



CORSO AVANZATO {AI}

LLMOps & AI Evals

Mandare un'applicazione AI in produzione è facile. Tenerla affidabile, valutarne la qualità in modo oggettivo e controllarne i costi è tutta un'altra storia, ed è esattamente dove la maggior parte dei progetti si arena. Senza valutazioni sistematiche, ogni modifica al prompt o cambio di modello è un salto nel buio.

Questo corso insegna la disciplina operativa dell'AI: costruire dataset di valutazione, eseguire eval automatiche, versionare prompt e configurazioni, individuare regressioni prima del deploy e monitorare comportamento, qualità e costi in produzione. È il complemento naturale di chi già costruisce agenti e applicazioni AI.

Livello

Avanzato

Durata

Corso da 16 ore

Lingua

Italiano

A chi si rivolge

- Sviluppatori e team che hanno già applicazioni o agenti AI, anche in prototipo
- ML/AI engineer e tech lead responsabili della qualità in produzione
- Chi ha seguito “Building Production AI Agents” o ha esperienza equivalente

Cosa faremo

Corso pratico con live coding: costruiremo una pipeline di valutazione e osservabilità attorno a un’applicazione AI esistente, dal primo dataset di eval fino al monitoring in produzione.

Lavoreremo con piattaforme di observability ed evaluation di riferimento (es. Langfuse, promptfoo o equivalenti), confermate a ridosso di ogni edizione.

Takeaway:

- Definire cosa significa “qualità” per il proprio sistema AI
- Costruire dataset di valutazione e metriche significative
- Automatizzare le eval (incluso LLM-as-judge) e integrarle nella CI
- Versionare e confrontare prompt, modelli e configurazioni
- Monitorare qualità, comportamento e costi in produzione

Numero massimo di partecipanti: 20

Skill minime necessarie

- Esperienza consolidata con almeno un linguaggio (Python o JavaScript/TypeScript)
- Aver già lavorato con LLM API o applicazioni AI
- Familiarità con Git, testing e pipeline CI/CD

Argomenti Trattati

- Modulo 1 — Fondamenti di valutazione
 - Perché le eval sono il collo di bottiglia dell'AI in produzione
 - Definire qualità e criteri di successo per il proprio caso
 - Tipi di eval: deterministiche, basate su riferimento, LLM-as-judge
- Modulo 2 — Costruire dataset ed eval
 - Creare e curare un dataset di valutazione rappresentativo
 - Metriche utili (e metriche fuorvianti)
 - LLM-as-judge: quando funziona, limiti e calibrazione
- Modulo 3 — Eval nel ciclo di sviluppo
 - Versionamento di prompt, modelli e configurazioni
 - Regression testing: cogliere i peggioramenti prima del deploy
 - Integrare le eval in CI/CD
- Modulo 4 — Observability e operatività
 - Tracing e logging di un sistema AI/agentico
 - Monitorare comportamento, qualità e drift in produzione
 - Controllo e ottimizzazione dei costi: routing, caching, scelta del modello
 - Feedback degli utenti e miglioramento continuo
 - Monitoraggio dei guardrail in produzione: violazioni, falsi positivi e tuning

Cosa è necessario

- Il proprio computer portatile
- Tanta buona volontà e voglia di imparare
- Connessione ad Internet
- Account gratuiti su piattaforme di terze parti (indicati prima del corso)
- Una API key per un provider LLM

Cosa è incluso

- Corso pratico con live coding
- Supporto setup environment
- Slide in formato PDF
- Repository del Progetto
- Attestato di Partecipazione
- Follow Up di fine corso
- Canale Slack dedicato ai partecipanti

Dove si svolge

Full-Remote

È possibile svolgere il corso in modalità *full-remote* con gli strumenti messi a disposizione da Devmy suddividendo, se lo si desidera, il tutto in sessioni da 4h.