Intelligence Artificielle

RAG en 2026 : Guide Architecture, Vec

7 décembre 2025 • Mis à jour le 12 juillet 2026 • 20 min de lecture • 8225 mots

Télécharger le PDF

Le RAG (Retrieval Augmented Generation) combine la recherche documen
réponses factuelles et sourcées, réduisant les hallucinations.

TL;DR — EN RÉSUMÉ

RAG (Retrieval Augmented Generation) 2026 : architecture, embeddings, vectorbase, chunking, hybric

Définition et origines

Plutôt que de s'appuyer uniquement sur les connaissances encodées durant le pr
récupère dynamiquement des documents pertinents depuis une base de connais
LLM pour générer une réponse informée et factuelle. RAG (Retrieval Augmented G
Architecture | Guide Complet 2025. Expert en cybersécurité et.

📋 EN BREF

- Architecture technique et principes de fonctionnement du modèle
- Cas d'usage concrets en cybersécurité et performance mesurée
- Limites, biais potentiels et considérations éthiques
- Guide d'implémentation et ressources recommandées

Formule Conceptuelle du RAG

Réponse = LLM(Question + Documents_Récupérés)

Au lieu de : Réponse = LLM(Question)