What are vector and embedding weaknesses?

Vector and embedding weaknesses (OWASP LLM08:2025) are flaws in how retrieval-augmented generation (RAG) systems create, store, and query embeddings. Embeddings can be inverted to recover sensitive source text, vector stores that mix tenants can leak data across customers, and unvalidated documents can poison retrieval so the model is fed attacker-controlled context. Defend with per-tenant isolation, access control on the vector store, source validation, and treating retrieved content as untrusted.

HOW IT WORKS

01 How the attacks work

Three main paths. Embedding inversion: given access to embeddings, an attacker reconstructs approximate original text, so storing embeddings of sensitive data can leak that data. Cross-tenant leakage: a multi-tenant vector store without strict isolation returns one customer's documents in another's queries, or an attacker crafts queries that surface data they should not see. Retrieval poisoning: an attacker plants documents through any ingestion path so their content is retrieved and injected into the prompt, an indirect prompt injection through the knowledge base. Shown for defensive testing.

SOURCES

[1] OWASP Top 10 for LLM Applications (2025)

[2] Morris et al., Text Embeddings Reveal (Almost) As Much As Text

Test your AI application before an attacker does.

securelayer7.net/learn/ai-security/vector-and-embedding-weaknesses

Open online