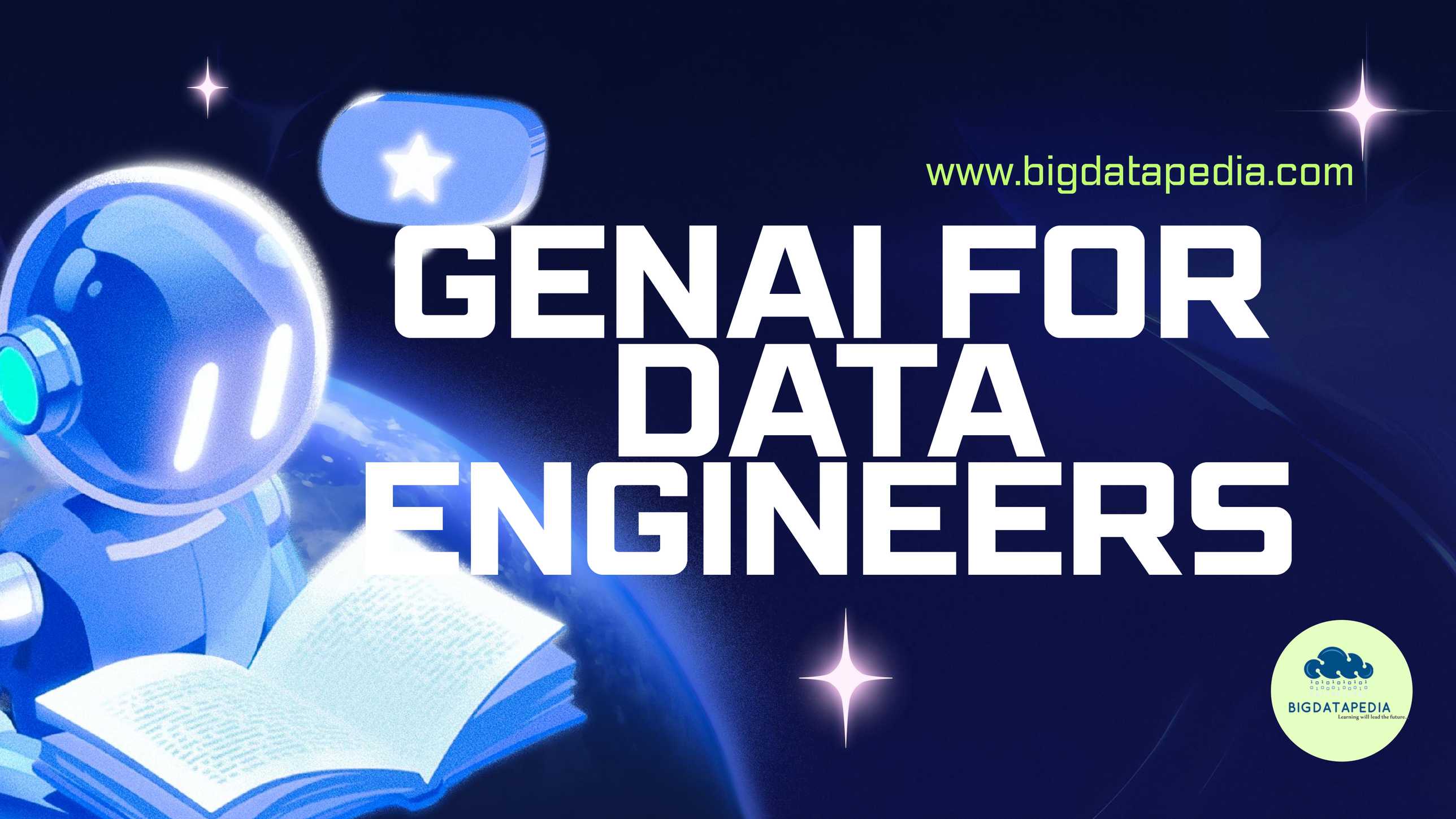

www.bigdatapedia.com
GENAI FOR DATA ENGINEERS
BIGDATAPEDIA
Learning will lead the future.

LEARNING GOAL FOR AI ENGINEERS

- Understand foundations of generative AI & LLMs.
- Apply LLMs to data engineering workflows (ETL, SQL, metadata).
- Build retrieval-augmented pipelines for data access.
- Fine-tune & deploy small models for domain-specific use cases.
- Integrate LLMOps with Airflow, Spark, and data warehouses.

MORE INFORMATION

TOPICS EXPLAINED

DATA

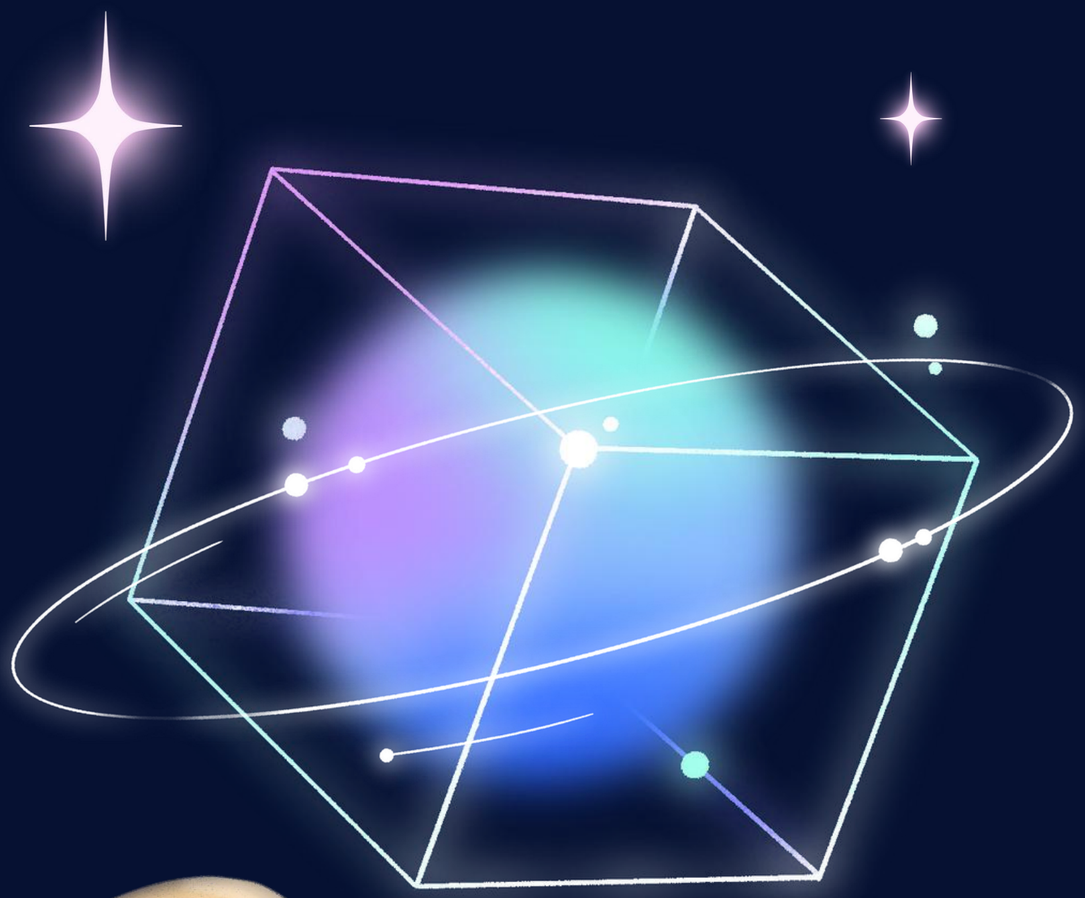
- Hive SQL queries
- Iceberg SQL Queries
- Kafka Streaming
- Airflow workflows
- PySpark Machine Learning

GENERAL AI

- NLP Introduction
- nltk feature extractions
- Deep Learning Introduction
- HuggingFace Transformers
- Vector Databases

SUPER AI

- Prompt Engineering
- Tokenization
- Popular LLMs
- RAG
- Fine Tuning
- Langchains



MACHINE LEARNING

- All About PySpark Dataframe Transformations
- PySpark ML with Supervised Learning Example
- PySpark ML with UnSupervised Learning Example
- PySpark ML Pipelines

INCLUDES SUPERVISED

USED IN FRAUD DETECTION

DEEP LEARNING

- NEURAL NETWORK POWER
- TRANSFORMERS
- DL ENCODERS/DECODERS

NATURAL LANGUAGE PROCESSING

- NLP Introduction
- nltk feature extractions
- Bag Of Words
- TF-IDF
- Word2Vec
- Tokenization
- Stemming
- Lemmatization

REALTIME PRACTICAL

SENTIMENTS

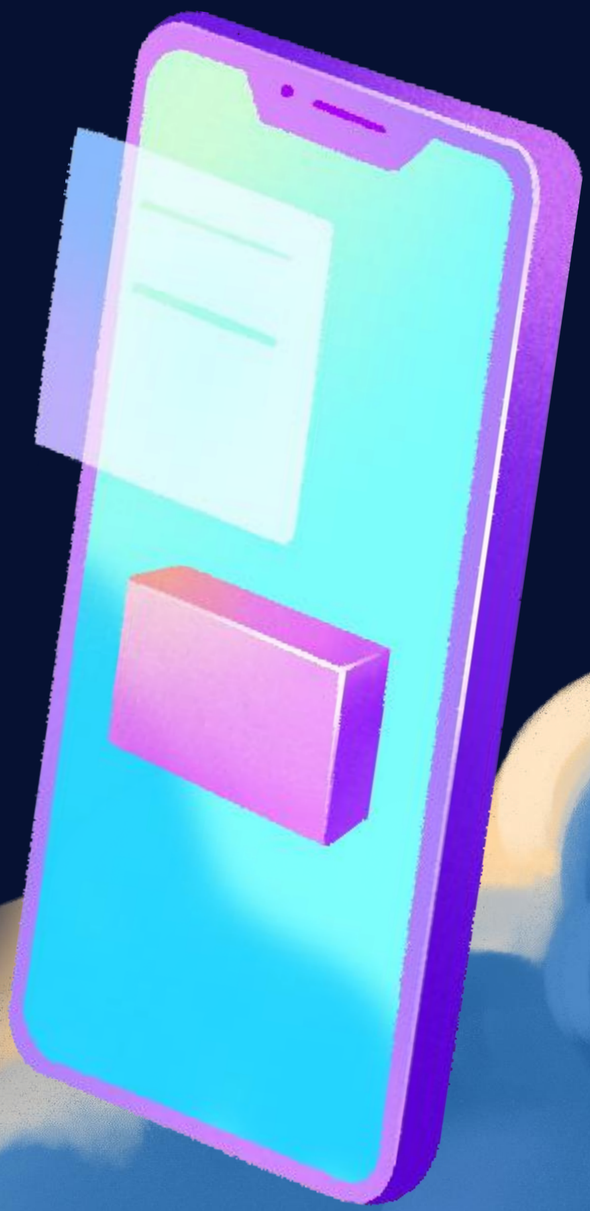
CLASSIFICATION

TEXT SUMMARIZATION

TEXT GENERATIONS

TRANSLATIONS

REAL WORLD NLP PROCESS



CHAT PROMPTING

VECTOR DB

HUGGINGFACE

MULTI MODEL

ENCODERS /DECODES MODELS

LLM

LARGE LANGUAGE MODEL

AI AGENTS

- RAG
- LANGCHAINS
- AI AGENTS
- FINE TUNING
- MULTI MODEL
- MCP

CAPSTONE PROJECTS

WWW.BIGDATAPEDIA.COM

inquiry@bigdatapedia.com

+919715010010

THANK YOU!