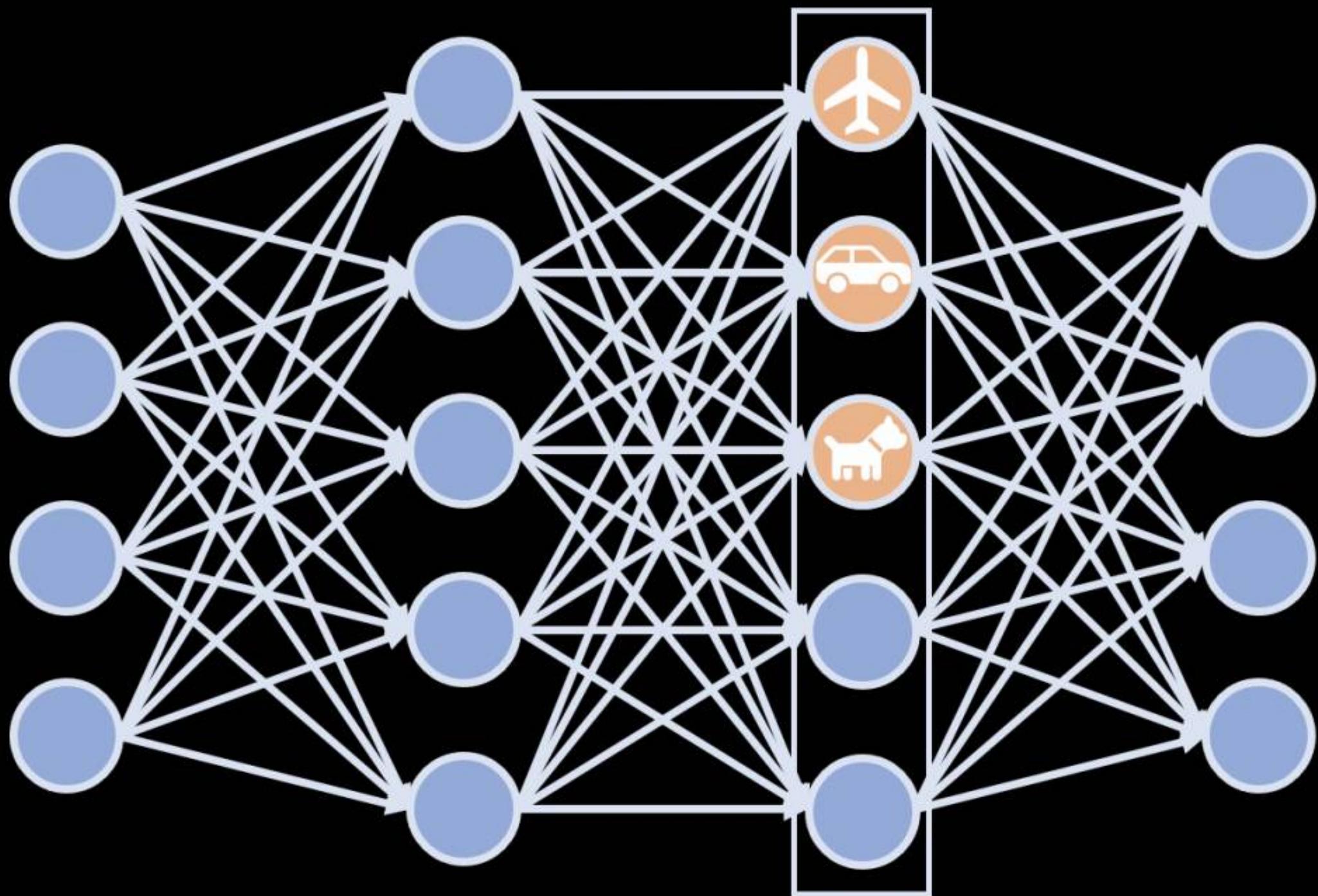
The core objective of DocHelper is to facilitate users in interacting with PDF documents in a more intuitive and dynamic manner. With the application of the BERT model (Bidirectional Encoder Representations from Transformers), this project demonstrates the potential of deep learning models in extracting meaningful insights from intricate textual data.

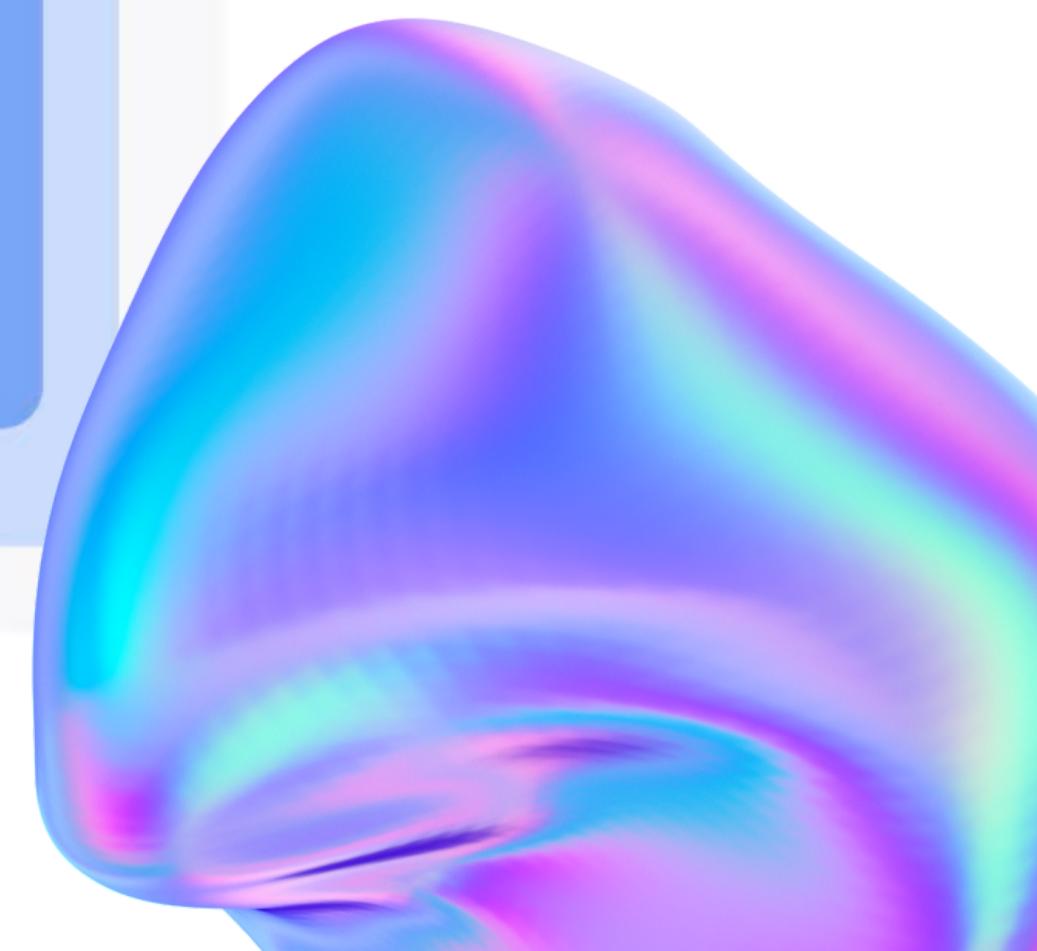
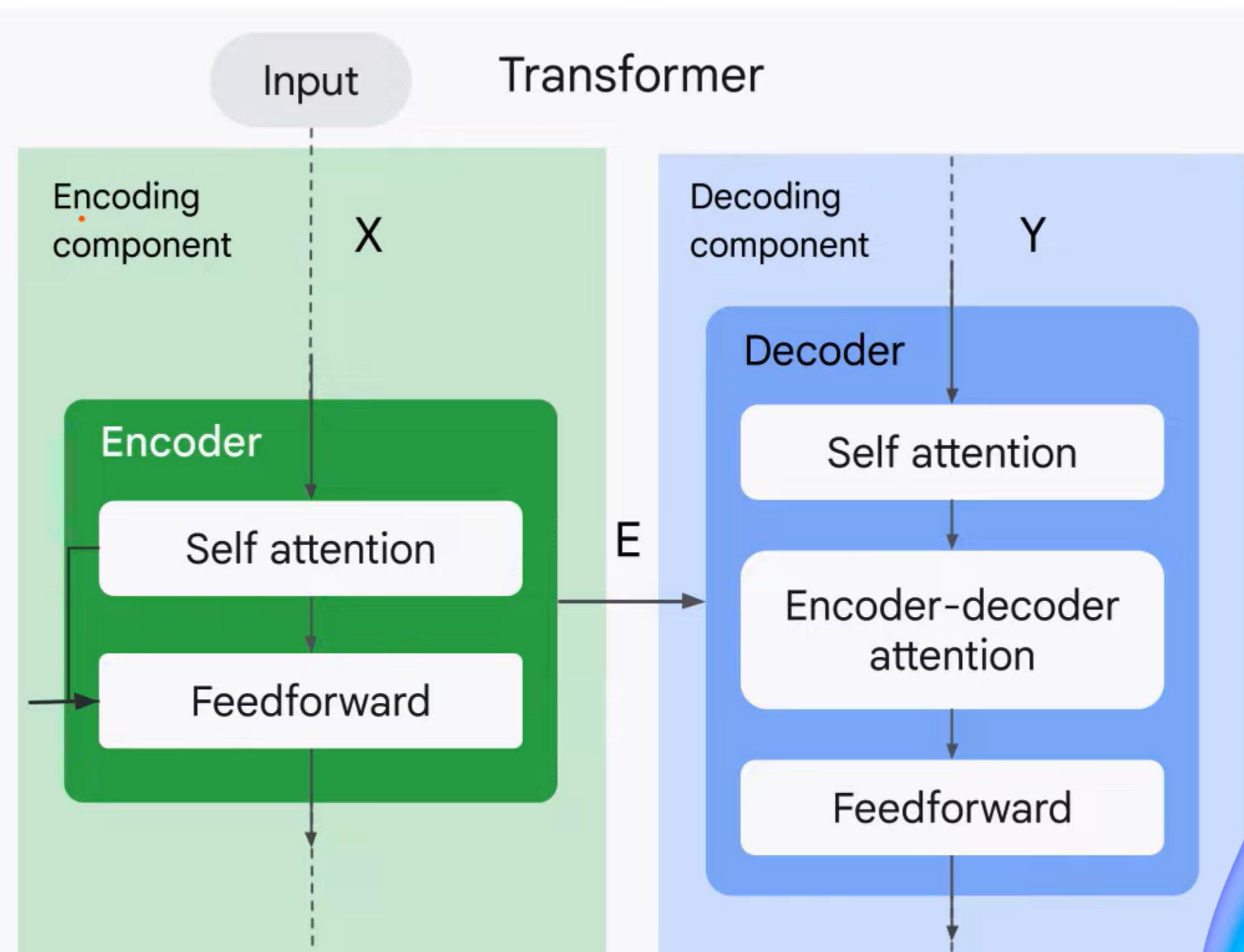
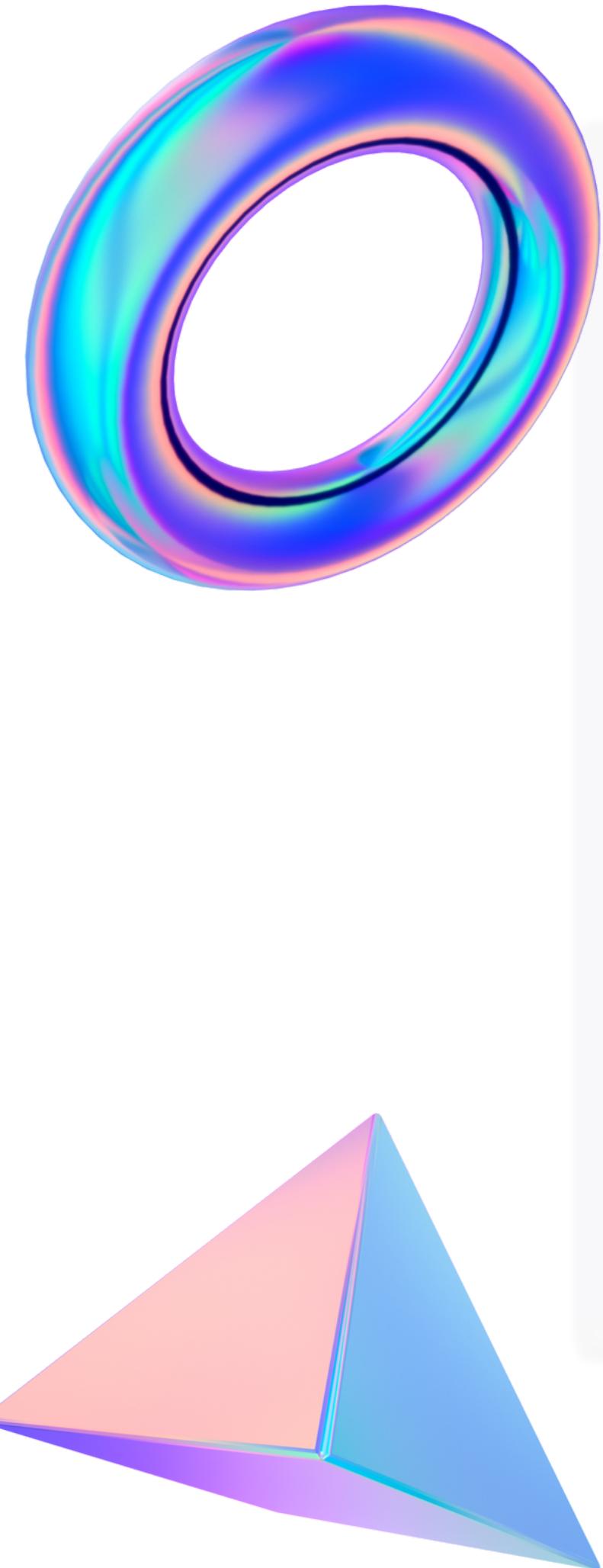


OBJECTIVES

- Enhance Information Retrieval from PDFs
- Streamline Document Understanding
- Promote Learning and Research
- User-Friendly Document Exploration
- Efficient Document Review

Deep Learning

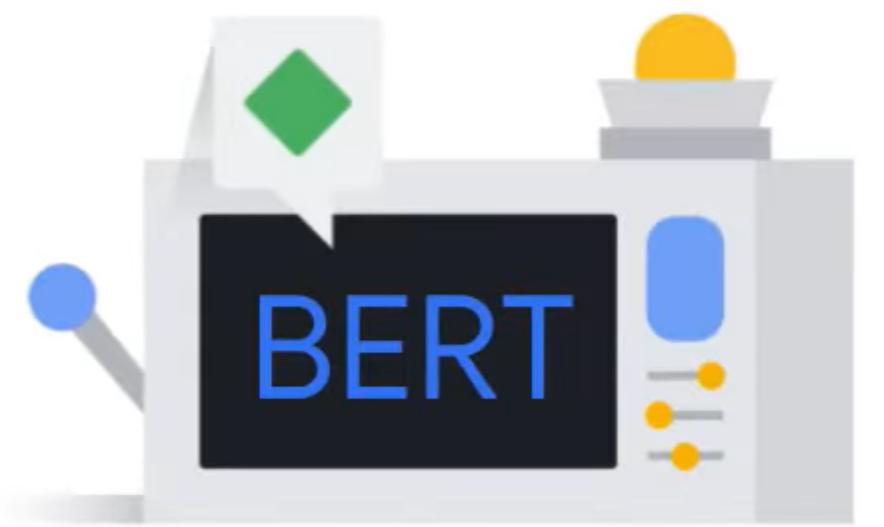




BERT Overview

- Trained in two variations
- Able to handle long input context
- Trained on entire Wikipedia and BookCorpus
- Trained for one million steps
- Targeted at multi-task objective
- Trained on TPU
- Works at both sentence-level and token-level tasks
- Be fine-tuned for many different tasks

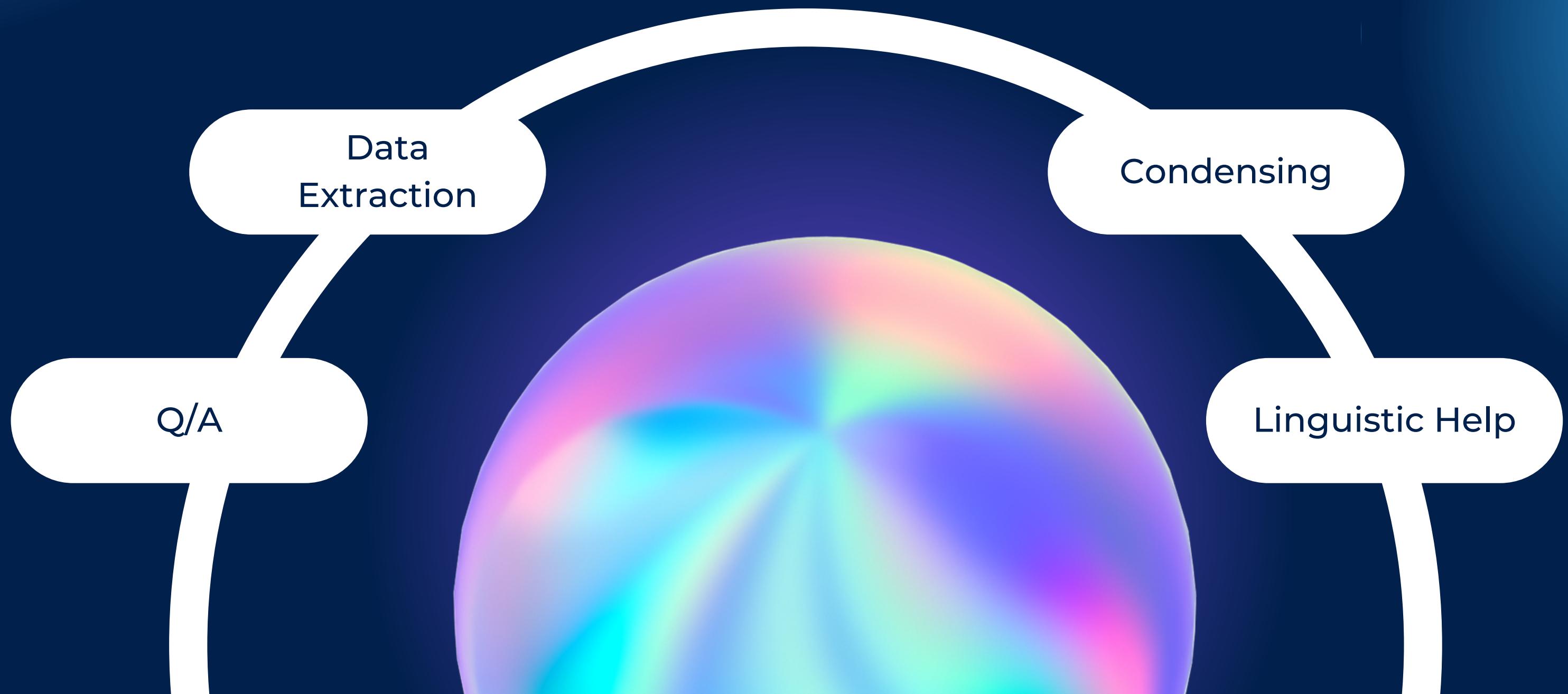
Fine tune



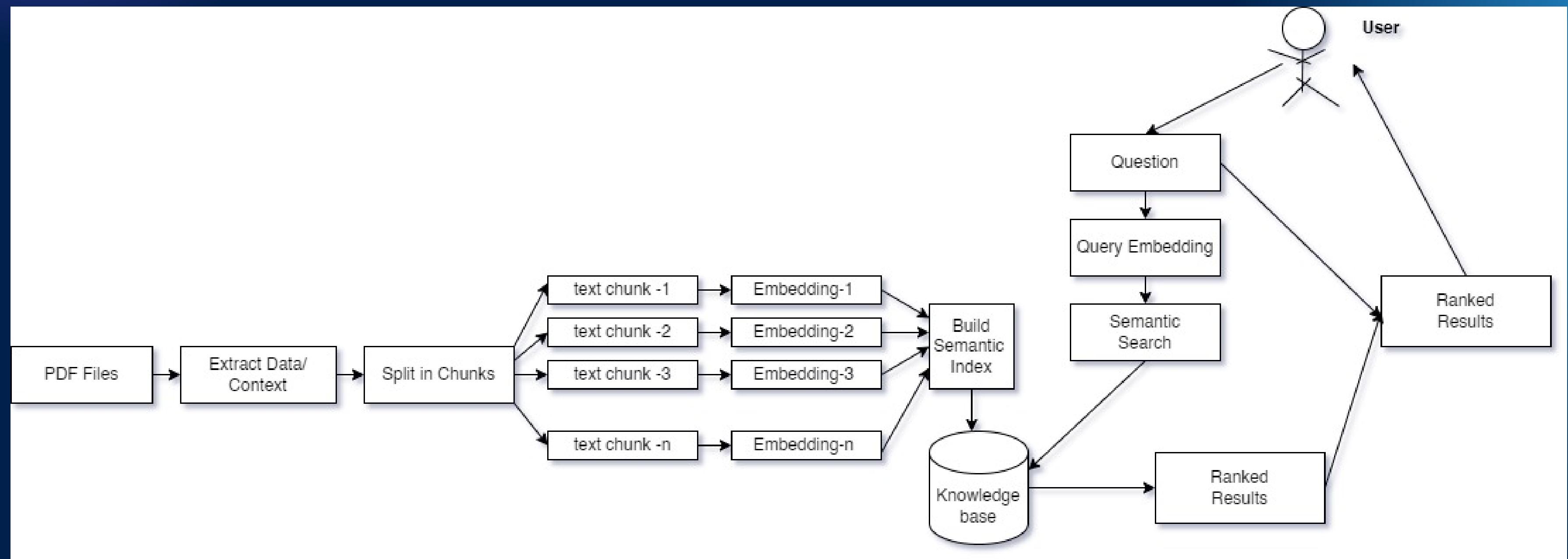
What we have achieved

- Data Preprocessing
- BERT Model Integration
- Streamlit Web Framework Deployment
- BERT Inference and Response Generation
- User Interface Design
- Testing and Evaluation

Functionality



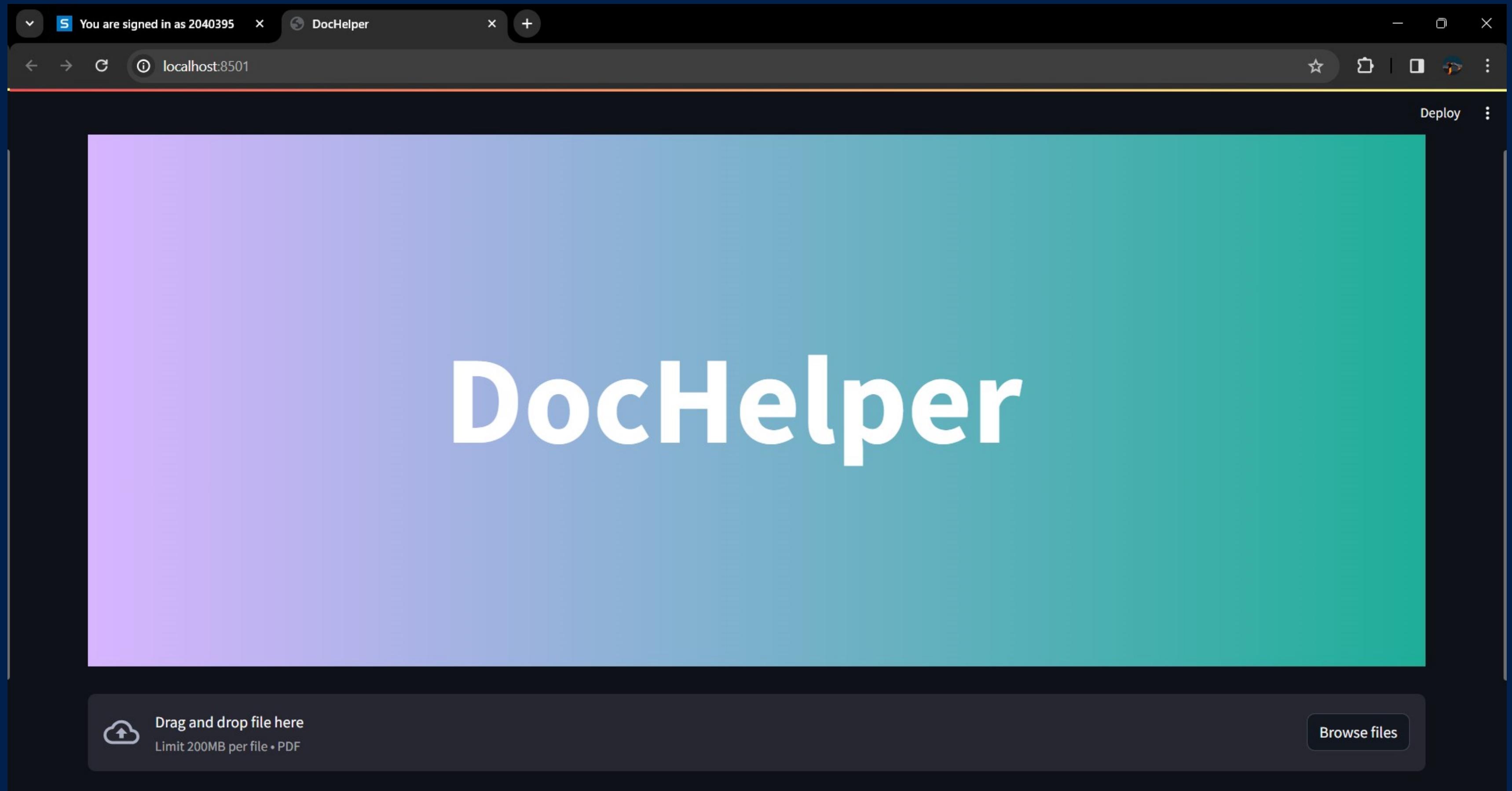
Flow Chart



Technologies

- 1. Langchain
- 2. Python
- 3. PDF Processing Tools (Eg-py2PDF)
- 4. Large Language model (LLM)
- 5. HTML/CSS
- 6. PyTorch
- 7. Streamlit

Interface



Q/A Chat

You are signed in as 2040395 DocHelper

localhost:8501

warming of the planet—a phenomenon commonly referred to as global warming.

The consequences of climate change are multifaceted and extend to various aspects of our environment. One of the most evident impacts is the alteration of weather patterns, resulting in more frequent and severe extreme weather events. Hurricanes, droughts, floods, and wildfires have become more intense and unpredictable, causing widespread destruction and displacing communities. Rising sea levels, attributed to the melting of polar ice caps and glaciers, pose a direct threat to coastal regions, jeopardizing the livelihoods of millions of people and the existence of low-lying island nations.

Beyond the environmental realm, climate change poses a significant risk to global economies.

Deploy :

Embeddings are created successfully!

Chat Here

Hey there! 

 I am ready to help you

Made with Streamlit

Limitations

- Monolingual (English)
- Can't parse images
- Limited Document Format Compatibility
- Response Length Constraint



HARDWARE REQUIREMENTS

- Minimum 8GB RAM (16GB recommended).
- Adequate storage with SSD recommended.
- Optional but recommended GPU for accelerated computations.
- Internet connectivity for updates and potential cloud integration.

SOFTWARE REQUIREMENTS

- Linux-based OS, or Windows/macOS (preferred).
- Python 3.x with TensorFlow or PyTorch for BERT model.
- Virtual environment and package manager for dependency management.
- Web browser compatibility (Chrome, Firefox, Safari).

Results:

1. Precision: 0.654
2. Recall: 0.4047
3. F1 Score: 5.0479
4. EMScore (Exact Match Score):
0.716

Future Prospect

- Text from image
- Extending to other formats
- Support for regional languages
- Output as Infographics
- User Interface Enhancements
- Collaboration and Integration with External Tools
- Feedback Mechanism and Continuous Learning

Team

Abdul Rehman

Responsible for machine
learning tasks

**Chandramauli
Gupta**

Responsible for Web interface

Shashwat

Responsible for Deployment of
ML tasks



Thank You