Progetto GAINS

(Generative AI for Network Sustainability)

Deliverable D3.3

Technical Report su Progettazione e Implementazione Knowledge Manager

Giordano, A. – Basile, L. – De Giacomo A. – Haider, B.- Fruncillo, D.
2026-01

Work Package: WP3 Progettazione, Implementazione Prototipale e Valutazione degli Agenti

Stato: Definitivo | Versione: 1.0

Classificazione

Abstract

Questo lavoro descrive la progettazione e l'implementazione prototipale del Knowledge Manager, componente di knowledge-plane orientato all'operatività per ambienti telco e cloud, con l'obiettivo di trasformare documentazione, policy e artefatti tecnici in un sistema interrogabile, applicabile e verificabile. Il Knowledge Manager è concepito come servizio containerizzato, riproducibile in Docker, capace di operare con un endpoint esterno OpenAI-compatibile per embeddings e generazione, mantenendo disaccoppiamento dal provider e controllo sul perimetro informativo effettivamente trasmesso all'esterno. Il design è centrato su tre requisiti non negoziabili: applicabilità contestuale, grounding con citazioni verificabili e sicurezza contro contaminazione ed esfiltrazione. La soluzione adotta un'architettura a piani: pipeline di ingest deterministica, knowledge store con ricerca ibrida e filtri strutturati, e runtime agentico per retrieval e sintesi. L'ingest è implementato in Langflow mediante flow versionati, esportabili e importabili, che realizzano parsing, normalizzazione, chunking deterministico, hashing canonicalizzato, redaction obbligatoria per snapshot e controlli anti-leakage con rifiuto fail-closed. L'indicizzazione utilizza Weaviate con configurazione BYOV, in cui i vettori sono calcolati dall'applicazione e associati agli oggetti persistiti, così da preservare portabilità e governance del ciclo embeddings. Il retrieval combina keyword search e vector search tramite hybrid search, con fusion algorithms selezionabili e metadati di scoring ispezionabili, e applica filtri vincolanti basati su proprietà flattened di applicabilità come vendor, ruolo del device, servizio e vincoli di versione, riducendo la probabilità di drift semantico in contesti operativi. Il comportamento agentico è orchestrato tramite Elysia, adottando un decision tree che rende esplicite le transizioni tra classificazione della richiesta, gating, retrieval multi-sorgente, re-ranking deterministico, costruzione dell'evidence pack, sintesi grounded e validazione finale. La modalità operativa impone vincoli stringenti: nessun claim prescrittivo è emesso senza almeno una reference valida, e i conflitti tra fonti sono risolti con policy-first, degradando la risposta a non-operativa quando le evidenze sono insufficienti. A supporto dell'auditabilità, il sistema produce un Explainability Pack persistibile e esportabile, che include mapping claim-refs, parametri di retrieval, hash degli artefatti, e indicatori di qualità come citation coverage e unsupported claim rate, abilitando verifiche ripetibili e diagnosi

di regressioni. Sono definiti contratti API coerenti per ingest, query cited o structured, blueprint di servizio, snippet discovery, e explain pack, con envelope comune, model di errori, RBAC e controlli di export. L'implementazione prototipale include una reference repository, docker-compose completo, bootstrap automatico di schema e flow, gestione segreti via Docker secrets e smoke test riproducibile. La valutazione è strutturata su metriche di retrieval e grounding, includendo precision e stabilità del top-k, tasso di violazione dell'applicabilità, citation coverage e assenza di claim non supportati in modalità operativa, con una pipeline di regressione su dataset bloccato. La sicurezza è trattata come requisito end-to-end: minimizzazione dei payload verso l'LLM, difese contro prompt injection diretta e indiretta tramite evidence-pack isolation, policy di export control, rate limiting e fail-closed, in coerenza con le principali tassonomie di rischio per applicazioni LLM-integrate. Nel complesso, il Knowledge Manager fornisce una base implementativa solida per integrare conoscenza tecnica e governance operativa in un contesto agentico, con riproducibilità industriale, tracciabilità e controlli di qualità verificabili.

Indice