Admissions & Support: support@mockattempt.com
Phone / WhatsApp: +91 98765 43210
Course Code: GAIF • **Updated:** 01 Nov 2026
https://mockattempt.com/training/genai-full-stack-engineering-program

GAIF • GENERATIVE AI & LLM SYSTEMS **JOB-GUARANTEED CAREER MASTER PROGRAM**

100% PLACEMENT GUARANTEED* CAREER PROGRAM (6-MO REFUND*)

Generative AI, Large Language Models (LLM) & Agentic AI Full Stack Engineering

GenAI Full Stack Developer + AWS + DevOps | 100% Placement Guaranteed Career Program | 100% Placement Guaranteed* (6 Months Fees Refund*)

Duration: 36w (540h) • 15h/wk (560 Hrs) • **Modules:** 16 • **Live Topics:** 42 • **Labs & Tasks:** 84 • **Mocks / AI Vivas:** 84 / 84 • **Projects:** 16 Module + 1 Live Client Prod

Fixed Non-Negotiable Tuition Fee: Base Fee **₹59,999** + 18% GST (₹10,800) = **Total Incl. GST: ₹70,799**

Live Batch Tracks: Weekday Career Intensive Batch (Mon – Fri • 15h/wk) (Monday to Friday (5 Live Sessions / Week • 15h/wk Total) • 3.0 Hours / Day (90 Mins Live Architecture + 90 Mins Supervised Real-Repo Coding) • 37 Weeks (560 Hours • 15h/wk)) | Weekend Working-Pro Career Batch (Sat – Sun + Guided Sprint • 15h/wk) (Saturday & Sunday Live + Mid-Week Mentor PR Review (15h/wk Total) • 5.0 Hours Sat & Sun Live + 5.0 Hours Guided Real-Repository Sprint • 37 Weeks (560 Hours • 15h/wk))

✓ 6-Month Placement Guarantee*: Complete 100% assigned repositories & score ≥90% across all Topic Mock Tests & Voice AI Interviews → 100+ verified IT job openings within 6 months or 100% net tuition refund*.

Core Technologies & Engineering Stack: Python 3.12 Core & Advanced • FastAPI Asynchronous Backend • Django 5 & Django REST Framework • React 19 & TypeScript • Transformer & LLM Internals • Prompt Engineering & Structured Outputs • Hybrid RAG, Chunking & Rerankers • PostgreSQL pgvector & Qdrant Vector DBs • LangChain, LangGraph & CrewAI • Model Context Protocol (MCP) Servers • LoRA / QLoRA LLM Fine-Tuning • Multimodal Vision & Voice AI Pipelines • LLM Guardrails, Red-Teaming & RAGAS Evals • Redis Semantic Caching & Streaming SSE • vLLM / TGI Inference Optimization • AWS Bedrock, SageMaker & S3 • Docker & GPU/CPU Container Serving • Kubernetes & Terraform LLMOps • GitHub Actions CI/CD for AI Evals • OpenTelemetry & Langfuse Tracing

1. PROGRAM MASTERY OUTCOMES & ELIGIBILITY

End-to-End Engineering Outcomes:

1. Build and ship 14 real-world production GitHub repositories + 1 Live Client Production Cloud Deployment (zero dummy to-do apps)
2. Achieve ≥90% mastery across every Topic Mock Test and Spoken Gemini AI Technical Interview (unlimited free re-attempts allowed below 90%)
3. Earn Dual Barcode-Verifiable Credentials: Course Excellence Certificate + Official Project Internship Confirmation Letter
4. Unlock 6-Month Placement Support guaranteeing 100 curated relevant IT job openings — backed by our 100% Tuition Fee Refund* policy
5. Strong full-stack foundation: Python 3.12, async FastAPI, Django 5, React 19 + TypeScript 5, PostgreSQL 16, and Redis 7.
6. Advanced RAG & Vector Engineering: hybrid BM25 + dense retrieval (`pgvector`/Qdrant), reranking, query routing, and citations.
7. Multi-Agent Systems & LLM Ops: LangGraph stateful agents, tool-calling protocols (MCP), structured outputs, and human-in-the-loop gates.
8. Fine-tuning (LoRA/QLoRA), prompt-injection guardrails, RAGAS/LLM-as-judge evaluations, caching, and AWS Docker/K8s deployment.

Eligibility & Prerequisites: Basic computer literacy and commitment to 15 hours/week of live coding, daily tasks & real-repository sprints • Dedication to complete 100% of assigned tasks and re-attempt assessments until crossing ≥90% mastery • No prior Generative AI, Large Language Models (LLM) & Agentic AI Full Stack Engineering industry experience required — Module 1 starts from environment setup, core mental models, and plain-English fundamentals. • Basic computer & internet familiarity on a Windows, macOS, or Linux laptop (8 GB+ RAM recommended for smooth live coding labs). • Ideal for B.Tech / BCA / MCA / B.Sc IT students, non-CS beginners, career switchers, and working developers upgrading to 2026 production standards. • Commitment to attend live Weekday or Weekend classes, complete daily hands-on lab checklists, and practice topic-wise Mock Tests & AI Interviews.

2. DETAILED MODULE & TOPIC-WISE SYLLABUS (16 MODULES • 42 LIVE TOPICS)

MODULE 1 OF 16 • 38 LIVE + LAB HOURS • 3 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

Python 3.12 Advanced Internals, Asyncio & Scientific Computing

Real-repository engineering module covering Python 3.12 Advanced Internals, Asyncio & Scientific Computing with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ CPython Internals, Generators, Asyncio TaskGroups, Typing & Pydantic v2 Contracts — Purpose: Write rock-solid, high-concurrency Python systems capable of orchestrating 100s of LLM calls. • ✓ Vectorized NumPy, Pandas/Polars Dataframes, Linear Algebra & Probability for AI — Purpose: Manipulate embeddings, cosine similarities, and million-row evaluation datasets in milliseconds. • ✓ Python DSA & Algorithmic Optimization for AI/Product Engineering Interviews — Purpose: Clear Round-1 algorithmic coding gates at top AI labs, unicorns, and global GCCs.

1.1 CPython Internals, Generators, Asyncio TaskGroups, Typing & Pydantic v2 Contracts [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply CPython Internals, Generators, Asyncio TaskGroups, Typing & Pydantic v2 Contracts in real Generative AI, Large Language Models (LLM) & Agentic AI...

1.2 Vectorized NumPy, Pandas/Polars Dataframes, Linear Algebra & Probability for AI [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Vectorized NumPy, Pandas/Polars Dataframes, Linear Algebra & Probability for AI in real Generative AI, Large Language Models (LLM) & Agentic AI Fu...

1.3 Python DSA & Algorithmic Optimization for AI/Product Engineering Interviews [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Python DSA & Algorithmic Optimization for AI/Product Engineering Interviews in real Generative AI, Large Language Models (LLM) & Agentic AI Full S...

Module 1 Hands-On Capstone Project (100 Marks): Async Python Streaming Pipeline & Vectorized Tensor Math Repo — Architect, implement, test, and push the production GitHub repository 'Async Python Streaming Pipeline & Vectorized Tensor Math Repo' (strictly zero dummy scaffolding) demonstrating end-to-e...

MODULE 2 OF 16 • 38 LIVE + LAB HOURS • 3 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

Deep Learning Foundations, PyTorch & Transformer Internals

Real-repository engineering module covering Deep Learning Foundations, PyTorch & Transformer Internals with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ Tensors, Autograd Backpropagation, Optimizers (AdamW), Regularization & Schedulers — Purpose: Understand neural network training dynamics, vanishing gradients, and mixed precision. • ✓ Tokenizers (BPE/Tiktoken), Embeddings, Positional Encoding (RoPE) & Attention — Purpose: Demystify how GPT/Llama/Gemini transformers attend to context and generate tokens. • ✓ KV-Cache Memory Math, Context Windows, Sampling (`temperature`, `top_p`) & Quantization — Purpose: Calculate exact GPU VRAM budgets and latency trade-offs for production LLM serving.

2.1 Tensors, Autograd Backpropagation, Optimizers (AdamW), Regularization & Schedulers [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Tensors, Autograd Backpropagation, Optimizers (AdamW), Regularization & Schedulers in real Generative AI, Large Language Models (LLM) & Agentic AI...

2.2 Tokenizers (BPE/Tiktoken), Embeddings, Positional Encoding (RoPE) & Attention [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Tokenizers (BPE/Tiktoken), Embeddings, Positional Encoding (RoPE) & Attention in real Generative AI, Large Language Models (LLM) & Agentic AI Full...

2.3 KV-Cache Memory Math, Context Windows, Sampling (`temperature`, `top_p`) & Quantization [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply KV-Cache Memory Math, Context Windows, Sampling (`temperature`, `top_p`) & Quantization in real Generative AI, Large Language Models (LLM) & Agent...

Module 2 Hands-On Capstone Project (100 Marks): From-Scratch Multi-Head Self-Attention & PyTorch Trainer Repo — Architect, implement, test, and push the production GitHub repository 'From-Scratch Multi-Head Self-Attention & PyTorch Trainer Repo' (strictly zero dummy scaffolding) demonstrating end-to-e...

MODULE 3 OF 16 • 38 LIVE + LAB HOURS • 3 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

Prompt Engineering, Structured Outputs & Multimodal LLM APIs

Real-repository engineering module covering Prompt Engineering, Structured Outputs & Multimodal LLM APIs with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ System Prompts, Few-Shot, Chain-of-Thought, ReAct & Structured Pydantic Schemas — Purpose: Guarantee 100% schema-valid JSON extraction for downstream APIs and databases. • ✓ Multimodal Vision (OCR/Charts/UI), Audio/Speech Pipelines & Context Caching — Purpose: Build AI features that reason seamlessly across PDFs, screenshots, voice notes, and text. • ✓ Model Routing (Frontier vs Small Models), Retries, Fallbacks & Token Cost Controls — Purpose: Deliver flagship quality at 70% lower API cost with resilient multi-provider fallbacks.

3.1 System Prompts, Few-Shot, Chain-of-Thought, ReAct & Structured Pydantic Schemas [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply System Prompts, Few-Shot, Chain-of-Thought, ReAct & Structured Pydantic Schemas in real Generative AI, Large Language Models (LLM) & Agentic AI Fu...

3.2 Multimodal Vision (OCR/Charts/UI), Audio/Speech Pipelines & Context Caching [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Multimodal Vision (OCR/Charts/UI), Audio/Speech Pipelines & Context Caching in real Generative AI, Large Language Models (LLM) & Agentic AI Full S...

3.3 Model Routing (Frontier vs Small Models), Retries, Fallbacks & Token Cost Controls [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Model Routing (Frontier vs Small Models), Retries, Fallbacks & Token Cost Controls in real Generative AI, Large Language Models (LLM) & Agentic AI...

Module 3 Hands-On Capstone Project (100 Marks): Deterministic JSON Schema Extractor & Multimodal Vision/Audio Repo — Architect, implement, test, and push the production GitHub repository 'Deterministic JSON Schema Extractor & Multimodal Vision/Audio Repo' (strictly zero dummy scaffolding) demonstrating end...

MODULE 4 OF 16 • 38 LIVE + LAB HOURS • 3 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

Async FastAPI & Django 5 Enterprise AI Backend Engineering

Real-repository engineering module covering Async FastAPI & Django 5 Enterprise AI Backend Engineering with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ Django 5 + DRF Multi-Tenant RBAC, Billing, Audit Logs & Admin Workflows — Purpose: Manage enterprise users, API keys, workspaces, and usage quotas with battle-tested Django. • ✓ Async FastAPI Streaming (SSE / WebSockets), Background Tasks & Rate Limiters — Purpose: Stream real-time LLM tokens and tool-call events at sub-25ms gateway overhead. • ✓ PostgreSQL 16 ACID Schemas, Async SQLAlchemy 2, Celery + Redis 7 Worker Fleets — Purpose: Coordinate long-running document ingestion and multi-agent jobs reliably at scale.

4.1 Django 5 + DRF Multi-Tenant RBAC, Billing, Audit Logs & Admin Workflows [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Django 5 + DRF Multi-Tenant RBAC, Billing, Audit Logs & Admin Workflows in real Generative AI, Large Language Models (LLM) & Agentic AI Full Stack...

4.2 Async FastAPI Streaming (SSE / WebSockets), Background Tasks & Rate Limiters [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Async FastAPI Streaming (SSE / WebSockets), Background Tasks & Rate Limiters in real Generative AI, Large Language Models (LLM) & Agentic AI Full...

4.3 PostgreSQL 16 ACID Schemas, Async SQLAlchemy 2, Celery + Redis 7 Worker Fleets [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply PostgreSQL 16 ACID Schemas, Async SQLAlchemy 2, Celery + Redis 7 Worker Fleets in real Generative AI, Large Language Models (LLM) & Agentic AI Ful...

Module 4 Hands-On Capstone Project (100 Marks): Dual Django 5 SaaS Control Plane + Async FastAPI AI Gateway Repo — Architect, implement, test, and push the production GitHub repository 'Dual Django 5 SaaS Control Plane + Async FastAPI AI Gateway Repo' (strictly zero dummy scaffolding) demonstrating end-t...

MODULE 5 OF 16 • 38 LIVE + LAB HOURS • 3 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

Vector Databases (`pgvector` & Qdrant) & Enterprise Hybrid RAG

Real-repository engineering module covering Vector Databases (`pgvector` & Qdrant) & Enterprise Hybrid RAG with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ Document Parsing (Layout-Aware PDF/Tables), Semantic/Hierarchical Chunking & Metadata — Purpose: Preserve tables, headings, and clauses so retrieval never splits critical facts in half. • ✓ PostgreSQL `pgvector` (HNSW) & Qdrant Filtering, Hybrid BM25 + Dense Retrieval — Purpose: Combine exact keyword matching (SKUs, IDs, clauses) with deep semantic vector recall. • ✓ Query Rewriting, HyDE, Cross-Encoder Reranking & Inline Citation Grounding — Purpose: Boost top-3 retrieval precision above 95% with clickable page-level source proofs.

5.1 Document Parsing (Layout-Aware PDF/Tables), Semantic/Hierarchical Chunking & Metadata [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Document Parsing (Layout-Aware PDF/Tables), Semantic/Hierarchical Chunking & Metadata in real Generative AI, Large Language Models (LLM) & Agentic...

5.2 PostgreSQL `pgvector` (HNSW) & Qdrant Filtering, Hybrid BM25 + Dense Retrieval [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply PostgreSQL `pgvector` (HNSW) & Qdrant Filtering, Hybrid BM25 + Dense Retrieval in real Generative AI, Large Language Models (LLM) & Agentic AI Ful...

5.3 Query Rewriting, HyDE, Cross-Encoder Reranking & Inline Citation Grounding [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Query Rewriting, HyDE, Cross-Encoder Reranking & Inline Citation Grounding in real Generative AI, Large Language Models (LLM) & Agentic AI Full St...

Module 5 Hands-On Capstone Project (100 Marks): Citation-Grounded Hybrid BM25 + pgvector + Reranker RAG Repo — Architect, implement, test, and push the production GitHub repository 'Citation-Grounded Hybrid BM25 + pgvector + Reranker RAG Repo' (strictly zero dummy scaffolding) demonstrating end-to-en...

MODULE 6 OF 16 • 38 LIVE + LAB HOURS • 3 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

Advanced GraphRAG, Agentic Retrieval & Evaluation-Driven RAG

Real-repository engineering module covering Advanced GraphRAG, Agentic Retrieval & Evaluation-Driven RAG with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ Corrective RAG (CRAG), Self-RAG & Knowledge-Graph Entity Relationship Retrieval — Purpose: Answer complex multi-hop enterprise questions across interconnected company records. • ✓ Golden Datasets, RAGAS Metrics (Faithfulness, Context Precision/Recall) & CI Gates — Purpose: Block regressions automatically whenever chunking, prompts, or embedding models change. • ✓ Semantic Caching with Redis Vector Search & Latency/Throughput Optimization — Purpose: Serve repeated or similar customer queries in <20ms at zero marginal LLM cost.

6.1 Corrective RAG (CRAG), Self-RAG & Knowledge-Graph Entity Relationship Retrieval [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Corrective RAG (CRAG), Self-RAG & Knowledge-Graph Entity Relationship Retrieval in real Generative AI, Large Language Models (LLM) & Agentic AI Fu...

6.2 Golden Datasets, RAGAS Metrics (Faithfulness, Context Precision/Recall) & CI Gates [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Golden Datasets, RAGAS Metrics (Faithfulness, Context Precision/Recall) & CI Gates in real Generative AI, Large Language Models (LLM) & Agentic AI...

6.3 Semantic Caching with Redis Vector Search & Latency/Throughput Optimization [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Semantic Caching with Redis Vector Search & Latency/Throughput Optimization in real Generative AI, Large Language Models (LLM) & Agentic AI Full S...

Module 6 Hands-On Capstone Project (100 Marks): Automated RAGAS / LLM-as-Judge CI Benchmark & GraphRAG Repo — Architect, implement, test, and push the production GitHub repository 'Automated RAGAS / LLM-as-Judge CI Benchmark & GraphRAG Repo' (strictly zero dummy scaffolding) demonstrating end-to-end...

MODULE 7 OF 16 • 38 LIVE + LAB HOURS • 3 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

Agentic AI Systems — LangChain, LangGraph & Multi-Agent Orchestration

Real-repository engineering module covering Agentic AI Systems — LangChain, LangGraph & Multi-Agent Orchestration with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ LangGraph StateGraphs, Cycles, Conditional Routing, Persistence & Time-Travel Debug — Purpose: Build deterministic, resumable agent workflows instead of brittle infinite-loop scripts. • ✓ Multi-Agent Topologies (Supervisor, Swarm, Planner-Executor-Critic) & CrewAI — Purpose: Coordinate specialized research, SQL, coding, and QA agents on complex enterprise tasks. • ✓ Human-in-the-Loop Interrupts, Long-Term Episodic Memory & Cost/Step Circuit Breakers — Purpose: Require explicit human approval before agents execute refunds, emails, or DB writes.

7.1 LangGraph StateGraphs, Cycles, Conditional Routing, Persistence & Time-Travel Debug [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply LangGraph StateGraphs, Cycles, Conditional Routing, Persistence & Time-Travel Debug in real Generative AI, Large Language Models (LLM) & Agentic A...

7.2 Multi-Agent Topologies (Supervisor, Swarm, Planner-Executor-Critic) & CrewAI [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Multi-Agent Topologies (Supervisor, Swarm, Planner-Executor-Critic) & CrewAI in real Generative AI, Large Language Models (LLM) & Agentic AI Full...

7.3 Human-in-the-Loop Interrupts, Long-Term Episodic Memory & Cost/Step Circuit Breakers [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Human-in-the-Loop Interrupts, Long-Term Episodic Memory & Cost/Step Circuit Breakers in real Generative AI, Large Language Models (LLM) & Agentic...

Module 7 Hands-On Capstone Project (100 Marks): Stateful LangGraph Supervisor, Coder & Analyst Multi-Agent Repo — Architect, implement, test, and push the production GitHub repository 'Stateful LangGraph Supervisor, Coder & Analyst Multi-Agent Repo' (strictly zero dummy scaffolding) demonstrating end-to...

MODULE 8 OF 16 • 38 LIVE + LAB HOURS • 3 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

Model Context Protocol (MCP), Audited Tool Calling & Voice Agents

Real-repository engineering module covering Model Context Protocol (MCP), Audited Tool Calling & Voice Agents with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ Designing & Securing Model Context Protocol (MCP) Servers, Resources & Tools — Purpose: Connect LLMs cleanly to GitHub, Postgres, CRM, and internal APIs via standard MCP contracts. • ✓ Sandboxed Code Execution, SQL-Generation Guardrails & Idempotent Action Tools — Purpose: Let AI agents run analytics and workflow actions with zero SQL-injection or RCE risk. • ✓ Low-Latency Spoken Voice AI Agents (STT -> Agentic LLM -> Streaming TTS) — Purpose: Build natural real-time voice interviewers, support concierges, and sales copilots.

8.1 Designing & Securing Model Context Protocol (MCP) Servers, Resources & Tools [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Designing & Securing Model Context Protocol (MCP) Servers, Resources & Tools in real Generative AI, Large Language Models (LLM) & Agentic AI Full...

8.2 Sandboxed Code Execution, SQL-Generation Guardrails & Idempotent Action Tools [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Sandboxed Code Execution, SQL-Generation Guardrails & Idempotent Action Tools in real Generative AI, Large Language Models (LLM) & Agentic AI Full...

8.3 Low-Latency Spoken Voice AI Agents (STT -> Agentic LLM -> Streaming TTS) [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Low-Latency Spoken Voice AI Agents (STT -> Agentic LLM -> Streaming TTS) in real Generative AI, Large Language Models (LLM) & Agentic AI Full Stac...

Module 8 Hands-On Capstone Project (100 Marks): Custom MCP Server Suite + Real-Time Spoken Voice AI Agent Repo — Architect, implement, test, and push the production GitHub repository 'Custom MCP Server Suite + Real-Time Spoken Voice AI Agent Repo' (strictly zero dummy scaffolding) demonstrating end-to-...

MODULE 9 OF 16 • 38 LIVE + LAB HOURS • 2 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

Modern AI Frontend Engineering with React 19 & TypeScript 5

Real-repository engineering module covering Modern AI Frontend Engineering with React 19 & TypeScript 5 with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ Strict TypeScript 5, React 19 Concurrent UI, Optimistic Threads & Chat State — Purpose: Craft ChatGPT/Claude-grade multi-thread workspaces with instant keyboard workflows. • ✓ Generative UI Cards, Live Tool-Execution Timelines, PDF Citation Split-Panes & E2E Tests — Purpose: Render interactive charts, approval buttons, and highlighted PDF sources inline.

9.1 Strict TypeScript 5, React 19 Concurrent UI, Optimistic Threads & Chat State [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Strict TypeScript 5, React 19 Concurrent UI, Optimistic Threads & Chat State in real Generative AI, Large Language Models (LLM) & Agentic AI Full...

9.2 Generative UI Cards, Live Tool-Execution Timelines, PDF Citation Split-Panes & E2E Tests [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Generative UI Cards, Live Tool-Execution Timelines, PDF Citation Split-Panes & E2E Tests in real Generative AI, Large Language Models (LLM) & Agen...

Module 9 Hands-On Capstone Project (100 Marks): React 19 + TypeScript Generative UI, Streaming Canvas & Copilot Repo — Architect, implement, test, and push the production GitHub repository 'React 19 + TypeScript Generative UI, Streaming Canvas & Copilot Repo' (strictly zero dummy scaffolding) demonstrating e...

MODULE 10 OF 16 • 38 LIVE + LAB HOURS • 2 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

LLM Fine-Tuning (LoRA / QLoRA), Preference Alignment & vLLM Serving

Real-repository engineering module covering LLM Fine-Tuning (LoRA / QLoRA), Preference Alignment & vLLM Serving with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ Dataset Curation, Synthetic Data Distillation, SFT with LoRA / QLoRA (PEFT) & DPO — Purpose: Fine-tune compact 7B/8B open-weight models to beat giant models on domain tasks. • ✓ PagedAttention Inference Serving with vLLM / TGI, Batching & Quantization (AWQ/FP8) — Purpose: Maximize tokens/sec per GPU and deploy private self-hosted enterprise LLM endpoints.

10.1 Dataset Curation, Synthetic Data Distillation, SFT with LoRA / QLoRA (PEFT) & DPO [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Dataset Curation, Synthetic Data Distillation, SFT with LoRA / QLoRA (PEFT) & DPO in real Generative AI, Large Language Models (LLM) & Agentic AI...

10.2 PagedAttention Inference Serving with vLLM / TGI, Batching & Quantization (AWQ/FP8) [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply PagedAttention Inference Serving with vLLM / TGI, Batching & Quantization (AWQ/FP8) in real Generative AI, Large Language Models (LLM) & Agentic A...

Module 10 Hands-On Capstone Project (100 Marks): QLoRA Domain-Tuned Open-Weight Model & High-Throughput vLLM Repo — Architect, implement, test, and push the production GitHub repository 'QLoRA Domain-Tuned Open-Weight Model & High-Throughput vLLM Repo' (strictly zero dummy scaffolding) demonstrating end-t...

MODULE 11 OF 16 • 38 LIVE + LAB HOURS • 2 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

AI Safety, Red-Teaming, Prompt-Injection Defense & Governance

Real-repository engineering module covering AI Safety, Red-Teaming, Prompt-Injection Defense & Governance with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ Direct/Indirect Prompt Injection, Jailbreak Red-Teaming & Output Topic Guardrails — Purpose: Prevent malicious PDFs or user prompts from hijacking agent tools or leaking secrets. • ✓ PII Masking/Anonymization, RBAC Document ACLs in RAG & Audit Compliance Logs — Purpose: Ensure an employee's AI copilot never retrieves documents outside their permission scope.

11.1 Direct/Indirect Prompt Injection, Jailbreak Red-Teaming & Output Topic Guardrails [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Direct/Indirect Prompt Injection, Jailbreak Red-Teaming & Output Topic Guardrails in real Generative AI, Large Language Models (LLM) & Agentic AI...

11.2 PII Masking/Anonymization, RBAC Document ACLs in RAG & Audit Compliance Logs [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply PII Masking/Anonymization, RBAC Document ACLs in RAG & Audit Compliance Logs in real Generative AI, Large Language Models (LLM) & Agentic AI Full...

Module 11 Hands-On Capstone Project (100 Marks): Enterprise PII Redaction, Jailbreak Firewall & Guardrails Repo — Architect, implement, test, and push the production GitHub repository 'Enterprise PII Redaction, Jailbreak Firewall & Guardrails Repo' (strictly zero dummy scaffolding) demonstrating end-to-...

MODULE 12 OF 16 • 38 LIVE + LAB HOURS • 2 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

Cloud Infrastructure (AWS + Bedrock) & Docker/Kubernetes LLMOps

Real-repository engineering module covering Cloud Infrastructure (AWS + Bedrock) & Docker/Kubernetes LLMOps with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ Multi-Stage Dockerfiles, GPU/CPU Kubernetes Workloads, HPA & AWS VPC/RDS/S3/Bedrock — Purpose: Deploy scalable full-stack AI backends and managed Bedrock models on enterprise AWS. • ✓ Terraform IaC & GitHub Actions CI/CD with Automated LLM Evaluation Quality Gates — Purpose: Automate infrastructure provisioning, security scans, eval benchmarks, and zero-downtime releases.

12.1 Multi-Stage Dockerfiles, GPU/CPU Kubernetes Workloads, HPA & AWS VPC/RDS/S3/Bedrock [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Multi-Stage Dockerfiles, GPU/CPU Kubernetes Workloads, HPA & AWS VPC/RDS/S3/Bedrock in real Generative AI, Large Language Models (LLM) & Agentic A...

12.2 Terraform IaC & GitHub Actions CI/CD with Automated LLM Evaluation Quality Gates [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Terraform IaC & GitHub Actions CI/CD with Automated LLM Evaluation Quality Gates in real Generative AI, Large Language Models (LLM) & Agentic AI F...

Module 12 Hands-On Capstone Project (100 Marks): Terraform AWS + Docker + Kubernetes + Bedrock Production AI Repo — Architect, implement, test, and push the production GitHub repository 'Terraform AWS + Docker + Kubernetes + Bedrock Production AI Repo' (strictly zero dummy scaffolding) demonstrating end-t...

MODULE 13 OF 16 • 38 LIVE + LAB HOURS • 2 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

LLMOps Observability, Tracing & GenAI System Design Interviews

Real-repository engineering module covering LLMOps Observability, Tracing & GenAI System Design Interviews with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ OpenTelemetry & Langfuse Tracing, Token/Cost Attribution, Drift & Feedback Loops — Purpose: Inspect every multi-agent step, prompt version, latency span, and user thumbs-down in prod. • ✓ Generative AI & Full-Stack System Design Interviews (Enterprise Copilot, Voice AI, Search) — Purpose: Ace Principal/Senior AI Engineer architecture rounds with concrete latency and cost math.

13.1 OpenTelemetry & Langfuse Tracing, Token/Cost Attribution, Drift & Feedback Loops [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply OpenTelemetry & Langfuse Tracing, Token/Cost Attribution, Drift & Feedback Loops in real Generative AI, Large Language Models (LLM) & Agentic AI F...

13.2 Generative AI & Full-Stack System Design Interviews (Enterprise Copilot, Voice AI, Search) [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Generative AI & Full-Stack System Design Interviews (Enterprise Copilot, Voice AI, Search) in real Generative AI, Large Language Models (LLM) & Ag...

Module 13 Hands-On Capstone Project (100 Marks): Langfuse / OpenTelemetry LLM Trace, Cost & Drift Dashboard Repo — Architect, implement, test, and push the production GitHub repository 'Langfuse / OpenTelemetry LLM Trace, Cost & Drift Dashboard Repo' (strictly zero dummy scaffolding) demonstrating end-to...

MODULE 14 OF 16 • 38 LIVE + LAB HOURS • 2 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

Live Client GenAI Production Sprint & ≥90% Placement Guarantee Gate

Real-repository engineering module covering Live Client GenAI Production Sprint & ≥90% Placement Guarantee Gate with supervised live coding, ≥90% AI Mock Gates, and production GitHub PR verification.

Objectives: ✓ End-to-End Client Production Deployment, Red-Team Sign-Off & Official Internship Letter — Purpose: Launch NexusAgent live on HTTPS cloud infrastructure for real client operations. • ✓ ≥90% Mock + Voice AI Gate Verification & 6-Month 100-Opening Placement Activation — Purpose: Achieve ≥90% across all assessments and enter 100+ GenAI & Full-Stack job openings.

14.1 End-to-End Client Production Deployment, Red-Team Sign-Off & Official Internship Letter [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply End-to-End Client Production Deployment, Red-Team Sign-Off & Official Internship Letter in real Generative AI, Large Language Models (LLM) & Agent...

14.2 ≥90% Mock + Voice AI Gate Verification & 6-Month 100-Opening Placement Activation [90 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply ≥90% Mock + Voice AI Gate Verification & 6-Month 100-Opening Placement Activation in real Generative AI, Large Language Models (LLM) & Agentic AI...

Module 14 Hands-On Capstone Project (100 Marks): NexusAgent Live Client Multi-Agent Production & Placement Repo — Architect, implement, test, and push the production GitHub repository 'NexusAgent Live Client Multi-Agent Production & Placement Repo' (strictly zero dummy scaffolding) demonstrating end-to-...

MODULE 15 OF 16 • 14 LIVE + LAB HOURS • 3 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

Multi-Agent Orchestration, Tool-Use Guardrails & Production RAG Evaluation

Self-explanatory list-and-bullet production mastery module covering Multi-Agent Orchestration, Tool-Use Guardrails & Production RAG Evaluation in Generative AI, Large Language Models (LLM) & Agentic AI Full Stack Engineering, paired with hands-on labs, capstone repository, Topic Mock Tests, and Spoken Gemini AI Viva.

Objectives: ✓ Agentic Design Patterns: Planner-Executor, Reflection, Tool Sandboxing & State Persistence — Purpose: Engineer reliable multi-step AI agents with bounded retries, timeouts, and human approval gates. • ✓ Advanced Hybrid Multimodal RAG, Re-Ranking & Automated Faithfulness/Safety Evals — Purpose: Eliminate hallucinations on complex PDFs, tables, and enterprise knowledge bases. • ✓ Prompt-Injection Defense, PII Redaction, Output Schema Enforcement & Audit Logging — Purpose: Meet enterprise security and compliance standards before exposing LLM endpoints to users. • ✓ Build, verify, and document 'Evaluated Multi-Agent Research & Action Platform with LangGraph, MCP Tools & Guardrails' with clean Git commits and edge-case protection. • ✓ Clear 100% of Topic Timed Mock Tests and Spoken Voice + Google Gemini AI Technical Interviews.

15.1 Agentic Design Patterns: Planner-Executor, Reflection, Tool Sandboxing & State Persistence [65 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Agentic Design Patterns: Planner-Executor, Reflection, Tool Sandboxing & State Persistence in real Generative AI, Large Language Models (LLM) & Ag...

15.2 Advanced Hybrid Multimodal RAG, Re-Ranking & Automated Faithfulness/Safety Evals [65 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Advanced Hybrid Multimodal RAG, Re-Ranking & Automated Faithfulness/Safety Evals in real Generative AI, Large Language Models (LLM) & Agentic AI F...

15.3 Prompt-Injection Defense, PII Redaction, Output Schema Enforcement & Audit Logging [65 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Prompt-Injection Defense, PII Redaction, Output Schema Enforcement & Audit Logging in real Generative AI, Large Language Models (LLM) & Agentic AI...

Module 15 Hands-On Capstone Project (100 Marks): Evaluated Multi-Agent Research & Action Platform with LangGraph, MCP Tools & Guardrails — Architect, implement, test, and document 'Evaluated Multi-Agent Research & Action Platform with LangGraph, MCP Tools & Guardrails' applying every bullet-wise engineering concept from Module...

MODULE 16 OF 16 • 14 LIVE + LAB HOURS • 3 TOPICS • 1 CAPSTONE PROJECT (100 MARKS)

LLM Inference Optimization, Full-Stack Streaming UX & GenAI Placement Defense

Self-explanatory list-and-bullet production mastery module covering LLM Inference Optimization, Full-Stack Streaming UX & GenAI Placement Defense in Generative AI, Large Language Models (LLM) & Agentic AI Full Stack Engineering, paired with hands-on labs, capstone repository, Topic Mock Tests, and Spoken Gemini AI Viva.

Objectives: ✓ Semantic Caching, Model Routing, Quantization/LoRA Trade-Offs & Token FinOps — Purpose: Deliver sub-second time-to-first-token while cutting production LLM inference costs. • ✓ Full-Stack React 19 + TypeScript Streaming Interfaces, Citations & Feedback Loops — Purpose: Build intuitive AI product UX with live tool-status indicators and thumbs-up/down eval capture. • ✓ End-to-End GenAI & LLM System Design Defense for Top AI Engineering Roles — Purpose: Architect enterprise copilots, agent platforms, and evaluation pipelines in final interviews. • ✓ Build, verify, and document 'vLLM/FastAPI Streaming Gateway + React 19 Copilot UI + AWS Observability Repository' with clean Git commits and edge-case protection. • ✓ Clear 100% of Topic Timed Mock Tests and Spoken Voice + Google Gemini AI Technical Interviews.

16.1 Semantic Caching, Model Routing, Quantization/LoRA Trade-Offs & Token FinOps [65 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Semantic Caching, Model Routing, Quantization/LoRA Trade-Offs & Token FinOps in real Generative AI, Large Language Models (LLM) & Agentic AI Full...

16.2 Full-Stack React 19 + TypeScript Streaming Interfaces, Citations & Feedback Loops [65 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply Full-Stack React 19 + TypeScript Streaming Interfaces, Citations & Feedback Loops in real Generative AI, Large Language Models (LLM) & Agentic AI...

16.3 End-to-End GenAI & LLM System Design Defense for Top AI Engineering Roles [65 mins • Assignment (100M) • Hands-On Lab (50M) • 2 Mock Sets • 2 AI Vivas]

Purpose: To understand and apply End-to-End GenAI & LLM System Design Defense for Top AI Engineering Roles in real Generative AI, Large Language Models (LLM) & Agentic AI Full Sta...

Module 16 Hands-On Capstone Project (100 Marks): vLLM/FastAPI Streaming Gateway + React 19 Copilot UI + AWS Observability Repository — Architect, implement, test, and document 'vLLM/FastAPI Streaming Gateway + React 19 Copilot UI + AWS Observability Repository' applying every bullet-wise engineering concept from Module 16 (...

3. LIVE CLIENT PRODUCTION PROJECT, GLOBAL CERTIFICATIONS & CAREER SCOPE

End-of-Course Live Client Production Project + Official Internship Confirmation Letter

Project: NexusAgent Enterprise Multi-Agent RAG Copilot, Voice/MCP Gateway & Full-Stack LLMOps Cloud • **Confirmed Role:** Generative AI, Agentic LLM & Full-Stack Cloud Engineering Resident

Domain & Production Infra: Enterprise AI Copilot, FinTech & Autonomous Support Client (Python 3.12 + FastAPI + Django 5 + LangGraph + MCP + pgvector + React 19 TS + AWS)

- Engineered hybrid BM25 + pgvector semantic retrieval, cross-encoder reranking, and hallucination-free citation pipelines.
- Built autonomous LangGraph multi-agent orchestrators and custom MCP tool servers with human-in-the-loop approvals.
- Fine-tuned domain adapters with QLoRA, enforced prompt-injection guardrails, and automated RAGAS regression evals in CI.
- Shipped React 19 TypeScript streaming workspaces and async FastAPI/Django backends onto Docker, Kubernetes, and AWS.

Aligned Global Vendor Certifications: **NCA-GENL** (NVIDIA-Certified Associate: Generative AI & LLMs) • **AIF-C01** (AWS Certified AI Practitioner (Generative AI & Foundation Models)) • **PCPP1-32-101** (PCPP1™ – Certified Professional in Python Programming 1) • **SAA-C03** (AWS Certified Solutions Architect – Associate) • **EDB-PG16** (EDB Certified PostgreSQL 16 Associate)

Target Job Roles: Generative AI & Agentic LLM Full Stack Engineer, Applied AI / LLM RAG & Multi-Agent Engineer, Full Stack AI Product Engineer (Python + React + LLMOps), LLMOps & Enterprise AI Platform Engineer

Realistic Annual CTC Scope: **Placement-Guaranteed GenAI Full Stack Engineer (0–1 Yr):** ₹7.5 LPA – ₹16.5 LPA **Senior GenAI / Agentic AI Systems SDE-2 (1–3 Yrs):** ₹17.0 LPA – ₹34.0 LPA **Principal AI Architect / Staff LLM Engineer (3–6+ Yrs):** ₹35.0 LPA – ₹70.0+ LPA

MockAttempt Academy Pvt. Ltd. • Email: support@mockattempt.com • Phone/WhatsApp: +91 98765 43210 • Web: https://mockattempt.com/training/genai-full-stack-engineering-program

Try 2 Free Topic-Wise 50-MCQ Mock Tests & Spoken Gemini AI Interviews daily on every topic above before enrolling.