KRISHNAIK ACADEMY

Ultimate Data Science & GenAI Bootcamp V2.0

PYTHON + STATS

MACHINE LEARNING

DEEP LEARNING

GEN AI

RAG

AGENTIC AI

NATURAL LANGUAGE PROCESSING

MACHINE LEARNING

RETRIEVAL AUGUMENT GENERATION

AGENTIC AI

VECTOR DB

GEN AI

This course is a comprehensive journey through modern Data Science and AI, designed to take learners from Python programming basics all the way to next-generation Agentic AI systems. You'll begin by mastering Python, statistics, and core machine learning techniques then progress into deep learning, generative AI, retrieval-augmented generation (RAG), and LangGraph-based multi-agent systems. By the end of this year-long program, you'll not only be fluent in data analysis, ML, and DL, but also understand how the world is moving towards autonomous AI systems that reason, retrieve, and collaborate.

Learning Objectives

- Write efficient, reusable Python code using OOP, file handling, and concurrency.
- Manipulate and process structured and unstructured data with NumPy, Pandas, and Streamlit.
- Use statistical reasoning to explore, interpret, and validate real-world datasets.
- Understand the principles of supervised and unsupervised learning.
- Build, evaluate, and tune regression, classification, and clustering models.
- Work with RNNs, LSTMs, GRUs, Attention, and Transformer architectures (BERT, GPT).
- Manage and query data using SQL (PostgreSQL + Supabase) and NoSQL (MongoDB Atlas).
- Create data-driven web applications with Streamlit for interactive insights.
- Understand Generative AI, its model architectures, and real-world applications.
- Use vector databases (Chroma, Pinecone, FAISS) for embedding-based search and retrieval.
- Build Retrieval-Augmented Generation (RAG) systems using LangChain and LLMs.
- Create LLM workflows, agents, and chains using LangChain.
- Understand the concepts of Agentic AI, multi-agent collaboration, and autonomous decision-making.
- Design agent architectures for research, automation, and reasoning using LangGraph and MCP (Model Context Protocol).
- Integrate human feedback, memory, and RAG for adaptive, self-improving AI agents.

Course Information

Prerequisites

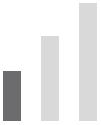
No prior experience is required this course is designed to take you from the ground up. The curriculum begins with Python programming and basic math, ensuring that even beginners can follow along. A general curiosity about technology, problem-solving, and AI systems is enough to get started.
For learners with prior experience in coding or data analysis, the early modules will serve as a refresher and skill alignment phase, before advancing into cutting-edge areas like Generative AI, RAG, LangChain, and Agentic AI.

The course is designed to be completed over a duration of approximately 10 to 12 months, providing an in-depth exploration from Python basics to GenAI (Including Agentic AI), with plenty of time for practical implementation and real-world applications.

Estimated Time

10-12 months 6hrs/week*

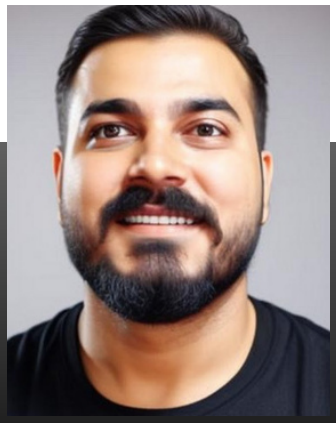
Required Skill Level



Begineer

Krish Naik Academy Team

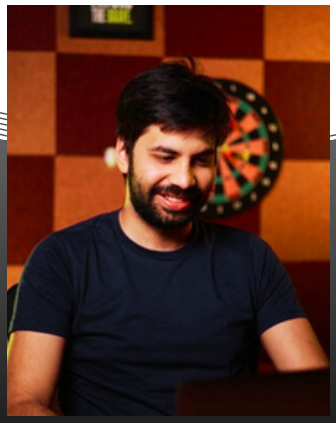
Sourangshu Pal
Senior Data Scientist
Linkedin



Krish Naik
Chief AI Engineer
Linkedin

Monal Kumar
Data Scientist
Linkedin

Sunny Savita
GenAI Engineer
Linkedin



Mayank Aggrawal
Senior ML Engineer
Linkedin

Darius B.
Head of Product
Linkedin

Module 1

Python Foundations

In this module, learners build a strong foundation in Python programming — understanding data types, variables, operators, and control flow. You'll learn to work with strings, lists, sets, and dictionaries, and write functions that make code modular and efficient. This module sets the stage for advanced programming and data manipulation tasks.

Topics	
Introduction to Python	Overview of Python and comparison with other programming languages, Python objects: Numbers, Booleans, and Strings
Core Language Concepts	Container objects and mutability, Operators: Arithmetic, Bitwise, Comparison, and Assignment, Operator precedence
Control Flow	Conditional statements (if, if-elif-else), Loops (for, while), Break and continue statements, Range function
Strings and Data Structures	String basics, inbuilt methods, splitting and joining, formatting, Basic data structures in Python: Lists, Tuples, Sets, and Dictionaries, List and Dictionary comprehensions, Dictionary view objects
Functions and Iterators	Function basics and parameter passing, Iterators and generator functions, Lambda functions, Map function and functional style programming

Module 2

Advanced Python Programming

This module builds on Python fundamentals by introducing object-oriented programming and code organization. You'll learn how to manage files, handle exceptions, and structure code for reusability. The module also introduces Python's concurrency concepts to prepare you for performance-driven applications.

Topics	
Object-Oriented Programming (OOP)	OOP concepts and class creation, Inheritance, Polymorphism, Encapsulation, and Abstraction, Decorators, class methods, and static methods, Special (Magic/Dunder) methods
File Handling	Reading and writing files, Buffered read and write operations, Other file handling methods
Modules and Exception Handling	Importing and managing modules, Using try-except blocks, Custom exceptions and best practices for error handling

Module 3

Numerical Computing with NumPy

NumPy introduces learners to efficient numerical computation in Python. You'll learn how to create, index, and manipulate arrays — the core data structure for numerical processing. This module is essential for performing fast mathematical and statistical operations that form the basis of data analysis and machine learning.

Topics	
Introduction to NumPy	Understanding NdArray objects, Data types and array attributes
Array Creation and Indexing	Creating arrays from existing data and numerical ranges, Indexing and slicing arrays
Array Operations	Broadcasting and iterating over arrays, Array manipulation techniques, Arithmetic and mathematical operations
Statistical and Logical Functions	Statistical functions, sorting, searching, and counting operations, Binary and string functions
Matrix Operations and Copies	Matrix library overview, Copy vs view behavior in NumPy arrays

Module 4

Data Analysis with Pandas

In this module, learners will explore the Pandas library for data manipulation and analysis. You'll learn to handle different data formats, filter and transform data efficiently, and visualize trends directly from DataFrames. Pandas is a cornerstone for building strong data-handling skills in data science.

Topics	
Pandas Core Concepts	Series, DataFrame, and Panel structures, Basic functionality and reindexing
Data Operations	Iteration, sorting, indexing, and selecting data
Text, Date, and Time Handling	Working with text data, Date and time functionality, Timedelta and categorical data
Statistical and Visualization Tools	Descriptive statistics and built-in statistical functions, Basic visualization with Pandas

Module 5

Data Visualization with Matplotlib & Seaborn

This module focuses on representing data visually using Matplotlib and Seaborn. You'll learn to create and customize visualizations to uncover insights, communicate results, and make data storytelling more effective.

Topics	
Data Visualization Fundamentals	Introduction to Matplotlib, Creating and customizing plots, charts, and figures
Advanced Visualizations	Introduction to Seaborn, Statistical visualizations such as histograms, pair plots, and heatmaps, Styling and aesthetics for better data communication

Module 6

Working with Databases (SQL & NoSQL)

In this module, you'll learn how to manage both structured and unstructured data using SQL and NoSQL databases. Starting with PostgreSQL on Supabase, you'll practice writing queries and building relationships across tables. Then, you'll move to MongoDB Atlas for handling flexible document-based data giving you a complete view of modern data storage systems.

Topics	
Relational Databases with SQL & Supabase	Understanding SQL queries, Primary and foreign keys, Joins and unions, Setting up Supabase with PostgreSQL, Connecting and executing SQL queries
NoSQL with MongoDB Atlas	Getting started with MongoDB, Creating databases and collections, Insert, find, query, and sort operations, Updating and deleting records, Dropping collections

Module 7

Building Interactive Data Apps with Streamlit

This module introduces Streamlit, a powerful library for turning Python scripts into interactive web applications. You'll learn to create dynamic dashboards, use widgets for user input, and manage session states. By the end, you'll be able to deploy simple, interactive data apps that bring your analysis to life.

Topics	
Streamlit Basics	Introduction to Streamlit and its core features, Creating interactive widgets
App Development and State Management	Using session state for dynamic data handling, Implementing async and await for responsive apps

Module 8

Foundations of Statistics

This module introduces the core ideas of statistics that power all data science techniques. Learners will understand different data types, measures of central tendency, and variability. The module also introduces probability distributions and estimation concepts forming the mathematical foundation for data analysis and machine learning.

Topics	
Introduction to Statistics	Basic statistical terms, Types of statistics: Descriptive and Inferential, Types of data and levels of measurement
Measures of Central Tendency and Dispersion	Mean, median, and mode, Range, variance, and standard deviation
Data Behavior and Relationships	Skewness and kurtosis, Covariance and correlation, Random variables and sets
Probability Concepts	Probability basics, Probability density and mass functions, Cumulative distribution function (CDF)
Probability Distributions	Binomial, Poisson, Normal (Gaussian), Bernoulli, and Uniform distributions, Understanding Z-statistics and the Central Limit Theorem, Examples of real-world normal distributions
Estimation	Concept of population and sample estimation, Confidence intervals

Module 9

Advanced Statistical Analysis

This module builds on foundational statistics and introduces inferential techniques for decision-making. You'll learn hypothesis testing, interpret p-values, and understand how Bayesian reasoning refines probability-based predictions. Using the Chi-square test, you'll also learn how to assess model fit and independence in categorical data.

Topics	
Hypothesis Testing	Steps and mechanism of hypothesis testing, P-values and statistical significance, Confidence levels and decision rules
Bayesian Statistics	Bayes theorem and its applications
Chi-Square Tests	Chi-square distribution, Goodness-of-fit test, Performing Chi-square tests using Python

Module 10

Feature Engineering for Machine Learning

Feature engineering bridges raw data and model-ready inputs. In this module, learners will explore techniques to clean, transform, and select features that improve model accuracy. By handling missing values, scaling data, encoding categories, and eliminating redundant features, you'll prepare datasets that enhance machine learning performance.

Topics	
Data Cleaning and Preparation	Handling missing data, Managing imbalanced datasets, Detecting and treating outliers
Feature Transformation	Feature scaling: normalization and standardization, Encoding categorical variables
Feature Selection	Backward and forward elimination, Correlation and covariance for feature relationships

Module 11

Exploratory Data Analysis (EDA)

This module focuses on uncovering insights from data using statistical and visual exploration. Through real-world datasets, learners will identify trends, detect anomalies, and validate hypotheses. By applying EDA, you'll gain a deeper understanding of data before modeling begins ensuring robust, insight-driven analysis.

Topics	
EDA Fundamentals	Purpose and process of exploratory analysis, Visualizing data patterns and relationships
Case Studies	Sentiment analysis of movie reviews, Analyzing wine quality and types, Forecasting stock and commodity prices

Module 12

Machine Learning Fundamentals

This module introduces the core supervised learning algorithms and evaluation techniques in machine learning. Learners will implement regression and classification models from scratch, understand performance metrics, and explore techniques like regularization and gradient descent. By the end, you'll be equipped to apply ML models to structured datasets confidently.

Topics	
Introduction to Machine Learning	Difference between AI, ML, DL, and Data Science, Types of learning: Supervised, Unsupervised, Semi-supervised, Reinforcement
Regression Analysis	Simple and Multiple Linear Regression with implementation, Evaluation metrics: MSE, MAE, RMSE, R^2, Linear Regression with Gradient Descent, Regularization methods: Ridge and Lasso
Classification Algorithms	Logistic Regression with implementation, Performance metrics: Confusion Matrix, Accuracy, Precision, Recall, F1/Fβ score
Support Vector Machines (SVM)	Concept and kernel functions, SVM classifier and regressor implementation
Other Supervised Algorithms	Naive Bayes classifier, K-Nearest Neighbors (KNN) classifier and regressor, Decision Tree classifier and regressor implementations

Module 13

Ensemble and Advanced ML Techniques

This module delves into ensemble and advanced learning techniques that enhance model accuracy and robustness. You'll learn how algorithms like Random Forests, Gradient Boosting, and XGBoost combine multiple models to improve predictions. The module concludes with an introduction to unsupervised learning through clustering.

Topics	
Ensemble Learning Concepts	Bagging and Boosting methods, Random Forest classifier and regressor with implementation, Out-of-Bag evaluation
Boosting Algorithms	Gradient Boosting classifier and regressor, XGBoost classifier and regressor with implementation
Unsupervised Learning	Introduction to clustering, K-Means clustering algorithm

Module 14

Natural Language Processing (NLP) for Machine Learning

This module introduces learners to Natural Language Processing the bridge between text data and machine learning. You'll learn how to clean, process, and represent text for modeling. Techniques like TF-IDF and Word2Vec will help you transform unstructured text into meaningful numerical features ready for ML applications.

Topics	
Introduction to NLP	Overview of NLP and its real-world use cases, Key terminology and learning roadmap
Text Preprocessing	Tokenization, Stemming, lemmatization, and stopword removal, Parts of Speech tagging using NLTK, Named Entity Recognition
Text Representation Techniques	One-hot encoding and N-grams (Bag of Words), TF-IDF intuition and implementation
Word Embeddings	Concept of word vectors, Introduction to Word2Vec and word embeddings for contextual understanding

Module 15

Deep Learning Foundations

This module introduces learners to the foundational principles of deep learning and neural networks. You'll explore how perceptrons evolve into complex architectures, understand the mathematics of backpropagation, and tackle training challenges like vanishing gradients. The module also covers essential components like activation functions, loss functions, and optimizers, using Keras and PyTorch frameworks for practical exposure.

Topics	
Introduction to Deep Learning	Overview and evolution of deep learning, Why deep learning has become popular
Perceptron and Artificial Neural Networks	Understanding perceptron intuition, Working of artificial neural networks, Forward and backward propagation, Chain rule of derivatives in neural networks
Training Challenges and Solutions	Vanishing and exploding gradient problems, Different activation functions and their roles, Types of loss functions and optimization algorithms
Regularization and Stability Techniques	Weight initialization methods, Dropout layers, Batch normalization
Deep Learning Frameworks	Introduction to Keras and PyTorch fundamentals, Visualizing neural network architecture and performance

Module 16

Advanced Deep Learning for NLP

This module explores advanced deep learning architectures for processing sequential and textual data. Learners will understand how RNNs, LSTMs, and GRUs capture dependencies in sequences, and how attention mechanisms and transformer architectures revolutionized NLP. By the end, you'll be ready to work with modern models like BERT and GPT for real-world NLP tasks.

Topics	
Sequential Models	Recurrent Neural Networks (RNNs), Understanding sequence modeling and context preservation
Memory-Based Architectures	Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRU) and their efficiency improvements
Attention Mechanisms	Concept of attention in neural networks, Encoders and decoders in sequence-to-sequence models, Self-attention and attention neural networks
Transformers and Modern Architectures	Introduction to transformers, Understanding BERT and GPT model architectures and applications

Module 17

Introduction to Generative AI

This module provides a foundational understanding of Generative AI the technology behind modern creative and intelligent systems. You'll learn the theory behind generative models, how they differ from discriminative ones, and explore recent innovations like text-to-image and large language models. Real-world applications highlight how these systems are shaping the next generation of AI.

Topics	
Fundamentals of Generative AI	What is Generative AI?, Why generative models are important, Understanding how generative models work
Generative vs Discriminative Models	Conceptual differences and use cases, Examples of generative vs discriminative tasks
Recent Advances and Applications	Latest research trends and breakthroughs, Key applications across industries, End-to-end generative AI project lifecycle

Module 18

Vector Databases for AI Applications

In this module, you'll dive into the data infrastructure powering Generative AI and LLMs. You'll understand how vector databases store embeddings, enable semantic search, and support retrieval-augmented systems. Hands-on exposure to popular tools like Chroma, Pinecone, and Faiss will prepare you to design and optimize vector-based AI applications.

Topics	
Introduction to Vector Databases	Understanding vector embeddings and similarity search, Comparison with SQL and NoSQL databases
Architecture and Capabilities	Data storage mechanisms and indexing, Types of vector databases: In-memory, on-disk, and cloud-based
Popular Vector Database Tools	Chroma DB, Faiss, Quadrant, Pinecone, LanceDB
Integrations and Applications	Vector search with NoSQL databases (MongoDB, Cassandra), Embedding-based retrieval for AI pipelines

Module 19

Retrieval-Augmented Generation (RAG)

This module focuses on RAG a key architecture that enhances LLM performance by integrating external data retrieval. You'll learn the complete RAG pipeline, combining vector search, language models, and frameworks like LangChain. By the end, you'll be able to design RAG systems that generate accurate, context-aware responses across text and multimodal domains.

Topics	
Introduction to RAG	Understanding Retrieval-Augmented Generation and its role in LLMs, The RAG pipeline overview
Implementing RAG Systems	Using LangChain, vector databases, and LLMs, Hybrid search and reranking techniques, Different retrieval methods for RAG
Advanced Concepts	Memory integration in RAG systems, Multimodal retrieval-augmented generation

Module 20

Comprehensive Guide to LangChain

LangChain is the backbone framework for building and managing LLM-powered applications. In this module, learners will explore how to connect data and APIs, design complex LLM workflows, and integrate tools like LangSmith and LangServe for monitoring and deployment. The hands-on experience will empower you to build scalable, production-ready AI applications.

Topics	
LangChain Foundations	Introduction to LangChain, Data connectors and API integrations
Building with LangChain	Chat models, tools, and toolkits, Prompt templating for context-aware LLMs
Chaining and Execution	LangChain Chains, LCEL, and Runnables, Synthetic data generation and memory management
Advanced Tools for Deployment	LangChain AI Agents, LangSmith for model monitoring, LangServe for model deployment

Module 21

Introduction to Agentic AI

This module introduces the emerging concept of Agentic AI autonomous systems that plan, reason, and collaborate. You'll learn how agentic frameworks differ from generative models, explore multi-agent architectures, and understand their growing role in intelligent automation and AI ecosystems.

Topics	
Understanding Agents	What are AI Agents?, Agentic AI vs traditional AI agents
Comparative Insights	Agentic AI vs Generative AI, Understanding multi-agent systems and collaboration
Frameworks and Ecosystem	Overview of Agentic AI frameworks, Applications and trends in multi-agent AI

Module 22

LangGraph Fundamentals

This module introduces the fundamentals of LangGraph, a framework for building AI workflow graphs that connect LLMs, tools, and agents. You'll learn how to create simple graphs, design chains and routers, and develop agents with memory. By the end, you'll understand how to deploy LangGraph applications and visualize workflow execution using LangGraph Studio.

Topics	
Getting Started with LangGraph	Introduction to LangGraph, Understanding Simple Graphs and Node Connections, LangGraph Studio Overview
Building and Managing Workflows	Chains and Routers in LangGraph, Creating and Managing Agents, Implementing Agents with Memory
Deployment Essentials	Introduction to Deployment and Execution in LangGraph

Module 23

State Management and Memory in LangGraph

In this module, you'll dive deep into LangGraph's state and memory system, learning how to structure data using schemas and reducers. You'll discover how multiple schemas can manage complex agent states and explore methods for trimming and filtering messages to optimize performance. This module builds the foundation for scalable and memory-efficient AI workflows.

Topics	
State Fundamentals	Introduction to LangGraph State System, Understanding State Schemas and Data Flow
Advanced State Handling	Creating and Managing State Reducers, Working with Multiple Schemas
Optimizing Memory	Trimming and Filtering Messages to Maintain Efficient Context

Module 24

UX & Human-in-the-Loop Systems with LangGraph

This module focuses on making AI systems interactive and human-centered using LangGraph. You'll learn how to enable real-time streaming, integrate human feedback at checkpoints, and utilize breakpoints for controlled execution. The "time travel" feature helps in debugging and analyzing past states making your AI workflows more explainable and testable.

Topics	
Interactive Workflow Design	Implementing Streaming for Real-Time Updates, Understanding and Managing Breakpoints
Human Feedback Integration	Editing State with Human Inputs, Dynamic Breakpoints and Controlled Execution
Debugging and Review	Time Travel in LangGraph for Workflow Rewind and Testing

Module 25

Agentic RAG Systems

In this module, you'll explore advanced RAG architectures enhanced with agentic intelligence. You'll implement adaptive, contextual (C-RAG), and self-retrieval (Self-RAG) systems both locally and with vector databases. This hands-on experience prepares you to design adaptive RAG systems capable of dynamic reasoning and context management for intelligent knowledge retrieval.

Topics	
Adaptive Retrieval-Augmented Generation (RAG)	Understanding Adaptive RAG and its Variants, Implementing Adaptive RAG with Cohere, Running Adaptive RAG Locally
Agentic RAG Architectures	Integrating Agents with RAG Pipelines, C-RAG (Contextual RAG) Concepts and Implementation
Self-RAG Implementations	Building Self-RAG Models, Integrating Self-RAG with Vector Databases, Deploying Self-RAG Locally

Module 26

Model Context Protocol (MCP)

This module covers the Model Context Protocol (MCP)a standard for seamless communication between AI models, tools, and development environments. You'll learn the internal workings of MCP, explore integrations with Claude and Cursor IDE, and build your own MCP servers using LangChain. By the end, you'll understand how MCP enables interoperability and modular design in LLM ecosystems.

Topics	
Understanding MCP	Introduction to Model Context Protocol, Core Components and Architecture
Inter-Component Communication	Data Flow and Communication between MCP Components, Integration with Tools like Claude Desktop and Cursor IDE
Building with MCP	Exploring Open MCP Repositories (e.g., Smithery.ai), Building MCP Servers with LangChain, Using Docker MCP Catalog and Toolkit

Module 27

Multi-Agent Systems for Research Automation

This capstone module teaches you how to build multi-agent systems that automate research and content generation. You'll design intelligent agents with specialized roles, integrate them through LangGraph workflows, and manage collaborative tasks. You'll also implement CI/CD pipelines using GitHub Actions and deploy agent systems via Docker and AWSbridging AI development with DevOps for full automation.

Topics	
Architecting Multi-Agent Systems	Introduction to Agentic AI and Multi-Agent Architectures, Designing Specialized Agents: Search, Reader, Analyst, Generator, Coordinator
Collaboration and Control	Managing Memory and Communication State, Prompt Engineering for Multi-Turn Collaboration, Integrating Human Feedback Checkpoints
Toolkits and Workflow Design	Connecting APIs like Arxiv, Search, and Paper Parsers, LangGraph-Structured Agent Workflows, Adding RAG for External Knowledge Integration
Deployment and Automation	FastAPI Backend for Multi-Agent Systems, UI for Logs, Graphs, and Reports, CI/CD Automation with GitHub Actions, Docker and AWS EC2 Deployment Pipeline

PROJECT

End-to-End AI & Data Projects

In this course, you'll gain hands-on experience in implementing end-to-end AI projects. You'll build end-to-end solutions spanning data engineering, machine learning, deep learning, and generative AI. Each project reinforces core concepts while giving you practical experience with CI/CD deployment, scalable pipelines, and production-ready model integration

Topics	
Python and Data Engineering Projects	End-to-End Review Scraper Project, Supabase and Streamlit Database Application
Machine Learning Projects	Network Intrusion Detection System (ML Project)
Deep Learning Projects	Text Summarization, Machine Translation, Question Answering
Generative AI & RAG Projects	RAG Q&A System with CI/CD Integration