Travel Guide Assistant

Sistema di Content Creation Assistita per il Travel

Ottobre 2025

Building a better working world

- *Jacopo Bonanno*
- *Pietro Montresori*
- *Roberto Parodo*
- *Luca Sangiovanni*
- *Monica Salvati*

Introduzione

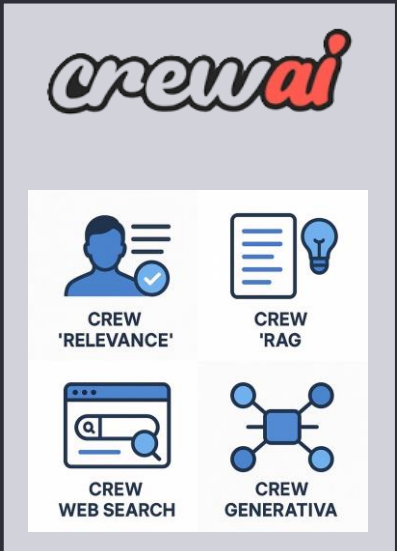
Una soluzione avanzata per automatizzare la creazione di contenuti turistici di qualità, combinando intelligenza artificiale e fonti affidabili.

Il sistema ha l'obiettivo di creare contenuti di guide turistiche e contenuti travel, ad esempio la ricerca dei migliori hotel in una destinazione, riducendo l'effort manuale e assicurando coerenza e qualità delle informazioni.

Architettura del sistema

- Il sistema è stato realizzato con CrewAI, un framework che consente lo sviluppo di applicazioni multi-agent

- Il sistema è composto da 4 Crew principali, ognuna formata da agenti con compiti specifici e ben definiti:



- *Crew Relevance*: Analizza e valuta automaticamente la pertinenza della domanda dell'utente rispetto al dominio travel, garantendo risposte mirate e appropriate
- *Crew Rag*: Effettua il recupero intelligente delle informazioni dai PDF indicizzati utilizzando tecniche di ricerca ibride avanzate
- *Crew Web Search*: Arricchisce il contesto con fonti online validate, attingendo esclusivamente da domini travel affidabili, specificati in una whitelist
- *Crew Generativa*: Sintetizza tutte le informazioni raccolte per generare la risposta finale e produce report dettagliati in formato Markdown

Rappresentazione grafica dell'architettura

Flusso operativo:

1. Input utente: L'utente inserisce una domanda specifica su turismo e viaggi attraverso l'interfaccia del sistema sviluppata su streamlit

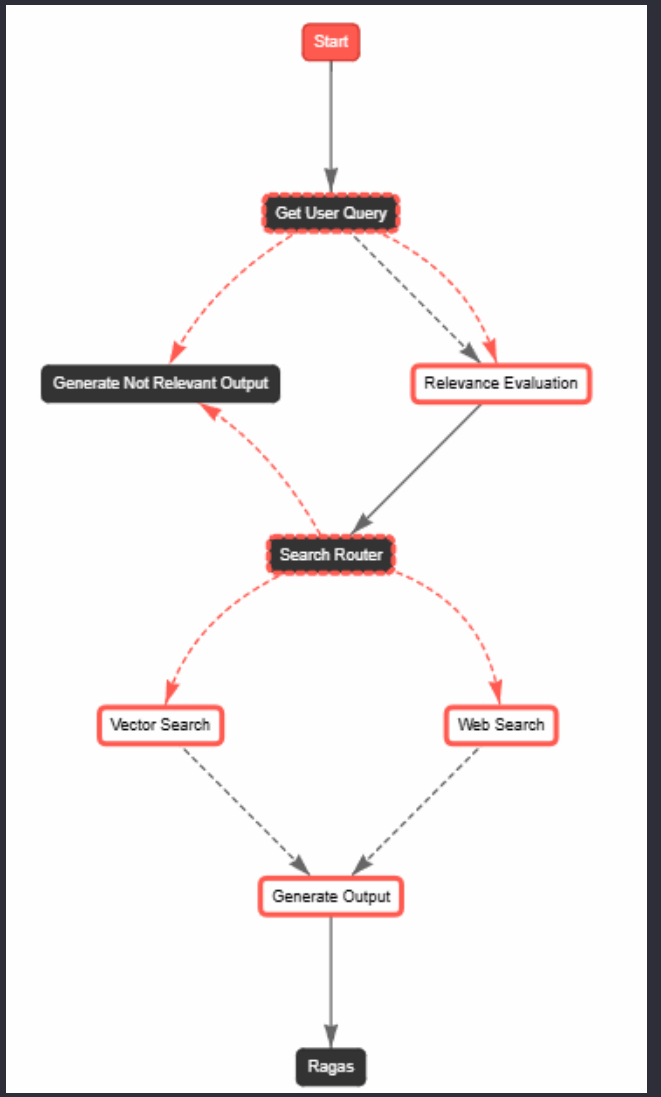
2. **Valutazione rilevanza**: Il Crew di valutazione determina automaticamente la pertinenza della richiesta al dominio travel

3. **Retrieval ibrido**: Sistema di ricerca avanzato che combina semantic search, keyword matching e MMR da database Qdrant

4. **Enrichment Web**: Arricchimento del contesto con risultati web filtrati da domini affidabili nella travel whitelist

5. **Generazione Finale**: La Crew generativa produce la risposta finale e il report dettagliato in formato Markdown

6. **Valutazione Ragas**: Viene impiegato ragas per effettuare la valutazione delle performance del sistema RAG



Controlli di sicurezza

I controlli di sicurezza sugli input garantiscono che i dati inseriti siano affidabili e privi di rischi, assicurando integrità, protezione e qualità delle informazioni elaborate dagli agenti

Nel nostro sistema servono a:

1. **Prevenire attacchi di prompt injection (Azure AI Content Safety)**
2. **PII detection (Azure AI Language):** rilevare ed eventualmente mascherare dati sensibili
3. **Normalizzare e Decodificare** correttamente i contenuti (es. UTF-8)
4. **Whitelist siti sicuri:** l'agente dedicato alle ricerche web accede esclusivamente a fonti affidabili
5. Verifica istruzioni malevoli **(Azure AI Content Safety)**: i documenti vengono analizzati per identificare contenuti potenzialmente dannosi
6. Controllo leggibilità: testi con lo stesso colore dello sfondo o nascosti vengono rilevati ed eliminati per ridurre ambiguità e tecniche di evasione
7. Esito del controllo: i documenti che non superano i controlli non vengono ingeriti nel sistema

EU AI Act

Il livello di rischio del nostro sistema, secondo le metriche utilizzate dall'EU AI Act, è il seguente:

Rischio Limitato

Avviso per l'utente: prima dell'esecuzione del flow, viene mostrato un messaggio che segnala la presenza di contenuti generati con AI:

⚠️ Nota: Il testo generato è prodotto dall'AI. Verifica sempre le informazioni prima di prenderle come definitive.

Trasparenza e consapevolezza: l'utente è sempre consapevole di stare interagendo con contenuti prodotti da AI

Sistema di valutazione

Sono stati adottati due framework open-source per monitorare e valutare sistemi di AI

Opik

Opik è una piattaforma **open-source** pensata per lo sviluppo, la valutazione e il monitoraggio di applicazioni basate su grandi modelli linguistici (LLM)

Ragas

RAGAS (Retrieval-Augmented Generation Assessment) è un framework open-source creato per valutare le pipeline **RAG** (cioè sistemi che combinano componenti di recupero d'informazioni e generazione con modelli linguistici) con metriche automatiche

Risultati metriche con Ragas

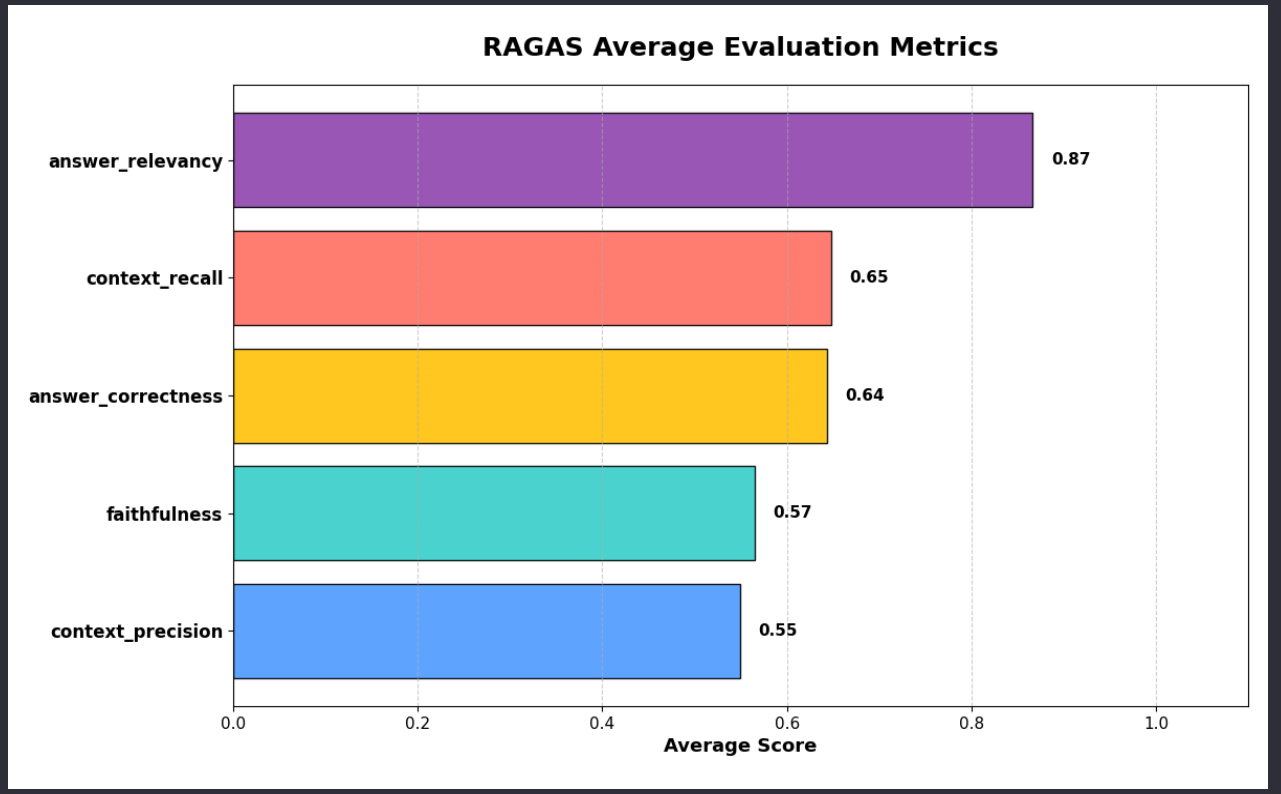
1. **Context Precision**: Valuta la capacità del sistema di classificare correttamente le informazioni rilevanti, valutando il rapporto segnale-rumore nel contesto recuperato

2. **Context Recall**: Valuta la capacità del sistema di recuperare le informazioni rilevanti necessarie per rispondere correttamente alla domanda

3. **Faithfulness**: Misura la coerenza della risposta rispetto al contesto, la risposta è considerata coerente se tutte le affermazioni possono essere dedotte dal contesto

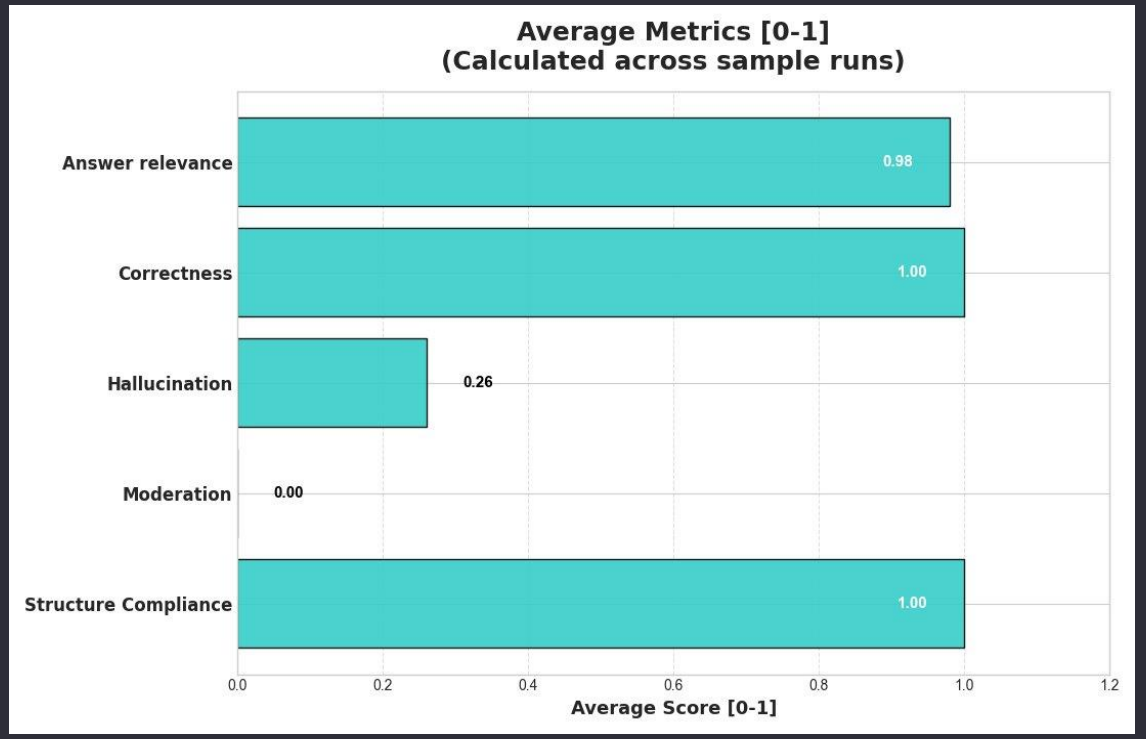
4. **Answer Relevancy**: Valuta se la risposta generata è pertinente alla domanda, ovvero se risponde effettivamente alla domanda dell'utente

5. **Answer Correctness**: Misura quanto la risposta generata dal modello si avvicini a quella corretta



Risultati metriche con Opik scala (0-1)

1. Answer Relevance: Grado di aderenza della risposta alla domanda
2. Correctness (bool): Indica se la risposta dell'assistente è utile o meno a soddisfare la richiesta dell'utente
3. Hallucination: Misura in cui il modello introduce informazioni non supportate dalle fonti o dal prompt. Un valore alto è indice di molta allucinazione e vice versa
4. Moderation: Valutazione della presenza di contenuti sensibili o violazioni (es. hate, violenza...)
5. Structure Compliance (bool): Indica se la risposta segue esattamente il formato o schema richiesto (nel nostro caso un file .md). Valore sì/no



Risultati metriche con Opik scala (0-5)

1. Argument Strength: Solidità del ragionamento, coerenza logica, supporto con evidenze, assenza di salti logici
2. Conciseness: Capacità di esprimere le informazioni necessarie senza prolissità o dettagli superflui
3. Document Redundancy: Quanta sovrapposizione informativa (duplicazione) esiste tra i documenti interni forniti come contesto
4. Internal Document Coverage: Indice di quanto la risposta si sia basata sui documenti interni
5. Redundancy: Indice delle ripetizioni inutili o concetti duplicati dentro la risposta generata. Un valore basso indica molta ripetitività
6. Relevance: Aderenza della domanda al topic
7. Reliance: Grado in cui la risposta è fondata (grounded) sulle fonti fornite rispetto a conoscenza esterna web