

LIVE COURSE

Generative AI and LLMs

Learn to build real Generative AI and LLM applications, from prompt engineering and RAG to fine-tuning and deployment, through live mentor-led sessions and hands-on projects.

Industry-Led Course

Batch details updating soon

Technical



DURATION

4 Months

LEVEL

Beginner to Advanced

MODE

Live Online

LANGUAGE

English

BATCH STARTS

To be announced

ENROLMENT CLOSES

Open

COURSE FEE

₹6,999

₹20,000 - save 65% OFF



LIVE COURSE BROCHURE

Inside this brochure

Everything on the Generative AI and LLMs page, in one document. Tap a section to jump to it.

01		Course Overview	03
02		How Our Courses Work	04
03		Eligibility & Prerequisites	05
04		Skills You Will Gain	05
05		How we designed our Product?	06
06		What You Will Learn	08
07		Our Placement Support Services	34
08		Meet Our SME Board Members	36
09		Key Tools and Techniques	38
10		Projects You Will Work On	39
11		Generative AI and LLMs	41
12		World's First Ed-Tech with 8C Appraisal	42
13		Frequently Asked Questions	43
14		Ready to take the next step?	44

01 · AT A GLANCE



Course Overview

Learn to build real Generative AI and LLM applications, from prompt engineering and RAG to fine-tuning and deployment, through live mentor-led sessions and hands-on projects.

This live, mentor-led course takes you from Generative AI basics to advanced LLM engineering. You will practise prompt design, embeddings, RAG, agents, fine-tuning and deployment through weekly projects, code reviews and feedback, finishing with a job-ready portfolio of working AI applications.



DURATION
4 Months



LEVEL
Beginner to Advanced



MODE
Live Online



LANGUAGE
English



BATCH STARTS
To be announced



ENROLMENT CLOSES
Open



LEARNERS
Trusted by ambitious learners nationwide

COURSE FEE

₹6,999

₹20,000 - save 65% OFF

Highlights

Industry-Led Upskilling Courses

Learn from industry professionals with years of real-world experience.

Personalized Mentorship

Get one-on-one guidance tailored to your career goals and aspirations.

Premium Resources

Access course materials, job portal, free events, challenges and learner 8C Appraisal reports.

Vibrant Community

Connect with like-minded peers and build lasting professional relationships.



02 · SIMPLE 3-STEP JOURNEY

How Our Courses Work

A structured, step-by-step path designed to take you from foundational learning to career success with industry guidance.

01

Step 1: Enroll & Get Started

Choose your course and complete the enrollment process. Access your learning materials instantly and join our community of learners.

✓ Instant access to course materials

✓ Personalized learning dashboard

✓ Welcome orientation session

02

Step 2: Learn & Practice

Engage with interactive lessons, live sessions, and hands-on projects. Learn from industry experts and apply your knowledge in real-world scenarios.

✓ Live classes with expert instructors

✓ Hands-on practical projects

✓ One-on-one mentorship support

03

Step 3: Complete & Get Certified

Finish your coursework, pass assessments, and receive your professional certificate. Launch your new career with confidence and credentials.

✓ Industry-recognized certificate

✓ Portfolio of completed projects

✓ Career placement assistance

03 · BEFORE YOU ENROLL



Eligibility & Prerequisites

Check who this course is designed for and what you should know before your first class.

WHO CAN JOIN**Eligibility**

- ✓ Students from any degree stream
- ✓ Freshers looking for AI roles
- ✓ Working professionals moving into AI
- ✓ Developers wanting to build LLM apps

BEFORE YOU START**Prerequisites**

- ✓ Basic computer and internet skills
- ✓ Willingness to learn Python from scratch
- ✓ Comfort with basic school-level maths
- ✓ A laptop with a stable internet connection

04 · CAREER-READY SKILLS



Skills You Will Gain

Hands-on skills you will practise in live classes, assignments and projects throughout the course.

Generative AI

Large Language Models

Prompt Engineering

Python for AI

Transformers

Embeddings

Vector Databases

RAG Pipelines

LangChain

Fine-Tuning

LoRA and PEFT

AI Agents

Hugging Face

LLM Evaluation

Model Deployment

05 · PRODUCT ECOSYSTEM ARCHITECTURE



How we designed our Product?

An all-in-one ed-tech ecosystem carefully engineered to bridge learning, practical skill assessment, career guidance, and employment.

WORLD'S FIRST FEATURE

8C Appraisal Report

Proprietary evaluation matrix measuring holistic performance across 8 career pillars for authentic proof-of-work.

Communication

Configuration

Chronicle

Completeness

Cognizance

Conception

Comprehension

Case Study

Industry Validated · 100% Data Score

SUBJECT MATTER EXPERT AI

Insyrio AI

Get instant 24/7 doubt solving and concept clarifications directly from our specialized AI trained on your enrolled courses.

- ✓ 24/7 Course-Specific SME Assistant
- ✓ Instant Step-by-Step Doubt Resolution

Instant Learning Support

GAMIFIED REWARDS

iSure Coins

Use iSure Coins for AI career tools, masterclasses and CertiQ - and part-pay your course fee with them.

- ✓ Earn Perks & Rewards · Pay with Coins

Redeem Rewards Store

AI ENGINE

AI Career Diagnosis

Algorithmic analysis mapping strengths, gaps, and tailored pathways for industry roles.

- ✓ Skill Gap Mapping

Personalized Roadmap

CURRICULUM

Industry-Led Courses

Handcrafted learning modules built directly alongside senior engineering, design, and management leads.

Practical Frameworks

LIVE

Live Classes

Interactive classrooms with live chat, code walk-throughs, and real-time engagement.

Real-time Q&A Support

AUTO-SCAN

AI Resume ATS

Instant ATS optimization scanner ensuring your resume passes screening software effortlessly.

High Parser Score

PRACTICAL

Assignments

Industry-grade project briefs with detailed feedback and grading from experienced mentors.

Code & Work Reviews

WEEKLY

Challenges

Weekly hackathons, coding sprints, and design contests to test your problem-solving skills.

Leaderboard Ranking

NETWORKING

Events

Exclusive expert keynotes, career fairs, and interactive networking workshops.

Free Community Pass

1-ON-1

Live Mentorship

Direct one-on-one video guidance sessions with experienced industry practitioners.

Personalized Growth

TRACKING

Attendance

Class-by-class attendance tracking to ensure disciplined, consistent progression.

Attendance Reports

HIRING

Jobs

Direct hiring pipeline connecting top-performing learners directly with recruiter managers.

Exclusive Placements

FREE ACCESS

Free Books

Curated digital library filled with essential tech, management, and soft-skill books.

Instant Digital Library

ARTICLES

Blog

In-depth technical tutorials, industry trend analysis, and career growth strategies.

Updated Weekly

FREE ACCESS

Free Study Material

Downloadable cheat sheets, interview preparation guides, templates, and revision notes.

One-Click Downloads

06 · COURSE CURRICULUM



What You Will Learn

A carefully sequenced curriculum - every section breaks into modules and hands-on topics taught live by industry experts.

6

SECTIONS

36

MODULES

180

TOPICS

9

PROJECTS

01

SECTION 1

Foundations of Generative AI and Python Essentials

This section builds the conceptual base of generative AI and the Python, data and machine learning skills needed for the rest of the course.

6 modules · 30 topics

M1.1 Introduction to Generative AI · 5 topics

Learners understand what generative AI is, how it evolved and where it is applied across industries.

1**What Is Generative AI**

Explore the definition of generative AI and how it differs from traditional predictive AI. Review real products that use it every day.

AI CONCEPTS

TERMINOLOGY

USE CASES

2**History and Evolution of Generative Models**

Trace the path from rule-based systems and Markov chains to GANs, transformers and modern large language models.

AI HISTORY

MODEL TYPES

TRENDS

3**Types of Generative Models**

Compare text, image, audio, video and code generation models at a conceptual level. Understand where each type fits.

MODEL FAMILIES

MULTIMODAL AI

COMPARISON

4**Industry Applications of Generative AI**

Study how healthcare, marketing, education, finance and software teams use generative AI to create value.

BUSINESS USE CASES

INDUSTRY ANALYSIS

INNOVATION

5

Ethics Risks and Responsible AI

Discuss bias, hallucinations, copyright, privacy and safety concerns. Learn the principles of responsible AI use.

AI ETHICS

RISK AWARENESS

RESPONSIBLE AI

M1.2

Python Programming Basics · 5 topics

Learners set up a Python environment and master the core syntax required for AI development.

1

Setting Up Python and Jupyter Notebooks

Install Python, create virtual environments and run code in Jupyter and Google Colab.

PYTHON SETUP

JUPYTER

GOOGLE COLAB

2

Variables Data Types and Operators

Work with numbers, strings, booleans and operators to write simple programs and expressions.

PYTHON SYNTAX

DATA TYPES

OPERATORS

3

Control Flow with Conditions and Loops

Use if statements, for loops and while loops to control how programs make decisions and repeat tasks.

CONDITIONALS

LOOPS

LOGIC

4

Functions and Modules

Write reusable functions, use parameters and return values, and import standard and third party modules.

FUNCTIONS

MODULES

CODE REUSE

5

Working with Files and Handling Errors

Read and write text and JSON files, and handle exceptions so programs fail gracefully.

FILE HANDLING

JSON

EXCEPTIONS

M1.3

Python Data Structures and Object Oriented Programming · 5 topics

Learners organize data effectively and structure larger programs using classes and clean coding practices.

1

Lists and Tuples

Create, slice and modify lists and tuples, and choose between them for different tasks.

LISTS

TUPLES

SLICING

2

Dictionaries and Sets

Store key value data in dictionaries and remove duplicates with sets. Practice lookups and updates.

DICTIONARIES

SETS

DATA LOOKUP

3

Comprehensions and Iterators

Write concise list and dictionary comprehensions and understand iterators and generators.

COMPREHENSIONS

ITERATORS

PYTHONIC CODE

4

Classes and Objects

Define classes, attributes and methods, and understand inheritance and encapsulation through small examples.

OOP

CLASSES

INHERITANCE

5

Clean Code and Debugging Practices

Apply naming conventions, docstrings and debugging tools to write readable and maintainable code.

CLEAN CODE

DEBUGGING

DOCUMENTATION

M1.4 Data Handling with NumPy and Pandas · 5 topics

Learners gain the data manipulation and visualization skills that underpin every machine learning workflow.

1

NumPy Arrays and Vectorized Operations

Create arrays, apply vectorized math and understand shapes and broadcasting for fast numerical computing.

NUMPY

ARRAYS

BROADCASTING

2

Linear Algebra Essentials with NumPy

Practice vectors, matrices, dot products and matrix multiplication, which are the building blocks of neural networks.

LINEAR ALGEBRA

MATRICES

VECTORS

3

Pandas Series and DataFrames

Load data from CSV files, inspect it and select rows and columns using Pandas.

PANDAS

DATAFRAMES

DATA LOADING

4

Data Cleaning and Transformation

Handle missing values, fix data types, filter records, group data and merge datasets.

DATA CLEANING

GROUPBY

MERGING

5

Data Visualization with Matplotlib

Build line, bar and scatter plots to explore data and communicate findings clearly.

MATPLOTLIB

VISUALIZATION

EDA

M1.5 Machine Learning Fundamentals · 5 topics

Learners understand core machine learning ideas and train simple models before moving to deep learning.

1 Introduction to Machine Learning

Learn the differences between supervised, unsupervised and reinforcement learning along with the standard ML workflow.

ML BASICS

SUPERVISED LEARNING

ML WORKFLOW

2 Probability and Statistics for AI

Review distributions, mean, variance and conditional probability, and see why they matter for generative models.

PROBABILITY

STATISTICS

DISTRIBUTIONS

3 Regression and Classification with Scikit-learn

Train linear regression and logistic regression models and make predictions on real datasets.

SCIKIT-LEARN

REGRESSION

CLASSIFICATION

4 Model Evaluation and Overfitting

Split data into train and test sets, use metrics such as accuracy and RMSE, and recognize overfitting and underfitting.

EVALUATION

OVERFITTING

METRICS

5 Feature Engineering and Mini Project

Prepare features through encoding and scaling, then complete a small end to end prediction project.

FEATURE ENGINEERING

PREPROCESSING

MINI PROJECT

M1.6 Neural Networks and Deep Learning Primer · 5 topics

Learners grasp how neural networks learn and build a first model with a modern deep learning framework.

1 Neurons Layers and Activation Functions

Understand how artificial neurons combine inputs and how activation functions add non linearity.

NEURAL NETWORKS

ACTIVATIONS

ARCHITECTURE

2 Loss Functions and Gradient Descent

See how models measure error and improve through gradient descent and learning rates.

LOSS FUNCTIONS

GRADIENT DESCENT

OPTIMIZATION

3 Backpropagation Explained

Follow the intuition behind backpropagation and how gradients flow through a network during training.

BACKPROPAGATION

GRADIENTS

TRAINING

4

Building Models with PyTorch

Use tensors, datasets and training loops in PyTorch to build and train a simple neural network.

PYTORCH

TENSORS

TRAINING LOOP

5

Regularization and Training Best Practices

Apply dropout, early stopping and batch training, and learn to monitor training with loss curves.

REGULARIZATION

DROPOUT

MODEL TUNING

02

SECTION 2

Transformers and How LLMs Work

Learners progress from sequence modeling basics to the full transformer architecture, then explore how large language models are trained, scaled, and used to generate text.

6 modules · 30 topics

M2.1

From Text to Numbers · 5 topics

Learners understand how raw language is converted into numerical representations that neural networks can process.

1

Tokenization Fundamentals

Explore how text is split into tokens using word, character, and subword approaches. Practice tokenizing sentences and inspecting the results.

TOKENIZATION

TEXT PROCESSING

NLP BASICS

2

Byte Pair Encoding and Vocabulary Building

Learn how BPE merges frequent symbol pairs to build a subword vocabulary. Train a small tokenizer on a sample corpus.

BPE

VOCABULARY

TOKENIZER TRAINING

3

Word Embeddings Explained

Understand how words are mapped to dense vectors that capture meaning. Compare Word2Vec and GloVe style embeddings on example words.

EMBEDDINGS

WORD2VEC

VECTOR SPACES

4

Measuring Similarity in Embedding Space

Use cosine similarity and nearest neighbor search to compare vectors. Visualize embeddings with dimensionality reduction.

COSINE SIMILARITY

PCA

VISUALIZATION

5 Neural Network Refresher for Language Tasks

Review layers, activations, loss functions, and backpropagation in the context of text. Build a tiny next-word predictor in PyTorch.

PYTORCH

BACKPROPAGATION

LOSS FUNCTIONS

M2.2 Sequence Models and the Road to Attention · 5 topics

Learners see why earlier sequence models struggled and how attention emerged to solve those problems.

1 Recurrent Neural Networks Basics

Learn how RNNs process sequences step by step with a hidden state. Implement a simple character-level RNN.

RNN

HIDDEN STATE

SEQUENCE MODELING

2 LSTMs and GRUs

Study gating mechanisms that help models remember long-range information. Compare LSTM and GRU performance on a small task.

LSTM

GRU

GATING

3 Vanishing Gradients and Long-Range Dependencies

Understand why recurrent models forget distant context. Observe gradient behavior through simple experiments.

VANISHING GRADIENTS

DEBUGGING

TRAINING DYNAMICS

4 Encoder-Decoder Models and Seq2Seq

Explore how encoder-decoder designs handle translation and summarization. Identify the bottleneck of a single context vector.

SEQ2SEQ

TRANSLATION

ENCODER-DECODER

5 The Birth of Attention

Learn how attention lets a decoder focus on relevant input positions. Walk through alignment weights on a translation example.

ATTENTION

ALIGNMENT

CONTEXT

M2.3 Inside the Transformer Architecture · 5 topics

Learners dissect each component of the transformer and implement the core mechanisms themselves.

1 Self-Attention with Queries Keys and Values

Break down how queries, keys, and values produce context-aware representations. Compute attention by hand on a short sentence.

SELF-ATTENTION

QKV

LINEAR ALGEBRA

2 Scaled Dot-Product and Multi-Head Attention

Understand scaling, softmax, and why multiple heads capture different relationships. Implement multi-head attention in code.

MULTI-HEAD ATTENTION

SOFTMAX

IMPLEMENTATION

3 Positional Encoding

Learn how transformers inject word order using sinusoidal and learned positions. Compare absolute and rotary approaches.

POSITIONAL ENCODING

ROPE

SEQUENCE ORDER

4 Feed-Forward Layers Residuals and Layer Normalization

Examine the supporting blocks that stabilize and deepen transformer training. See how they combine into a full layer.

RESIDUAL CONNECTIONS

LAYERNORM

FEED-FORWARD

5 Assembling a Mini Transformer

Combine all components into a small working transformer block stack. Run a forward pass and inspect tensor shapes.

MODEL BUILDING

PYTORCH

TENSOR SHAPES

M2.4 Transformer Families and Model Types · 5 topics

Learners compare encoder-only, decoder-only, and encoder-decoder models and when each is used.

1 Encoder-Only Models and BERT

Understand masked language modeling and bidirectional context. Use BERT for classification and embeddings.

BERT

MASKED LM

CLASSIFICATION

2 Decoder-Only Models and the GPT Family

Learn autoregressive generation and causal masking. Trace how GPT models predict one token at a time.

GPT

CAUSAL MASKING

AUTOREGRESSIVE

3 Encoder-Decoder Models like T5 and BART

Explore text-to-text frameworks for translation and summarization. Compare their strengths with decoder-only models.

T5

BART

TEXT-TO-TEXT

4 Working with Hugging Face Transformers

Load pretrained models and tokenizers using the Hugging Face library. Run inference pipelines for several tasks.

HUGGING FACE

PIPELINES

PRETRAINED MODELS

5 Choosing the Right Model for a Task

Evaluate model size, speed, cost, and licensing against project needs. Build a simple decision checklist.

MODEL SELECTION

TRADE-OFFS

EVALUATION

M2.5 How LLMs Are Trained · 5 topics

Learners follow the full training lifecycle from pretraining data to alignment with human preferences.

1 Pretraining Objectives and Data Sources

Learn how next-token prediction on massive corpora builds general language ability. Discuss data quality, filtering, and bias.

PRETRAINING

DATA CURATION

BIAS

2 Scaling Laws and Compute

Understand how model size, data, and compute relate to performance. Interpret scaling curves from published research.

SCALING LAWS

COMPUTE

RESEARCH READING

3 Distributed Training and Optimization

Get an overview of GPUs, data and model parallelism, and optimizers like Adam. Learn why mixed precision matters.

DISTRIBUTED TRAINING

ADAM

MIXED PRECISION

4 Instruction Tuning and Supervised Fine-Tuning

See how models are adapted to follow instructions using curated examples. Prepare a small instruction dataset.

INSTRUCTION TUNING

SFT

DATASETS

5 Alignment with RLHF and Preference Optimization

Learn how human feedback shapes helpful and safe behavior. Compare RLHF and Direct Preference Optimization at a conceptual level.

RLHF

DPO

ALIGNMENT

M2.6 Generating Text and Understanding Model Behavior · 5 topics

Learners control how LLMs generate output and learn to reason about their capabilities and limits.

1 Decoding Strategies

Compare greedy search, beam search, top-k, and nucleus sampling. Generate text and observe how each strategy changes output.

DECODING

TOP-P SAMPLING

BEAM SEARCH

2

Temperature and Generation Parameters

Experiment with temperature, repetition penalties, and max length. Learn how to tune for creativity versus accuracy.

TEMPERATURE

PARAMETER TUNING

EXPERIMENTATION

3

Context Windows and the KV Cache

Understand context length limits and how caching speeds up inference. Discuss memory and latency trade-offs.

CONTEXT WINDOW

KV CACHE

INFERENCE

4

Hallucinations and Failure Modes

Identify why LLMs produce confident but wrong answers. Practice spotting and testing for common failure patterns.

HALLUCINATION

RELIABILITY

TESTING

5

Evaluating LLM Outputs

Learn metrics such as perplexity and benchmark scores alongside human evaluation methods. Design a small evaluation set for a use case.

PERPLEXITY

BENCHMARKS

EVALUATION DESIGN

03

SECTION 3

Prompt Engineering and LLM APIs

Learners master prompt design techniques and build practical applications by integrating hosted and open LLM APIs into Python programs.

6 modules · 30 topics

M3.1

Prompt Design Fundamentals · 5 topics

Learners understand how prompts shape model behavior and write clear, reliable instructions.

1

Anatomy of an Effective Prompt

Explore the components of a prompt including role, task, context, constraints and output format. Practice rewriting vague prompts into precise ones.

PROMPT STRUCTURE

CLARITY

INSTRUCTION DESIGN

2

Roles, System Messages and Personas

Learn how system, user and assistant messages steer tone and behavior. Design personas for support, tutoring and analysis use cases.

SYSTEM PROMPTS

PERSONAS

CHAT ROLES

3

Context and Constraint Setting

Supply background information and limits such as length, style and audience to control responses. Compare outputs with and without constraints.

CONTEXT CONTROL

CONSTRAINTS

OUTPUT QUALITY

4

Controlling Output Format

Request tables, lists, markdown and structured text reliably. Use delimiters and templates to keep formats consistent.

FORMATTING

TEMPLATES

DELIMITERS

5

Sampling Parameters in Practice

Experiment with temperature, top-p, max tokens and stop sequences to see their effect on creativity and determinism.

TEMPERATURE

TOP-P

TOKEN LIMITS

M3.2

Advanced Prompting Techniques · 5 topics

Learners apply proven strategies that improve reasoning and accuracy on complex tasks.

1

Zero-Shot and Few-Shot Prompting

Decide when to give examples and how to select them. Measure how example quality and order change results.

FEW-SHOT

EXAMPLE SELECTION

IN-CONTEXT LEARNING

2

Chain-of-Thought Reasoning

Elicit step-by-step reasoning for math, logic and planning problems. Understand when explicit reasoning helps and when it adds noise.

CHAIN-OF-THOUGHT

REASONING

PROBLEM SOLVING

3

Self-Consistency and Prompt Chaining

Break big tasks into linked prompts and aggregate multiple samples for better answers.

PROMPT CHAINING

SELF-CONSISTENCY

TASK DECOMPOSITION

4

Reflection and Self-Critique Prompts

Ask models to review and improve their own drafts. Build critique-and-revise loops for writing and code.

SELF-CRITIQUE

ITERATION

REFINEMENT

5

Prompt Patterns Library

Study reusable patterns such as persona, template, question refinement and cognitive verifier. Build a personal pattern library.

PROMPT PATTERNS

REUSABILITY

BEST PRACTICES

M3.3 Working with LLM APIs · 5 topics

Learners call hosted model endpoints from Python and handle real-world API concerns.

1 API Basics and Authentication

Understand REST endpoints, API keys and secure secret handling. Make a first chat completion request from Python.

REST APIS

API KEYS

PYTHON REQUESTS

2 Chat Completion Requests and Responses

Build message lists, read response objects and manage multi-turn conversation history.

CHAT COMPLETIONS

MESSAGE HISTORY

RESPONSE PARSING

3 Streaming Responses and Async Calls

Stream tokens for responsive interfaces and use asynchronous calls to run many requests efficiently.

STREAMING

ASYNCIO

LATENCY

4 Error Handling, Retries and Rate Limits

Handle timeouts, rate limit errors and malformed responses with exponential backoff and logging.

ERROR HANDLING

RETRIES

RATE LIMITING

5 Cost, Tokens and Model Selection

Count tokens, estimate spend and choose between models based on quality, speed and price.

TOKENIZATION

COST CONTROL

MODEL SELECTION

M3.4 Structured Outputs and Function Calling · 5 topics

Learners turn free-form model text into dependable structured data and tool actions.

1 Generating JSON Reliably

Prompt for strict JSON and apply JSON mode. Parse and repair outputs in Python.

JSON

PARSING

OUTPUT RELIABILITY

2 Schema Validation with Pydantic

Define data schemas and validate model outputs automatically. Retry on validation failure.

PYDANTIC

VALIDATION

SCHEMAS

3 Function Calling and Tool Use

Describe functions to the model and execute the calls it requests. Return results to complete the conversation loop.

FUNCTION CALLING

TOOL USE

AGENTS BASICS

4 Information Extraction and Classification

Build pipelines that extract entities and classify text such as resumes, reviews and tickets.

EXTRACTION

CLASSIFICATION

NLP PIPELINES

5 Multi-Step Workflows with LLMs

Combine structured calls and tools into a small automated workflow with branching logic.

WORKFLOWS

ORCHESTRATION

AUTOMATION

M3.5 Open Source Models and Local Inference · 5 topics

Learners run and call open models using community tooling and compare them with hosted services.

1 Hugging Face Hub and Model Cards

Navigate the Hub, read model cards and licenses, and pick suitable models for a task.

HUGGING FACE

MODEL CARDS

LICENSING

2 Text Generation with Transformers Library

Load a pretrained model and generate text using pipelines and generate settings.

TRANSFORMERS LIBRARY

PIPELINES

INFERENCE

3 Running Models Locally with Ollama

Install and serve quantized models locally and call them through a local API.

OLLAMA

LOCAL LLMS

QUANTIZATION

4 Prompt Formats and Chat Templates

Learn how different model families expect instructions and apply chat templates correctly.

CHAT TEMPLATES

MODEL FAMILIES

INSTRUCTION TUNING

5 Hosted Versus Open Models Tradeoffs

Compare quality, privacy, latency, cost and maintenance to make informed deployment choices.

TRADEOFF ANALYSIS

PRIVACY

DEPLOYMENT CHOICES

M3.6 Prompt Evaluation, Safety and Capstone Application · 5 topics

Learners test prompts systematically, defend against misuse and ship a small prompt-driven application.

1 Building Prompt Test Sets

Create representative test cases and expected outputs to measure prompt performance objectively.

TEST SETS

EVALUATION

BENCHMARKING

2

Automated Evaluation and LLM-as-Judge

Score outputs using metrics and model-based graders, and understand their biases and limits.

LLM-AS-JUDGE

METRICS

SCORING

3

Prompt Injection and Jailbreak Defense

Recognize injection attacks and apply input filtering, instruction hierarchy and output checks.

PROMPT INJECTION

SECURITY

GUARDRAILS

4

Reducing Hallucinations and Bias

Ground answers in provided sources, request citations and test for biased or unsafe outputs.

HALLUCINATION CONTROL

GROUNDING

RESPONSIBLE AI

5

Capstone Build: Prompt-Driven Assistant App

Assemble prompts, API calls, structured outputs and evaluation into a working command line or web assistant and present it.

PROJECT BUILDING

INTEGRATION

PRESENTATION

04

SECTION 4

Embeddings, Vector Databases and RAG

Learners master text embeddings, vector search and retrieval-augmented generation to build grounded, production-ready LLM applications over private data.

6 modules · 30 topics

M4.1

Understanding Text Embeddings · 5 topics

This module explains how text is converted into dense vectors and how those vectors capture meaning.

1

From Words to Vectors

Trace the evolution from one-hot encoding and bag of words to dense representations. Understand why vector space captures semantic meaning.

VECTOR REPRESENTATIONS

NLP BASICS

SEMANTIC MEANING

2

Word and Sentence Embeddings

Compare Word2Vec and GloVe with modern sentence-level embedding models. Generate embeddings for sample text in Python.

WORD2VEC

SENTENCE TRANSFORMERS

PYTHON

3 Embedding Models from Providers

Use hosted embedding APIs and open-source models from Hugging Face. Compare dimensions, context limits, cost and quality.

EMBEDDING APIS

HUGGING FACE

MODEL SELECTION

4 Similarity Metrics

Learn cosine similarity, dot product and Euclidean distance and when each one fits. Compute and compare them with NumPy.

COSINE SIMILARITY

NUMPY

DISTANCE METRICS

5 Visualizing and Evaluating Embeddings

Project embeddings into two dimensions with PCA and t-SNE to inspect clusters. Evaluate embedding quality on small benchmark tasks.

PCA

T-SNE

EMBEDDING EVALUATION

M4.2 Vector Search Fundamentals · 5 topics

This module covers the algorithms that make searching millions of vectors fast and accurate.

1 Exact Versus Approximate Search

Understand brute-force nearest neighbor search and why it fails at scale. Explore the speed and recall trade-off of approximate methods.

NEAREST NEIGHBOR

RECALL

SCALABILITY

2 Indexing with HNSW

Learn how hierarchical navigable small world graphs organize vectors for fast lookup. Tune key parameters such as M and ef.

HNSW

GRAPH INDEXES

PARAMETER TUNING

3 IVF and Product Quantization

Study inverted file indexes and vector compression to reduce memory use. Weigh accuracy loss against storage savings.

IVF

QUANTIZATION

MEMORY OPTIMIZATION

4 Hands-on with FAISS

Build and query indexes with the FAISS library on a real text dataset. Benchmark different index types for latency and recall.

FAISS

BENCHMARKING

PYTHON

5

Hybrid and Keyword Search

Combine BM25 keyword search with vector search for better results. Merge ranked lists using reciprocal rank fusion.

BM25

HYBRID SEARCH

RANK FUSION

M4.3

Vector Databases in Practice · 5 topics

This module introduces production vector databases and how to design, load and query them.

1

Vector Database Landscape

Survey Chroma, Pinecone, Weaviate, Qdrant, Milvus and pgvector. Compare hosting models, features and typical use cases.

CHROMA

PINECONE

TOOL COMPARISON

2

Collections, Schemas and Metadata

Design collections that store vectors alongside documents and metadata. Plan identifiers and fields for later filtering.

SCHEMA DESIGN

METADATA

DATA MODELING

3

Ingesting and Updating Data

Batch insert, upsert and delete vectors while keeping the index consistent. Handle re-embedding when documents change.

UPSERTS

BATCH INGESTION

DATA FRESHNESS

4

Filtered and Namespaced Queries

Run similarity queries with metadata filters and multi-tenant namespaces. Understand how filtering affects performance.

METADATA FILTERING

MULTI-TENANCY

QUERYING

5

Postgres with pgvector

Add vector search to a relational database using pgvector. Combine SQL joins with similarity queries in one system.

PGVECTOR

SQL

POSTGRES SQL

M4.4

Document Processing and Chunking · 5 topics

This module teaches how to turn raw documents into high-quality retrievable chunks.

1

Loading Documents from Many Sources

Extract text from PDFs, web pages, Word files and databases using document loaders. Clean noisy text and handle tables and layouts.

DOCUMENT LOADERS

PDF PARSING

DATA CLEANING

2

Chunking Strategies

Compare fixed-size, recursive, sentence and semantic chunking. Learn how chunk size and overlap affect retrieval quality.

CHUNKING

TEXT SPLITTING

OVERLAP TUNING

3

Structure-Aware Chunking

Use headings, markdown and code structure to produce coherent chunks. Preserve context with parent-child and hierarchical chunks.

HIERARCHICAL CHUNKS

MARKDOWN PARSING

CONTEXT PRESERVATION

4

Metadata Enrichment

Attach titles, sources, dates and generated summaries to each chunk. Use this metadata to improve filtering and citations.

METADATA ENRICHMENT

CITATIONS

SUMMARIZATION

5

Building an Ingestion Pipeline

Assemble loading, chunking, embedding and storage into a repeatable pipeline. Add logging and error handling for reliability.

DATA PIPELINES

AUTOMATION

ERROR HANDLING

M4.5

Building RAG Systems · 5 topics

This module shows how to connect retrieval with LLMs to produce grounded and cited answers.

1

RAG Architecture Overview

Understand the retrieve, augment and generate flow and why RAG reduces hallucination. Compare RAG with fine-tuning and long-context prompting.

RAG ARCHITECTURE

HALLUCINATION REDUCTION

DESIGN TRADE-OFFS

2

Building a Basic RAG Pipeline

Implement an end-to-end question answering app over your own documents. Use LangChain or LlamaIndex to wire the components together.

LANGCHAIN

LLAMAINDEX

QUESTION ANSWERING

3

Prompting for Grounded Answers

Write prompts that force the model to use retrieved context and admit when it does not know. Return answers with source citations.

GROUNDING

PROMPT DESIGN

CITATIONS

4

Conversational RAG with Memory

Add chat history and question rewriting so follow-up queries retrieve correctly. Manage context window limits across turns.

CHAT MEMORY

QUERY REWRITING

CONTEXT MANAGEMENT

5

Building a RAG User Interface

Wrap the pipeline in a simple Streamlit or Gradio app with streaming responses. Display retrieved sources alongside answers.

STREAMLIT

GRADIO

STREAMING

M4.6

Advanced RAG and Evaluation · 5 topics

This module covers techniques to improve retrieval quality and methods to measure and monitor RAG performance.

1

Query Transformation Techniques

Apply query expansion, multi-query generation and HyDE to improve recall. Decompose complex questions into sub-queries.

HYDE

MULTI-QUERY

QUERY DECOMPOSITION

2

Reranking and Context Compression

Use cross-encoder rerankers to reorder results and compress context to the most relevant passages. Reduce token cost without losing accuracy.

RERANKING

CROSS-ENCODERS

CONTEXT COMPRESSION

3

Agentic and Graph-Based RAG

Explore agents that decide when and what to retrieve and knowledge graph approaches to RAG. Understand when added complexity pays off.

AGENTIC RAG

KNOWLEDGE GRAPHS

ROUTING

4

Evaluating RAG Quality

Measure retrieval with recall and MRR and generation with faithfulness and relevance. Use tools such as RAGAS and LLM-as-judge methods.

RAGAS

EVALUATION METRICS

LLM AS JUDGE

5

Production RAG Considerations

Address latency, caching, cost, security and access control for deployed systems. Monitor failures and set up feedback loops for continuous improvement.

CACHING

OBSERVABILITY

ACCESS CONTROL

COST OPTIMIZATION

05

SECTION 5

Fine-Tuning, Agents and Evaluation

Learners adapt open models to custom tasks, build tool-using agents, and measure and safeguard LLM applications before deployment.

6 modules · 30 topics

M5.1 Fine-Tuning Fundamentals · 5 topics

Understand when and how to adapt a pretrained model instead of relying on prompting or retrieval alone.

1 Prompting vs RAG vs Fine-Tuning

Learn a decision framework for choosing the right adaptation strategy based on cost, data, and task needs. Review real examples where each approach wins.

DECISION MAKING

MODEL ADAPTATION

COST ANALYSIS

2 Preparing Datasets for Fine-Tuning

Collect, clean, and format instruction and chat datasets in JSONL. Cover deduplication, quality filtering, and train and validation splits.

DATA CLEANING

JSONL

DATASET DESIGN

3 Supervised Fine-Tuning Workflow

Walk through the full SFT pipeline from loading a base model to training and saving checkpoints. Use Hugging Face Transformers and the Trainer API.

HUGGING FACE

SFT

TRAINING LOOPS

4 Hyperparameters and Training Diagnostics

Tune learning rate, batch size, and epochs while reading loss curves. Recognize overfitting, underfitting, and catastrophic forgetting.

HYPERPARAMETERS

LOSS CURVES

OVERFITTING

5 Fine-Tuning Hosted Models via API

Fine-tune a hosted model using a provider API and compare it to open-source options. Discuss pricing, limits, and data privacy.

FINE-TUNING APIS

PRIVACY

VENDOR COMPARISON

M5.2 Parameter-Efficient Fine-Tuning · 5 topics

Adapt large models on modest hardware using LoRA, quantization, and related techniques.

1 Why Parameter-Efficient Methods Matter

Explore the memory and compute limits of full fine-tuning and how adapters reduce them. Survey the PEFT landscape.

PEFT

GPU MEMORY

EFFICIENCY

2 LoRA Explained and Implemented

Understand low-rank adaptation mathematically and intuitively, then apply LoRA adapters with the PEFT library. Choose rank and alpha values sensibly.

LORA

PEFT LIBRARY

LOW-RANK MATH

3 Quantization and QLoRA

Load models in 4-bit and 8-bit precision and fine-tune them with QLoRA on a single GPU. Examine the quality and speed tradeoffs.

QUANTIZATION

QLORA

BITSANDBYTES

4 Merging, Saving and Serving Adapters

Merge adapters into base weights, publish models to the Hub, and serve multiple adapters efficiently. Cover versioning practices.

MODEL MERGING

MODEL HUB

SERVING

5 Hands-On Lab: Domain-Specific Assistant

Fine-tune a small open model on a domain dataset such as support or legal text. Compare its outputs with the base model.

APPLIED LAB

DOMAIN ADAPTATION

COMPARISON

M5.3 Alignment and Preference Tuning · 5 topics

Shape model behavior using human feedback and modern preference optimization methods.

1 Instruction Tuning and Chat Templates

See how raw models become helpful assistants through instruction data and chat formatting. Avoid template mismatches that degrade output.

INSTRUCTION TUNING

CHAT TEMPLATES

FORMATTING

2 RLHF Concepts

Learn the stages of reinforcement learning from human feedback including reward models and policy optimization.

Understand its strengths and challenges.

RLHF

REWARD MODELS

PPO BASICS

3

Direct Preference Optimization

Train with chosen and rejected response pairs using DPO as a simpler alternative to RLHF. Build a small preference dataset.

DPO

PREFERENCE DATA

TRL

4

Synthetic Data Generation

Use strong models to generate and filter training data for smaller models. Discuss quality control, licensing, and bias amplification.

SYNTHETIC DATA

DISTILLATION

DATA QUALITY

5

Safety Alignment and Red Teaming

Probe models for harmful outputs, jailbreaks, and prompt injection. Apply mitigation strategies and guardrail layers.

RED TEAMING

GUARDRAILS

SAFETY

M5.4 Building LLM Agents · 5 topics

Design systems where models plan, call tools, and act toward goals.

1

Agent Concepts and the ReAct Pattern

Understand the reasoning and acting loop that underpins most agents. Compare agents with fixed chains and workflows.

REACT

AGENT DESIGN

REASONING LOOPS

2

Function Calling and Tool Use

Define tool schemas, let models choose functions, and handle results and errors. Build agents that call search, calculators, and APIs.

FUNCTION CALLING

TOOL SCHEMAS

API INTEGRATION

3

Agent Memory and Planning

Implement short-term and long-term memory and task decomposition strategies. Explore planning and reflection techniques.

MEMORY

PLANNING

REFLECTION

4

Agent Frameworks with LangGraph

Build stateful agent workflows as graphs with branching, loops, and human approval steps. Compare with other frameworks.

LANGGRAPH

STATE MACHINES

ORCHESTRATION

5 Multi-Agent Systems and Model Context Protocol

Coordinate specialized agents and connect tools through the Model Context Protocol. Discuss when multi-agent designs are worth the complexity.

MULTI-AGENT

MCP

COLLABORATION PATTERNS

M5.5 Evaluating LLM Systems · 5 topics

Measure quality systematically using metrics, benchmarks, and automated judges.

1 Evaluation Challenges and Benchmarks

Learn why evaluating generative output is hard and review benchmarks such as MMLU and HumanEval. Understand contamination and benchmark limits.

BENCHMARKS

MMLU

EVALUATION THEORY

2 Automatic Metrics for Text Quality

Apply BLEU, ROUGE, and BERTScore and understand when they mislead. Choose metrics that fit each task.

ROUGE

BERTSCORE

METRIC SELECTION

3 LLM-as-a-Judge Techniques

Use models to grade outputs with rubrics and pairwise comparison. Detect and reduce judge bias.

LLM JUDGES

RUBRICS

BIAS MITIGATION

4 Evaluating RAG and Agents

Measure retrieval quality, faithfulness, and answer relevance with tools like Ragas. Trace agent trajectories and tool-call accuracy.

RAGAS

FAITHFULNESS

AGENT EVALUATION

5 Building Test Sets and Regression Suites

Create golden datasets and automated evaluation pipelines that run on every change. Track results over time.

GOLDEN DATASETS

REGRESSION TESTING

CI PIPELINES

M5.6 Production, Monitoring and Responsible AI · 5 topics

Ship LLM applications reliably while managing cost, risk, and ethics.

1 Deploying and Serving Models

Serve models with vLLM and managed endpoints while balancing latency and throughput. Cover batching, streaming, and caching.

VLLM

INFERENCE SERVING

LATENCY

2

Observability and Tracing

Log prompts, responses, latency, and token usage using tracing tools such as LangSmith or Langfuse. Debug failures in production.

TRACING

LANGFUSE

MONITORING

3

Cost and Performance Optimization

Reduce spend through model routing, prompt compression, caching, and smaller models. Set budgets and alerts.

COST CONTROL

MODEL ROUTING

CACHING

4

Security, Privacy and Governance

Address prompt injection, data leakage, PII handling, and compliance requirements. Establish review and approval processes.

PROMPT INJECTION

PII

GOVERNANCE

5

Responsible AI and Capstone Planning

Examine bias, transparency, and societal impact, then scope a capstone project that combines fine-tuning, agents, and evaluation.

ETHICS

FAIRNESS

PROJECT PLANNING

06

SECTION 6

Deployment, Safety and Capstone Project

Learners take LLM applications from prototype to production by mastering serving, monitoring, safety and governance, then apply everything in a capstone project and prepare for their careers.

6 modules · 30 topics

M6.1

Serving and Deploying LLM Applications · 5 topics

Learners package LLM applications as APIs and deploy them reliably to cloud platforms.

1

Building LLM Backends with FastAPI

Wrap prompts, RAG pipelines and model calls inside a clean REST API. Cover request validation, streaming responses and error handling.

FASTAPI

REST APIS

STREAMING

2

Containerizing with Docker

Package an LLM service with its dependencies into a portable Docker image. Learn environment variables, secrets handling and image optimization.

DOCKER

CONTAINERS

SECRETS MANAGEMENT

3 Deploying to the Cloud

Deploy a containerized LLM app to a managed cloud service and expose it securely. Compare serverless, container and VM options.

CLOUD DEPLOYMENT

AWS

HOSTING

4 Hosting Open Source Models

Serve open source models using tools such as vLLM and Ollama. Understand GPU requirements, batching and throughput tradeoffs.

VLLM

OLLAMA

GPU SERVING

5 Building User Interfaces with Streamlit and Gradio

Create quick interactive front ends for chat and document apps. Connect the UI to the deployed backend.

STREAMLIT

GRADIO

PROTOTYPING

M6.2 Optimization, Cost and Performance · 5 topics

Learners reduce latency and spending while keeping output quality high.

1 Token Economics and Cost Control

Estimate and track the cost of LLM calls across providers. Apply budgets, limits and model selection strategies.

TOKEN PRICING

COST TRACKING

BUDGETING

2 Caching and Request Batching

Use semantic caching, prompt caching and batching to cut repeated work. Measure savings against freshness requirements.

CACHING

BATCHING

LATENCY

3 Quantization and Model Compression

Shrink models with 8-bit and 4-bit quantization and distillation ideas. Evaluate the quality and speed tradeoff.

QUANTIZATION

GGUF

DISTILLATION

4 Model Routing and Fallbacks

Route simple queries to small models and hard ones to larger models. Add retries and fallbacks for provider outages.

ROUTING

FALLBACKS

RESILIENCE

5 Scaling and Load Testing

Simulate traffic against an LLM service and find bottlenecks. Apply rate limiting, queues and autoscaling concepts.

LOAD TESTING

RATE LIMITING

AUTOSCALING

M6.3 LLMOps, Monitoring and Observability · 5 topics

Learners operate LLM systems in production with tracing, monitoring and continuous improvement.

1 Introduction to LLMOps

Understand the lifecycle of an LLM product from experiment to production. Review versioning of prompts, models and datasets.

LLMOPS

VERSIONING

LIFECYCLE

2 Tracing and Logging with LangSmith and Langfuse

Capture traces of chains, retrievals and tool calls to debug behavior. Inspect latency and token usage per step.

TRACING

LANGSMITH

LANGFUSE

3 Monitoring Quality and Drift

Track output quality, user feedback and retrieval health over time. Detect drift and set up alerts.

MONITORING

DRIFT DETECTION

ALERTS

4 CI/CD for LLM Applications

Automate tests, evaluations and deployments using GitHub Actions. Gate releases on evaluation scores.

CI/CD

GITHUB ACTIONS

REGRESSION TESTING

5 Feedback Loops and Continuous Improvement

Collect user feedback and turn it into better prompts, data and evaluations. Plan iteration cycles for a live product.

USER FEEDBACK

ITERATION

A/B TESTING

M6.4 Safety, Security and Guardrails · 5 topics

Learners defend LLM applications against misuse, attacks and harmful outputs.

1 LLM Threat Landscape and OWASP Top 10

Survey the main risks to LLM apps including injection, data leakage and excessive agency. Map each risk to a mitigation.

OWASP LLM

THREAT MODELING

RISK ANALYSIS

2 Prompt Injection and Jailbreak Defense

Practice attacking and hardening prompts and agents. Apply input isolation, output checks and least privilege for tools.

PROMPT INJECTION

RED TEAMING

HARDENING

3 Guardrails and Content Moderation

Add input and output filters using moderation APIs and guardrail libraries. Enforce topic limits and structured output rules.

GUARDRAILS

MODERATION

OUTPUT VALIDATION

4

Hallucination Mitigation and Grounding

Reduce fabricated answers with grounding, citations and confidence checks. Design graceful refusal behavior.

HALLUCINATIONS

GROUNDING

CITATIONS

5

Privacy and PII Protection

Detect and redact personal data before it reaches models or logs. Review data retention and consent practices.

PII REDACTION

DATA PRIVACY

COMPLIANCE

M6.5

Responsible AI, Ethics and Governance · 5 topics

Learners apply ethical principles, legal awareness and governance practices to AI products.

1

Bias and Fairness in Generative Models

Identify sources of bias in training data and outputs. Test applications with fairness probes and mitigation methods.

BIAS TESTING

FAIRNESS

MITIGATION

2

Copyright, Licensing and Data Rights

Understand open model licenses, dataset provenance and content ownership questions. Learn practical steps to reduce legal exposure.

LICENSING

COPYRIGHT

DATA PROVENANCE

3

AI Regulation and Governance Frameworks

Review the EU AI Act, NIST AI RMF and organizational policy basics. Translate requirements into project checklists.

EU AI ACT

NIST AI RMF

GOVERNANCE

4

Transparency, Model Cards and Documentation

Document model behavior, limits and intended use with model cards and system cards. Communicate risk clearly to stakeholders.

MODEL CARDS

DOCUMENTATION

TRANSPARENCY

5

Human Oversight and Safe Product Design

Design human in the loop workflows and escalation paths for high stakes use cases. Balance automation with accountability.

HUMAN IN THE LOOP

PRODUCT DESIGN

ACCOUNTABILITY

M6.6 Capstone Project and Career Launch · 5 topics

Learners build, present and showcase an end to end GenAI application and prepare for job opportunities.

1 Capstone Planning and Architecture

Choose a real world problem and define scope, success metrics and system design. Select models, data and tools for the build.

PROJECT SCOPING

SYSTEM DESIGN

PLANNING

2 Capstone Build Sprint: Core Pipeline

Implement the main RAG or agent pipeline with prompts, retrieval and tool use. Receive mentor feedback on code and design.

RAG

AGENTS

IMPLEMENTATION

3 Capstone Build Sprint: Deploy, Evaluate and Secure

Deploy the app, add evaluations, monitoring and guardrails, and fix issues found in testing. Prepare it for public demonstration.

DEPLOYMENT

EVALUATION

GUARDRAILS

4 Demo Day and Technical Presentation

Present the capstone to peers and mentors with a live demo and architecture walkthrough. Practice explaining tradeoffs and results.

PRESENTATION

DEMOING

TECHNICAL COMMUNICATION

5 Portfolio, Resume and Interview Preparation

Publish the project on GitHub with a strong README and tailor the resume for GenAI roles. Practice common LLM interview questions and case discussions.

PORTFOLIO

RESUME

INTERVIEW PREP

07 · PLACEMENT TRAINING



Our Placement Support Services

Placement Training comes with the Placement Assistance plan. Once your course is completed, our placement trainers take you through these six live sessions one by one - finish all six to earn the Placement Training Completion Certificate.

01

Complete your course

Placement Training starts after course completion

02

6 sessions, one by one

Each step is scheduled after the previous one is done

03

Get certified

Placement Training Completion Certificate

STEP 1 OF 6 · LIVE SESSION

1:1 Mock Interview

A live one-to-one mock interview with an industry expert, followed by honest feedback on your answers, confidence and presence.

- ✓ Technical & HR round simulation
- ✓ Questions from your course domain
- ✓ Body language & confidence pointers
- ✓ Personal improvement checklist

1:1 Live

Online

Feedback

After course completion · live with a placement trainer

STEP 2 OF 6 · EXPERT REVIEW

Resume Optimization

Your resume reviewed line by line and reshaped into a clear, ATS-friendly profile that highlights your projects and skills.

- ✓ Structure & format review
- ✓ Impact-focused bullet points
- ✓ Course projects showcased
- ✓ Keywords for your target roles

ATS-Ready

Projects

Rewrite

After course completion · live with a placement trainer

STEP 3 OF 6 · PROFILE MAKEOVER

LinkedIn Optimization

Headline, About, skills and featured work tuned so recruiters can find you and quickly see what you can do.

- ✓ Headline & About section
- ✓ Skills & endorsements plan
- ✓ Featured projects & certificates
- ✓ Networking & outreach tips

Discoverable

Skills

Network

After course completion · live with a placement trainer

STEP 4 OF 6 · INDUSTRY-LED

Career-Led Workshop

A practical workshop on job search strategy, applications, interviews and how hiring works in your domain.

- ✓ Finding the right openings
- ✓ Applications that get replies
- ✓ Interview rounds explained
- ✓ Offer & joining essentials

Job Search

Live

Hiring Process

After course completion · live with a placement trainer

STEP 5 OF 6 · MENTOR-LED

Career-Led Mentorship

Personal guidance from an industry mentor on roles, growth paths and your next career moves.

- ✓ Role & domain fit discussion
- ✓ Personal career roadmap
- ✓ Skill gaps to close next
- ✓ Answers to your career questions

1:1 Mentor

Career Path

Growth

After course completion · live with a placement trainer

STEP 6 OF 6 · COMMUNICATION

Industrial Communication Training

Workplace communication that employers expect: emails, meetings, presentations and professional etiquette.

- ✓ Professional emails & messages
- ✓ Meetings & stand-up updates
- ✓ Presenting your work clearly
- ✓ Workplace etiquette & teamwork

Emails

Speaking

Workplace

After course completion · live with a placement trainer

Placement Training Completion Certificate

Complete all six sessions to receive a verifiable certificate listing every training you finished - share it with recruiters along with your course certificate.

08 · SMES



Meet Our SME Board Members

Expert professionals guiding your journey through real-world industry knowledge and practical experience.

**SME BOARD MEMBER · SENIOR SME**

Karteek Ch

Associate

Tata Consultancy Services · Subject Matter Expert

Technology professional, mentor, and engineer specializing in Oracle, AI, IoT, and emerging technologies.

Technology professional and TCS Engineer specializing in Oracle, Java, Python, AI, and IoT. An experienced Subject Matter Expert and Mentor at Intern Stump, passionate about practical learning, emerging technologies, and industry-ready student development.

**SME BOARD MEMBER · SENIOR SME**

Sri Satya Lokesh

Lead Engineer

Dhan AI · Subject Matter Expert

Strategic Lead Engineer specializing in scalable microservices, AI automation, and Conversational AI.

Strategic Lead Engineer with 6+ years of experience in microservices, AI automation, and Conversational AI, specializing in Node.js, TypeScript, Python, Kafka, scalable architectures, performance optimization, mentoring, and enterprise technology modernization.

**SME BOARD MEMBER · SENIOR SME**

Gowtham P

Regional HR Business Partner

Policy Bazaar · Subject Matter Expert

Regional HR Business Partner specializing in strategic HR management, employee relations, and talent development.

Gowtham is a Regional HR Business Partner at Policybazaar, specializing in strategic human resource management, employee engagement, talent acquisition, performance management, workforce planning, and organizational development to support business growth and employee success.

09 · TOOLS AND TECHNIQUES



Key Tools and Techniques

Master industry-standard methodologies used by top-tier professionals, gaining practical exposure to tools, frameworks, and strategies for results.

INDUSTRY STANDARD

01

Python

The core programming language used to build, test and automate every LLM workflow in the course.

MOST USED

02

OpenAI API

Access powerful hosted models for chat, completion, function calling and embeddings in real applications.

OPEN SOURCE

03

Hugging Face Transformers

Load, run and fine-tune open-source models and datasets using the most popular model hub and library.

POPULAR FRAMEWORK

04

LangChain

Chain prompts, memory, tools and retrievers together to build structured LLM applications faster.

DATA FRAMEWORK

05

LlamaIndex

Connect LLMs to your own data sources and build efficient indexing and retrieval pipelines.

VECTOR SEARCH

06

FAISS and Chroma

Store and search embeddings quickly with vector databases that power retrieval-augmented generation.

EFFICIENT TRAINING

07

LoRA and PEFT

Fine-tune large models efficiently on limited hardware by training only small adapter layers.

DEPLOYMENT READY

08

Streamlit and FastAPI

Wrap your models in interactive interfaces and production-style APIs for easy sharing and deployment.

CORE TECHNIQUE

09

Prompt Engineering Techniques

Apply few-shot, chain-of-thought and structured output methods to get reliable and accurate model responses.

Many More Tools Integrated in Curriculum

10 · COURSE PROJECTS



Projects You Will Work On

Build a strong portfolio by working on industry-relevant projects that simulate real job scenarios, helping you gain confidence and career-ready skills.

INDUSTRY-LED PROJECT

01

Smart Prompt Playground

Build a web app that lets users test and compare prompts across different LLM APIs. Add templates, temperature controls and response logging.

Duration: 6-10 hours

INDUSTRY-LED PROJECT

02

Customer Support Chatbot

Create a conversational chatbot with memory that answers product queries in a consistent tone. Add fallback handling and conversation summaries.

Duration: 8-12 hours

INDUSTRY-LED PROJECT

03

Document Q&A with RAG

Build a system that answers questions from uploaded PDFs using embeddings and a vector database. Show source citations with every answer.

Duration: 10-14 hours

INDUSTRY-LED PROJECT

04

Semantic Search Engine

Develop a search tool that finds results by meaning instead of keywords. Compare embedding models and measure retrieval quality.

Duration: 6-10 hours

INDUSTRY-LED PROJECT

05

Resume Analyzer and Job Matcher

Use an LLM to extract skills from resumes and match them to job descriptions. Generate a score and improvement suggestions for each candidate.

Duration: 8-12 hours

INDUSTRY-LED PROJECT

06

Text Summarization Dashboard

Build a dashboard that summarizes articles, meeting notes and reports in several styles. Include length controls and a quality comparison view.

Duration: 6-10 hours

INDUSTRY-LED PROJECT

07

Fine-Tuned Domain Model

Fine-tune a small open-source model with LoRA on a custom dataset. Compare its outputs with the base model using clear metrics.

Duration: 10-16 hours

INDUSTRY-LED PROJECT

08

Multi-Tool AI Agent

Build an agent that plans tasks and calls tools such as web search, calculators and databases. Add guardrails and step-by-step logging.

Duration: 10-14 hours

INDUSTRY-LED PROJECT

09

Deployed LLM Application Capstone

Combine RAG, evaluation and a clean interface into a complete product and deploy it to the cloud. Document the architecture and present it as a portfolio piece.

Duration: 12-18 hours



11 · COURSE ENROLLMENT

Generative AI and LLMS

Learn to build real Generative AI and LLM applications, from prompt engineering and RAG to fine-tuning and deployment, through live mentor-led sessions and hands-on projects.

Starts	To be announced	Enrolment closes	Batch details soon
Duration	4 Months	Timings	To be announced
Batch	Batch details updating soon	Batch status	Batch open · Batch dates announced soon

Your plan

One-time fee · inclusive of GST

Plan 1

Course + Live Work Experience

Build real projects and earn a live work experience certificate.

₹6,999

₹20,000 · 65% OFF

Live classes & completion certificate	✓ Included
---------------------------------------	------------

Live projects	✓ Included
---------------	------------

Live work experience certificate	✓ Included
----------------------------------	------------

8C appraisal report	✓ Included
---------------------	------------

SME AI (Insyrio)	✓ Included
------------------	------------

Live chat with SME	✗ Not included
--------------------	----------------

Placement assistance	✗ Not included
----------------------	----------------

Plan perks	15% off across Intern Stump - coins, self-paced courses, masterclasses, mock interviews 20% off membership plans
------------	---

Also included

- Live classes with industry SMEs
- Course completion certificate
- Live projects reviewed by your SME

Good to know

Completion certificate

Verifiable ID - a sample certificate can be previewed on the course page.

8C Appraisal System

How the eight Cs work - see the 8C Appraisal section of this brochure.

Part-pay with iSure Coins

Use your coin balance at checkout.

Need help choosing?

See the FAQs or book a free demo.

12 · 8C APPRAISAL



World's First Ed-Tech with 8C Appraisal

Verified 8C Appraisal Certificate: every candidate receives an official corporate readiness report shared directly with hiring partners upon course completion.

The 8C Appraisal is a proprietary evaluation matrix measuring holistic performance across 8 career pillars for authentic proof-of-work:

Communication

Configuration

Chronicle

Completeness

Cognizance

Conception

Comprehension

Case Study



Frequently Asked Questions

Find answers to common questions about our learning tracks, placement assistance, and 8C evaluation model, or connect directly with an academic consultant.

Q1 Do I need prior coding experience to join Generative AI and LLMS?

No. The course starts with Python essentials and builds up step by step, so beginners can follow along while advanced learners get deeper challenges in later modules.

Q2 How long is the course?

The course runs for 4 Months with live mentor-led sessions, hands-on assignments and project reviews throughout.

Q3 When does the next batch start?

The upcoming batch is the next batch and starts on soon. Sessions are held at to be announced.

Q4 What is the course fee?

The course fee is ₹6,999 after a 65% discount from the original ₹20,000. Check the website for the latest offers.

Q5 Will I build real projects?

Yes. You will complete 9 hands-on projects, including RAG systems, chatbots, agents and a deployed capstone, which you can showcase in your portfolio.

Q6 Do I need a powerful computer or GPU?

No. A normal laptop with internet is enough. We use cloud notebooks and hosted APIs for tasks that need heavy computing.

Q7 Are sessions recorded if I miss a class?

Yes. Recordings and shared materials are available so you can revise or catch up, and mentors help clear doubts.

Q8 Will this course help me get a job?

The Technical track focuses on job-ready skills, portfolio projects and interview preparation. Outcomes depend on your effort, but you will have practical proof of your abilities.

Ready to take the next step?

Generative AI and LLMs · Live Course



Scan to view details & enrol

internstump.com/courses/generative-ai-and-llms



WEBSITE

internstump.com



EMAIL

info@internstump.com



CALL / WHATSAPP

+91 7396168247



Intern Stump Solutions · Generated on 26 Oct 2026, 07:06 PM IST from the live details on internstump.com. Prices, batches and schedules can change.