PRACTICAL ARCHITECTURE • REAL-WORLD DELIVERY • ZERO FLUFF

AI/ML & Agentic Systems Engineering

Realtime Live Sessions • Tailored to Your Stack • Production Project Focus

Designed for **engineers at all career stages**—from **junior and mid-level developers** looking to fast-track their career into high-demand AI engineering, to **senior practitioners and tech leads** aiming to master enterprise-scale agent architecture and hands-on project delivery.

ZERO FILLER. ACCELERATED PATH TO PRODUCTION.

Whether you are stepping into AI for the first time or leveling up your existing systems, we skip the slow academic theory. We provide clear foundational bridge materials whenever needed and focus 100% of our live time on building real production projects, debugging agents, and understanding enterprise trade-offs.

Four Pillars of Real-World AI Engineering

THE CORE CAPABILITIES YOU WILL DEVELOP

Architectural Decision Making

Know when to use Classical ML, GenAI, autonomous agents, or simple deterministic logic. Clearly justify buy vs. build choices based on latency, cost, and maintainability.

`System Design • Cost vs. Quality`

Modern Retrieval & Context

Build enterprise RAG pipelines that actually work in production: multi-stage retrieval, semantic reranking, graph databases, and context window optimization at scale.

`Hybrid Search • Context Pipelines`

Autonomous Agents & Tooling

Design coordinated multi-agent workflows using open protocols (MCP), practical function calling, persistent memory, and human-in-the-loop safeguards.

`MCP • Planning • Self-Correction`

Production Reliability & Safety

Ship with confidence: harden against prompt injection, sandbox code execution, implement automated quality evaluations, and trace production token consumption.

`Security • Automated Evals • MLOps`

PRIMARY OBJECTIVE: ENTERPRISE PROJECT HANDLING EXPERIENCE

Move beyond isolated notebook scripts to full-stack project ownership. You will build, debug, and deliver resilient AI systems that withstand real production workloads, pass enterprise security reviews, and satisfy tight latency and cost requirements.

Hands-On Engineering Modules

Each module focuses on practical architecture decisions, real failure modes, and code you can ship.

01

AI Decision Making & Architecture Selection

System Strategy

Knowing when not to use an LLM is a superpower. Learn how to systematically evaluate problem requirements and pick between deterministic code, classical ML, or agents.

- Classical ML vs. GenAI vs. Agents
- Deterministic vs. Probabilistic Logic
- Build vs. Buy Cost Analysis
- Latency, Privacy & Accuracy Tradeoffs

02

Modern Context Engineering & Advanced RAG

Accurate Retrieval

Move far beyond basic vector search. Construct multi-stage retrieval pipelines with reranking, hybrid search, and graph indexing that eliminate hallucinations.

- Reasoning Models in Production (o1/o3, R1)
- Multi-Stage RAG Pipelines
- Hybrid Search & Cross-Encoder Reranking
- Graph RAG & Dynamic Model Routing

03

Agentic AI, Orchestration & Memory Systems

Agent Architecture

Build autonomous systems capable of multi-step reasoning, dynamic tool selection, session memory, and graceful self-correction when tasks fail.

- Single vs. Multi-Agent Topologies
- Planning Trees (ReAct, Plan-and-Solve)
- Short-term & Long-term Memory Stores
- Human-in-the-Loop & Circuit Breakers

04

Model Context Protocol (MCP) & Tool Integration

Open Standard

Adopt the open standard transforming enterprise agent integration. Build, host, and secure standard MCP servers connecting models to company databases and APIs.

- MCP Client & Server Implementation
- Tools, Resources & Prompt Expositions
- Role-Based Access & Permission Scopes
- Connecting Enterprise Databases & Services

05

AI Quality Evaluation & Automated Testing

Quality Assurance

Never deploy an AI system on guesswork. Build automated evaluation pipelines, synthetic test harnesses, and regression benchmarks to verify quality before release.

- Groundedness, Faithfulness & Relevance
- Tool-Calling Accuracy & Task Completion
- LLM-as-a-Judge Calibration & Golden Sets
- Continuous Regression Testing in CI/CD

06

AI Security, Red-Teaming & Safety Controls

Security & Hardening

Protect against real threats: direct and indirect prompt injection, data exfiltration, poisoned context, and unauthorized agent privilege escalation.

- Direct & Indirect Injection Defense
- Context Poisoning & Dependency Risks
- Sandboxed Execution (Docker/Containers)
- Structured Audit Logging & Guardrails

Production Scale, Economics & Governance

Turn working prototypes into reliable, cost-controlled, and compliant platforms.

07

Production AI Infrastructure & Gateways

Platform Operations

Deploy robust model gateways handling streaming, token rate limits, automated fallbacks across providers, and end-to-end distributed telemetry.

- Model Gateways & Prompt Version Control
- Semantic Caching & Token Throttling
- Streaming Retries, Backoffs & Fallbacks
- OpenTelemetry & Distributed Agent Tracing

08

Data Engineering & Real-Time Context Stores

Data Pipelines

Construct continuous pipelines for unstructured company knowledge, multi-tenant vector databases, and real-time freshness guarantees.

- Document Chunking, Parsing & Extraction
- Vector Database Indexing (HNSW, Quantization)
- Knowledge Graphs & Context Linking
- Real-Time Ingestion & Data Freshness

09

Inference Economics & Cost Optimization

Cost Engineering

Prevent runaway cloud invoices. Learn how to combine prompt caching, model routing, and smaller specialized models to cut costs by 5x to 10x.

- Token Economics & GPU Cost Modeling
- Self-Hosted Inference vs. Commercial APIs
- Prompt Compression & Speculative Decoding
- Cost/Accuracy Trade-Off Optimization

10

Governance, Compliance & Auditability

Risk Management

Ensure your enterprise agents comply with emerging privacy regulations, internal data protection standards, and corporate audit requirements.

- Responsible AI & Model Explainability
- PII Masking & Privacy-Preserving Inference
- Model Risk Management & Access Controls
- Regulatory Preparedness & Compliance Audits

11

AI-Assisted Software Engineering

Developer Velocity

Supercharge development teams using repo-level reasoning agents, automated test generation, and intelligent pull request verification.

- Repository-Level AST & Dependency Graphs
- Autonomous Bug Detection & Patching
- Automated Unit & Integration Test Suites
- CI/CD Code Review & Architecture Checks

12

Applied Research & Engineering Workflows

Continuous Growth

Build structured routines to digest technical research papers, benchmark new models quickly, and turn cutting-edge ideas into reliable code.

- Technical Whitepaper & ArXiv Synthesis
- Rapid Experimentation & Benchmarking
- Source Verification & Technical Discipline
- Building Reusable Internal Engineering Tools

Built Around Real Project Execution

Personalized Mentorship • High Technical Depth • Direct Relevance to Your Stack

01

Live Problem Solving

Work directly on active architectural challenges and live debugging rather than watching passive pre-recorded slides.

02

Tailored to Your Level

Customized to your background—whether fast-tracking from junior engineering into applied AI, or solving senior architecture challenges in your company's stack.

03

Architecture Defense

Practice presenting and defending your system choices in realistic whiteboard discussions with experienced engineers.

04

Production Tooling

Walk away with tested blueprints for MCP servers, evaluation suites, gateway rate-limiters, and telemetry setups.

Practical Capability Assessment

Where most teams get stuck vs. what you will command after this program

CORE ENGINEERING CAPABILITY	COMMON INDUSTRY GAP	YOUR POST-TRAINING CAPABILITY
Architecture Selection: Classical ML vs. GenAI vs. Agents	Trial & Error / Hype Driven	Clear Decision Matrix & ROI Justification
Enterprise Context: Multi-Stage RAG & Hybrid Indexing	Basic Vector Search / Frequent Hallucinations	High-Precision Production RAG & Reranking
Agent Engineering: Tool Use, Model Context Protocol (MCP)	Fragile, Hardcoded Function Calls	Standardized Scalable MCP Implementations
Reliability & Security: Testing, Sandboxing & Guardrails	Manual Spot Checks / Unhandled Injections	Automated CI/CD Evals & Red-Teaming
Inference Economics: Cost, Latency & Rate Budgeting	Unpredictable Cloud Bills & Throttles	Predictable Cost/Token & 5×–10× Savings
Technical Defense: Whiteboard & Stakeholder Reviews	Vague High-Level Explanations	Confident, Defensible Technical Rationale

Who Should Join: Junior & Mid-Level Engineers upgrading into AI • Senior Developers & Tech Leads • Data Scientists & ML Practitioners • Anyone wanting real-world enterprise project handling experience on their resume.

Course Inquiry & Personalized Program Details

DIRECT FACULTY CONSULTATION

Connect for a short technical consultation to discuss your background, chosen project, and schedule.

DELIVERY FORMATS

1-on-1 Mentorship / Small Cohort Track

PROGRAM DURATION

6 to 8 Weeks (Flexible for Working Professionals)

PREREQUISITES

Basic Programming (Python/JS) • No prior AI theory needed

INQUIRIES & ENROLLMENT

Direct Technical Consultation Available Upon Request