Profice

AI ACADEMY

Applicare l’Intelligenza Artificiale nello sviluppo software

AI ACADEMY

Hackaton agentic RAG
26/06/2025

INTRODUZIONE DELL'ISTRUTTORE

Tamas Szakacs

Formazione

- Laureato come programmatore matematico
- MBA in management

Principali esperienze di lavoro

- Amministratore di sistemi UNIX
- Oracle DBA
- Sviluppatore di Java, Python e di Oracle PL/SQL
- Architetto (solution, enterprise, security, data)
- Ricercatore tecnologico e interdisciplinare di IA

Dedicato alla formazione continua

- Teorie, modelli, framework IA
- Ricerche IA
- Strategie aziendali
- Trasformazione digitale
- Formazione professionale

email: tamas.szakacs@proficegroup.it

MOTIVI E RIASSUNTO DEL CORSO

L'**Intelligenza Artificiale (AI)** è oggi il motore dell'innovazione in ogni settore, grazie alla sua capacità di analizzare dati, automatizzare processi e generare nuove soluzioni. Questo corso offre una panoramica completa e pratica sullo sviluppo di applicazioni AI moderne, guidando i partecipanti dall'ideazione al rilascio in produzione.

Attraverso una **combinazione di teoria chiara ed esercitazioni pratiche**, saranno affrontate le tecniche e gli strumenti più attuali: **machine learning, deep learning, reti neurali, Large Language Models (LLM), Transformers, Retrieval Augmented Generation (RAG)** e progettazione di agenti AI.
Le competenze acquisite saranno applicate in progetti concreti, dallo sviluppo di chatbot all'integrazione di modelli generativi, fino al deploy di soluzioni AI in ambienti reali e collaborativi.

Il percorso è pensato per chi vuole imparare a progettare, valutare e integrare sistemi AI di nuova generazione, con particolare attenzione alle best practice di programmazione, collaborazione in team, sicurezza, valutazione delle performance ed etica dell'AI.

DURATA: 17 GIORNI

OBIETTIVI

Il percorso formativo è progettato per **giovani consulenti junior**, con una conoscenza base di programmazione, che stanno iniziando un percorso professionale nel settore AI.

L’obiettivo centrale è fornire una panoramica pratica, completa e operativa sull’intelligenza artificiale moderna, guidando ogni partecipante attraverso tutte le fasi fondamentali.

OBIETTIVI

- Allineare conoscenze AI, ML, DL di tutti i partecipanti
- Saper usare e orchestrare modelli LLM (closed e open-weight)
- Costruire pipeline RAG complete (retrieval-augmented generation)
- Progettare agenti AI semplici con strumenti moderni (LangChain, tool calling)
- Capire principi di valutazione, robustezza e sicurezza dei sistemi GenA
- Migliorare la produttività come sviluppatori usando tool GenAI-driven
- Padroneggiare best practice di sviluppo, versioning e deploy AI
- Introdurre i fondamenti di Graph Data Science e Knowledge Graph
- Ottenere capacità di valutazione dei modelli e metriche
- Comprensione dell'etica e dei bias nei modelli di intelligenza artificiale
- Approfondire le normative di riferimento: AI Act, compliance e governance AI

Il corso è **estremamente pratico** (circa il 40% del tempo in esercitazioni hands-on, notebook, challenge e hackathon), con l'utilizzo di Google Colab, GitHub, e tutti gli strumenti necessari per lavorare su progetti reali e simulati.

STRUTTURA DELLE GIORNATE – PROGRAMMA BREVE

Tutte le giornate sono di 8 ore (9:00-17:00), con 1 ora di pausa suddivisa (mezz'ora pranzo, due pause da 15 min durante la mattina e il pomeriggio).

La progettazione sintetica delle giornate:

Giorno	Tema	Breve descrizione
1	Git & Python clean-code	Collaborazione su progetti reali, versionamento, codice pulito e testato
2	Machine Learning Supervised	Modelli supervisionati per predizione e classificazione
3	Machine Learning Unsupervised	Clustering, riduzione dimensionale, scoperta di pattern
4	Prompt Engineering avanzato	Scrivere e valutare prompt efficaci per modelli generativi
5	LLM via API (multi-vendor)	Uso pratico di modelli LLM via API, autenticazione, deployment
6	Come costruire un RAG	Pipeline end-to-end per Retrieval-Augmented Generation
7	Tool-calling & Agent design	Progettare agenti AI che usano strumenti esterni
8	Hackathon: Agentic RAG	Challenge pratica: chatbot agentico RAG in team

STRUTTURA DELLE GIORNATE – PROGRAMMA BREVE

Tutte le giornate sono di 8 ore (9:00-17:00), con 1 ora di pausa suddivisa (mezz'ora pranzo, due pause da 15 min durante la mattina e il pomeriggio).

La progettazione sintetica delle giornate:

Giorno	Tema	Breve descrizione
9	Hackathon: Rapid Prototyping	Da prototipo a web-app con Streamlit e GitHub
10	AI Productivity Tools	Workflow con IDE AI-powered, automazione e refactoring assistito
11	Docker & HF Spaces Deploy	Deployment di app GenAI containerizzate o su HuggingFace Spaces
12	AI Act & ISO 42001 Compliance	Fondamenti di compliance e governance AI
13	Knowledge Base & Graph Data Science	Introduzione a Knowledge Graph e query con Neo4j
14	Model evaluation & osservabilità	Metriche avanzate, explainability, strumenti di valutazione
15	AI bias, fairness ed etica applicata	Analisi dei rischi, metriche e mitigazione dei bias
16-17	Project Work & Challenge finale	Lavoro a gruppi, POC/POD, presentazione e votazione progetti

METODOLOGIA DEL CORSO

1. Approccio introduttivo ma avanzato

Il corso è introduttivo nei concetti base dell'AI applicata allo sviluppo, ma affronta anche tecnologie, modelli e soluzioni avanzate per garantire un apprendimento completo.

2. Linguaggio adattato

Il linguaggio utilizzato è chiaro e adattato agli studenti, con spiegazioni dettagliate dei termini tecnici per favorirne la comprensione e l'apprendimento graduale.

3. Esercizi pratici

Gli esercizi pratici sono interamente svolti online tramite piattaforme come Google Colab o notebook Python, eliminando la necessità di installare software sul proprio computer.

4. Supporto interattivo

È possibile porre domande in qualsiasi momento durante le lezioni o successivamente via email per garantire una piena comprensione del materiale trattato.

NOTA

Il corso segue un **approccio laboratoriale**: ogni giornata combina sessioni teoriche chiare e concrete con molte attività pratiche supervisionate, per sviluppare *competenze reali* immediatamente applicabili.

I partecipanti lavoreranno spesso in gruppo, useranno notebook in Colab e versioneranno codice su GitHub, vivendo una vera simulazione del lavoro in azienda AI.

Nessun prerequisito avanzato richiesto: si partirà dagli strumenti e flussi fondamentali, con una crescita graduale verso le tecniche più attuali e richieste dal mercato.

ORARIO TIPICO DELLE GIORNATE

Orario	Attività	Dettaglio
09:00 – 09:30	Teoria introduttiva	Concetti chiave, schema della giornata
09:30 – 10:30	Live coding + esercizio guidato	Esempio pratico, notebook Colab
10:30 – 10:45	*Pausa breve*	
10:45 – 11:30	Approfondimento teorico	Tecniche, best practice
11:30 – 12:30	Esercizio hands-on individuale	Sviluppo o completamento di codice
12:30 – 13:00	Discussione soluzioni + Q&A	Condivisione e correzione
13:00 – 14:00	*Pausa pranzo*	
13:30 – 14:15	Teoria avanzata / nuovi tools	Nuovi strumenti, pattern, demo
14:15 – 15:30	Esercizio a gruppi / challenge	Lavoro di squadra su task reale
15:30 – 15:45	*Pausa breve*	
15:45 – 16:30	Sommario teorico e pratico	
16:30 – 17:00	Discussioni, feedback	Riepilogo, best practice, domande aperte

DOMANDE?

Cominciamo!

OBIETTIVI DELLA GIORNATA

Obiettivi della giornata

- Analizzare la richiesta e il perimetro progettuale dell'hackathon sul chatbot agentico RAG.
- Pianificare i lavori e distribuire i compiti tra i partecipanti in modo collaborativo.
- Utilizzare e integrare gli elementi già sviluppati durante il corso (pipeline RAG, modelli, agenti, utility, notebook).
- Individuare, progettare e realizzare le componenti mancanti per completare il progetto.
- Integrare le diverse parti e testare il funzionamento end-to-end del chatbot agentico.
- Redigere una documentazione funzionale essenziale (architettura, flussi, setup, esempi d'uso).
- Preparare la consegna finale e la presentazione del progetto realizzato.

ANALISI DOCUMENTALE E PROTEZIONE DATI

SmartDocs Srl – Analisi documentale e protezione dati

Scenario:
SmartDocs Srl, media azienda europea, deve gestire email e documenti contenenti dati sensibili di clienti (es: IBAN, codice fiscale, indirizzi, nomi, numeri di telefono).

L’azienda vuole automatizzare:

- Estrazione di entità e dati sensibili (NER, pattern matching)
- Riepilogo automatico e risposta alle richieste clienti
- Tutelare la privacy (alcuni dati NON devono mai lasciare il server locale)
- Minimizzare i costi cloud e garantire risposte rapide

ANALISI DOCUMENTALE E PROTEZIONE DATI

Useremo due modelli LLM per la soluzione aziendale

Architettura semplificata

1. **Modello locale open source (es: TinyLLaMA)**
    1. Viene eseguito localmente, direttamente sul vostro computer o su server aziendali.
    2. Si occupa delle operazioni più “sensibili”, come l’estrazione di dati personali e la protezione della privacy (Named Entity Recognition e masking dei dati).
    3. È veloce, economico, e mantiene i dati riservati all’interno dell’azienda.

2. **Modello cloud avanzato (es: Azure OpenAI GPT-3.5/4)**
    1. Viene utilizzato tramite API esterne, in cloud.
    2. Si occupa di attività più complesse come il riepilogo automatico, l’analisi semantica e la generazione di risposte ai clienti.
    3. Permette di gestire testi lunghi e offre maggiore potenza di calcolo e qualità delle risposte.

ANALISI DOCUMENTALE E PROTEZIONE DATI

L’obiettivo:

Sfruttare **i punti di forza di entrambi i modelli**:

- La privacy e la velocità del modello locale
- La potenza e la flessibilità del modello cloud

Garantire che **i dati più delicati non escano dall’azienda**, mentre sfruttiamo le migliori tecnologie disponibili per la produttività e l’automazione.

Distribuzione dei lavori

- Tutti lavorano sullo **stesso problema reale** ma con strumenti diversi.

Lavoro locale autonomo

- privacy, sicurezza, regex/NER, masking.

Lavoro cloud con l’aiuto dell’istruttore

- prompt, parametri, API, gestione delle risposte.

Altri componenti futuri per i giorni successivi

- RAG, contestualizzazione, configurazione, deployment ecc.
- La **collaborazione** tra i modelli con una architettura AI.

RUOLI DEI DUE MODELLI

1. Modello locale (TinyLLaMA o simili) — “Difensore/controllore”
Viene **eseguito localmente.**
Obiettivi:

- **NER (Named Entity Recognition)**: trova nomi, indirizzi, IBAN, codice fiscale, numeri, ecc.
- **Pattern Detection**: segnala se sono presenti dati sensibili o “red flag”.
- **RAG** (Retrieval Augmented Generation) opzionale: usa una knowledge base aziendale locale per arricchire le risposte.
- **Controlla l’output** prima che venga inviato al modello cloud, mascherando o anonimizzando i dati sensibili (es. sostituendo “Mario Rossi” con “[NOME]”).

2. Modello cloud (Azure GPT-3.5, GPT-4, ecc) — “Analista/colloquio”
Viene usato tramite API cloud, guidato passo-passo.
Obiettivi:

- Riceve documenti **già “ripuliti”** o parzialmente anonimizzati dal modello locale.
- Esegue:
  - Riepilogo (summary)
  - Analisi semantica
  - Generazione di risposte per i clienti
- Gestisce solo dati che non violano la privacy.

ESEMPIO DI PIPELINE COLLABORATIVA

Input:
L'utente carica un documento/email.

Passo 1 (locale):
Il modello locale fa NER, segnala dati sensibili e li anonimizza (es: "Mario Rossi" → "[NOME]", "IT60X0542811101000000123456" → "[IBAN]").

Passo 2 (controllo):
L'output ripulito viene controllato/validato (gli esperti possono anche vedere che i dati sono davvero rimossi).

Passo 3 (cloud):
Il testo anonimizzato viene inviato all'API Azure che fa il riepilogo, classifica la richiesta e prepara una risposta.

Passo 4 (output):
L'output finale può essere ricomposto localmente, reinserendo alcune entità dove permesso, oppure consegnato così.

RUOLI DI PROGETTO NEL TEAM AGENTIC RAG CHATBOT

Introduzione
Per lavorare in modo efficace su un progetto complesso come lo sviluppo di un chatbot agentico basato su RAG, è fondamentale definire chiaramente i ruoli all'interno del team. Ogni ruolo contribuisce con competenze specifiche per garantire qualità, sicurezza, organizzazione e successo della soluzione.

Principali ruoli di progetto:

Project Manager (PM)
Coordina le attività, pianifica tempi e risorse, facilita la comunicazione e monitora l'avanzamento del progetto.

Business Analyst (BA)
Raccoglie i requisiti del cliente, definisce il perimetro funzionale, traduce i bisogni in specifiche tecniche e funzionali.

Solution Architect
Progetta l'architettura generale del sistema, definisce le scelte tecnologiche e l'integrazione tra le componenti (LLM, RAG, agenti, API, interfacce).

RUOLI DI PROGETTO NEL TEAM AGENTIC RAG CHATBOT

Introduzione
Per lavorare in modo efficace su un progetto complesso come lo sviluppo di un chatbot agentico basato su RAG, è fondamentale definire chiaramente i ruoli all'interno del team. Ogni ruolo contribuisce con competenze specifiche per garantire qualità, sicurezza, organizzazione e successo della soluzione.

Principali ruoli di progetto:

Developer (Frontend/Backend)
Implementa le funzionalità del chatbot, gestisce l'integrazione tra moduli, scrive codice e risolve bug.

Prompt Engineer
Progetta, ottimizza e testa i prompt per LLM e agenti, curando l'efficacia e la sicurezza delle interazioni AI.

Security Specialist
Analizza rischi, implementa misure di sicurezza (es. controllo accessi, protezione dati, guardrail su input/output).

Tester / QA
Esegue test funzionali, di integrazione e sicurezza, valida i risultati, individua errori e propone correzioni.

DELIVERABLES E DOCUMENTI MINIMI PER IL PROGETTO

Per consegnare i lavori bisogna preparare questi deliverable:

- **Risposta progettuale**
  Breve descrizione di come il team ha risolto il problema proposto, con riferimento alle scelte principali.
- **Schema architetturale**
  Schema essenziale (testuale o diagramma semplice) dell'architettura della soluzione e dei principali componenti (modello, agenti, pipeline, API, ecc.).
- **Codice sorgente**
  Tutto il codice sviluppato, con commenti chiari e autoesplicativi (documentazione del codice generata direttamente dai commenti).
- **Dataset di test**
  File di dati usati per le prove e le demo, rappresentativi dei casi d'uso.
- **Metodo di test**
  Breve descrizione del metodo di test applicato: quali casi, quali dati, come è stato valutato il funzionamento.
- **(Opzionale – da aggiungere dopo lezione su etica/EU AI Act)**
  Eventuali note su conformità etica, privacy e regole AI.

DOMANDE?

Lavoro in gruppi

DOMANDE?

PAUSA

GRAZIE PER L'ATTENZIONE