

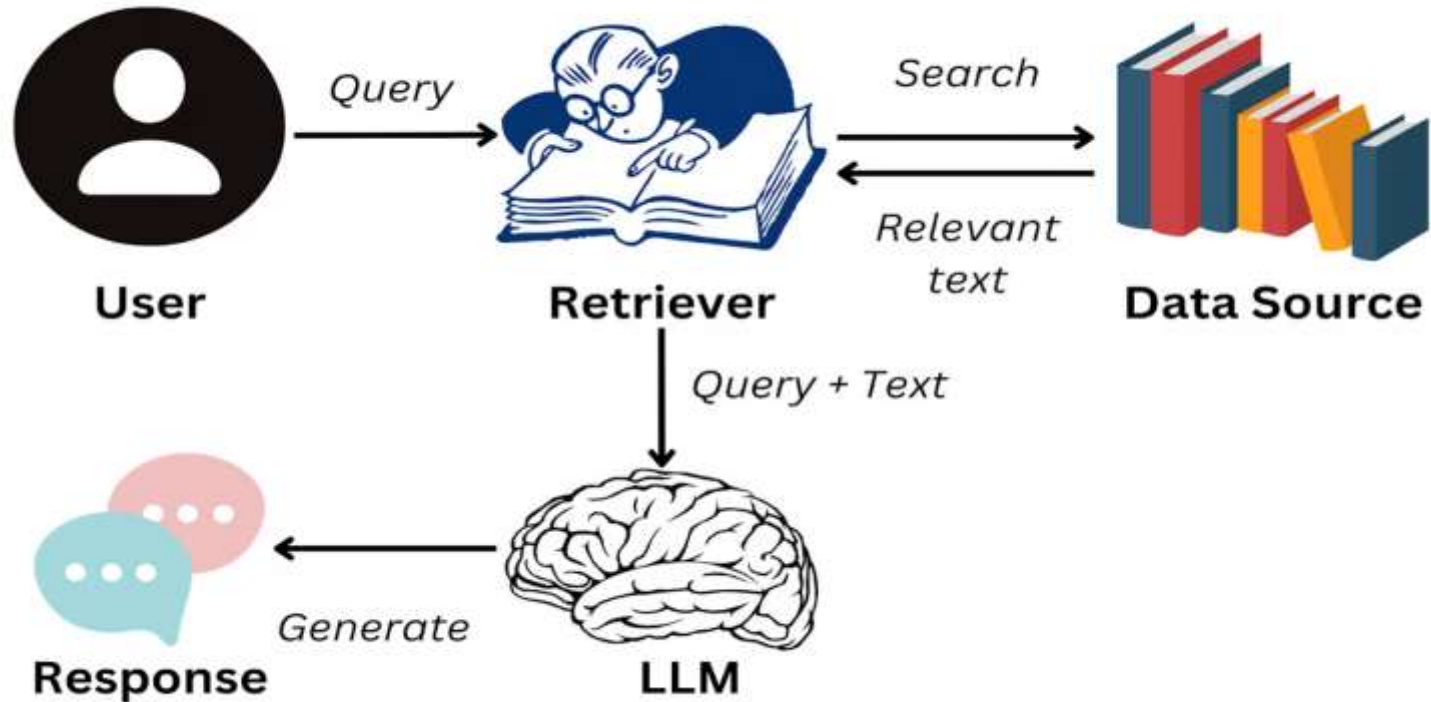
RAG and Image Classification

Presented By Akshit Jain

Part A

(A) Demonstrate LangChain with a suitable example involving RAG, input/output parsing, output formatting, etc.

Retrieval Augmented Generation



Retrieval Augmented Generation

Query Input – User submits a query or question to the RAG system

Document Retrieval – Retriever component searches a large corpus to find relevant documents based on the query.

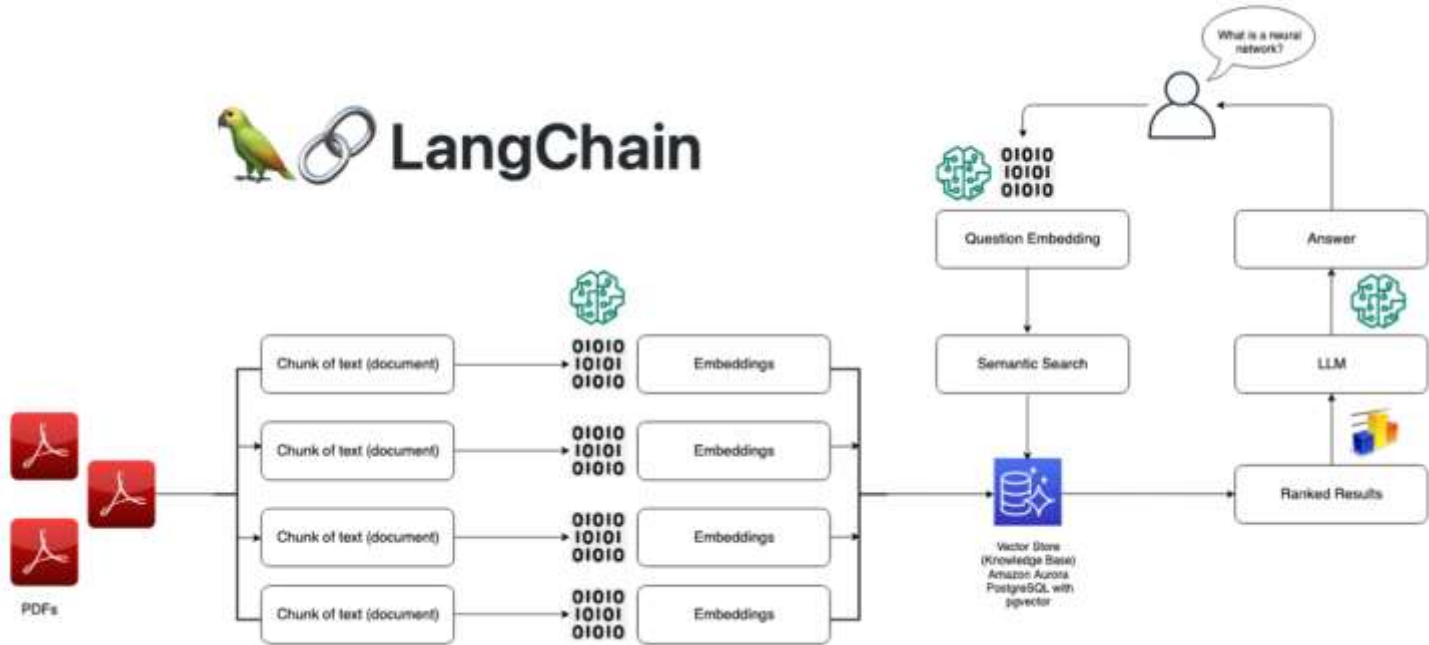
Context Augmentation – Retrieved documents provide additional context and background information to the query.

Response Generation – Generator component uses the augmented query to produce a coherent and informative response.

Output Parsing – Raw → Cleaned answer string

Output Formatting – Final string includes *Answer* + *Source*

LangChain Ecosystem



LangChain RAG

Components

Retriever – FAISS (Vector search over internal documents)

LLM – LangChain's ChatOpenAI

RAG Chain – Combines query → Document Retrieval → Answer Generation

Input Parsing: Manual Query Implementation

RAG Execution – LangChain RetrievalQA Chain

Output Parsing – Raw → Cleaned answer string

Output Formatting – Final string includes *Answer* + *Source*

Results

```
a.hasmukhkumarjain@c0244 X Assignment3_Part_5.ipynb X AkshithHasmukhKumarjain, A X
Python 3 (ipykernel)

==== NON-RAG RESPONSE ====
There is no specific US Department of Government Efficiency. However, there are various government agencies and offices dedicated to improving government efficiency and effectiveness, such as the Office of Management and Budget (OMB) and the Government Accountability Office (GAO).

The head of the OMB is the Director, who is appointed by the President and confirmed by the Senate. As of October 2021, the Director of the OMB is Shalanda Young. The head of the GAO is the Comptroller General, who is appointed by the President and confirmed by the Senate. As of October 2021, the Comptroller General of the GAO is Gene L. Dodaro.

==== RAG RESPONSE ====
=== INPUT PARSING ===
Original user query: What is the US Department of Government Efficiency who heads it?
Augmented user query: What is the US Department of Government Efficiency who heads it?Be specefic with your answer

=== OUTPUT PARSING ===
Raw RAG answer: The U.S. Department of Government Efficiency (DOGE) is a temporary organization within the U.S. federal government aimed at modernizing federal technology and software to enhance governmental efficiency and productivity. It was established by President Donald Trump on January 20, 2025. Elon Musk leads the DOGE organization.
Augmented RAG answer: The answer to your question is: The U.S. Department of Government Efficiency (DOGE) is a temporary organization within the U.S. federal government aimed at modernizing federal technology and software to enhance governmental efficiency and productivity. It was established by President Donald Trump on January 20, 2025. Elon Musk leads the DOGE organization.

=== OUTPUT FORMATTING ===
Answer: The answer to your question is: The U.S. Department of Government Efficiency (DOGE) is a temporary organization within the U.S. federal government aimed at modernizing federal technology and software to enhance governmental efficiency and productivity. It was established by President Donald Trump on January 20, 2025. Elon Musk leads the DOGE organization.
Source: DOGE_Government_Announcement.docx

=== FINAL RAG OUTPUT ===
Answer: The answer to your question is: The U.S. Department of Government Efficiency (DOGE) is a temporary organization within the U.S. federal government aimed at modernizing federal technology and software to enhance governmental efficiency and productivity. It was established by President Donald Trump on January 20, 2025. Elon Musk leads the DOGE organization.
Source: DOGE_Government_Announcement.docx
```

Part B

(B) Classify CIFAR 10 images with a CNN model but study the effect on the performance by varying the following modeling aspects:

- 1) Batch Normalization and Dropout
- 3) Number of convolution and pooling layers
- 4) Activation function

CIFAR-10 Classification

- **Dataset:** CIFAR-10 has 60,000 32x32 RGB images
- **Output Class:** 10 Classes
- **Frequency:** Each class has exactly 5,000 rows.

Training Setup

- **Optimizer:** Adam (0.001)
- **Epochs:** 20
- **Batch Size:** 64
- **Loss:** Categorical Crossentropy
- **Metric:** Accuracy

Convolution Neural Network Model

Input Variables

- **num_conv_layers:** [2,3,4]
- **Activation:** ['relu', 'tanh']
- **use_batch_norm:** [True, False]
- **dropout_rate:** [0.0, 0.25, 0.50]

Output Variables

- Final Training Accuracy
- Final Validation Accuracy

Convolution Neural Network Model

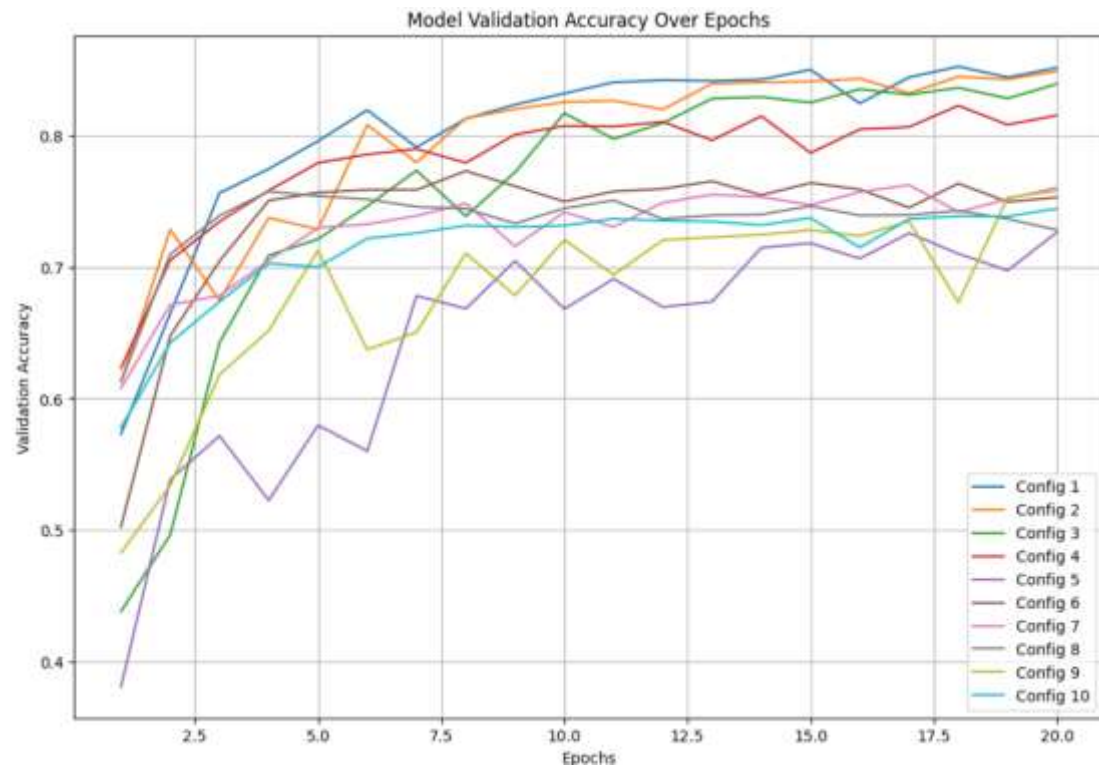
Variables Used:

- Batch Normalization and Dropout
- Number of Convolution/Pooling layers
- Activation Functions

Insights

- Batch Normalization improves validation accuracy & training stability
- Moderate dropout 0.25 performs best
- No regularization leads to overfitting

Model Validation Accuracy



Configurations Results

	Config	Num Conv Layers	Activation	Batch Norm	Dropout Rate	Final Train Accuracy	Final Val Accuracy
0	1	4	relu	True	0.25	0.93714	0.8512
1	2	3	relu	True	0.25	0.92888	0.8485
2	3	4	relu	True	0.50	0.84076	0.8389
3	4	2	relu	True	0.25	0.94836	0.8150
4	5	4	tanh	True	0.25	0.73384	0.7262
5	6	4	relu	False	0.00	0.95688	0.7527
6	7	3	tanh	False	0.25	0.75466	0.7601
7	8	2	relu	False	0.00	0.98652	0.7280
8	9	3	tanh	True	0.00	0.89492	0.7577
9	10	2	tanh	False	0.25	0.80810	0.7442

Evaluation

- **Training Dataset Size:** 50,000 rows.
- **Testing Dataset Size:** 10,000 rows.
- **Output Class:** [0,1,2,3,4,5,6,7,8,9]
- **Output Class Count:** Each class has 5,000 rows
- **Final Train Accuracy:** 93.7%
- **Final Test Accuracy:** 85.7%

Conclusion

- BatchNorm + Dropout combo yields strong generalization.
- ReLU > Tanh in deeper models.
- **Actionable Insights:** Useful for classifying images.
- **Next Steps:** Fine-tuning and adding more epochs.

Thank You