

Le RAG expliqué simplement

Le RAG (retrieval augmented generation, en français génération augmentée par récupération) est une méthode qui fait d'abord chercher une IA dans vos documents, puis lui fait rédiger sa réponse à partir des passages trouvés. Ce guide explique son fonctionnement sans jargon, ce qui le distingue des autres approches, ses usages en entreprise, ses limites et la manière de l'évaluer.



PARIS • SALLE DES CATALOGUES DE LA BIBLIOTHÈQUE MAZARINE • MARIE-LAN NGUYEN • CC BY 2.0 FR

- 01 Qu'est-ce que le RAG en IA ?
 - 02 Une recherche documentaire suivie d'une réponse
 - 03 Comment fonctionne un RAG, étape par étape ?
 - 04 RAG, fine-tuning ou long contexte : quelle différence ?
 - 05 Quels cas d'usage du RAG en entreprise ?
 - 06 Quels documents donner à un RAG et comment les tenir à jour ?
 - 07 Une citation n'est pas une preuve suffisante
 - 08 Comment évaluer un RAG avant de le déployer ?
 - 09 Pourquoi un RAG peut-il encore halluciner ?
 - 10 Protéger les droits jusque dans la réponse
 - 11 Combien coûte un RAG : quels postes prévoir ?
 - 12 Qu'est-ce que le RAG agentique ?
 - 13 Le choix Blink : une réponse reliée à une source et à une action
-

Qu'est-ce que le RAG en IA ?

Le RAG est une technique qui relie un modèle de langage à une collection de documents choisie : à chaque question, le système retrouve les passages utiles et les transmet au modèle, qui s'en sert pour répondre. Un modèle de langage, comme ceux qui font fonctionner ChatGPT ou Claude, a appris à rédiger à partir d'immenses volumes de textes. Il ne connaît pas pour autant vos procédures, vos contrats ou vos fiches produits, et ses connaissances s'arrêtent à une date donnée. Le RAG lui apporte, au moment de la question, l'information qui lui manque.

Le terme vient d'un article de recherche publié en 2020 par Patrick Lewis et ses coauteurs, présenté à la conférence NeurIPS. Les auteurs combinaient deux mémoires : celle que le modèle a acquise pendant son entraînement, et un index de documents interrogé à la demande, en l'occurrence Wikipédia. Dans une entreprise, l'index contient vos propres documents. Vous posez une question en langage courant, par exemple « que faire si un client conteste une livraison ? », et obtenez une réponse rédigée avec les passages qui la justifient. La qualité de ces documents limite d'abord celle des réponses.

Une recherche documentaire suivie d'une réponse

Un RAG enchaîne trois temps : comprendre la question, retrouver les passages pertinents que la personne est autorisée à consulter, puis rédiger une réponse qui conserve les références de ces passages. Le modèle ne lit pas toute votre bibliothèque à chaque demande. Il reçoit une sélection de quelques extraits, jugés les plus proches de la question, avec la consigne de s'appuyer dessus. Ce que la recherche n'a pas remonté n'existe pas pour lui au moment de répondre.

Cette séparation explique la plupart des erreurs. Si la recherche manque le bon passage, la meilleure rédaction ne peut pas le réparer : le modèle répondra à côté ou complétera avec ses connaissances générales. Si deux documents se contredisent, il faut savoir lequel fait référence. Un RAG assemble donc deux problèmes distincts, retrouver la bonne information puis la restituer fidèlement, et chacun se teste séparément.



Repère pédagogique à adapter à votre organisation.

03

Comment fonctionne un RAG, étape par étape ?

Un RAG fonctionne en deux phases : une préparation des documents, faite à l'avance et refaite à chaque mise à jour, puis une recherche suivie d'une génération, à chaque question. Les termes techniques qui reviennent dans toutes les présentations désignent des opérations assez simples. Les voici dans l'ordre où elles interviennent, avec ce qu'elles changent pour la qualité des réponses.

Chaque étape se règle. Anthropic a par exemple publié en 2024 une méthode qui ajoute à chaque passage une courte description de son contexte, le document et la section dont il vient, et qui réduisait les échecs de recherche dans ses essais. OpenAI décrit son outil de recherche dans des fichiers comme combinant recherche par le sens et recherche par mots-clés. Ces réglages relèvent de la conception ; ce qui compte pour vous, c'est de les juger sur vos propres questions.

- ✓ **Extraction** : le texte est tiré des fichiers (PDF, pages web, documents bureautiques). Un PDF scanné doit d'abord passer par une reconnaissance de caractères pour devenir du texte.

Découpage (en anglais chunking) : chaque document est coupé en passages de taille

- ✓ limitée. Trop court, un passage sépare une règle de son exception ; trop long, il noie l'information utile.

Embeddings : un modèle transforme chaque passage en une liste de nombres qui

- ✓ représente son sens. Deux textes de sens proche obtiennent des listes proches, même sans mot en commun.

Base vectorielle : une base de données qui stocke ces listes de nombres avec le passage, sa

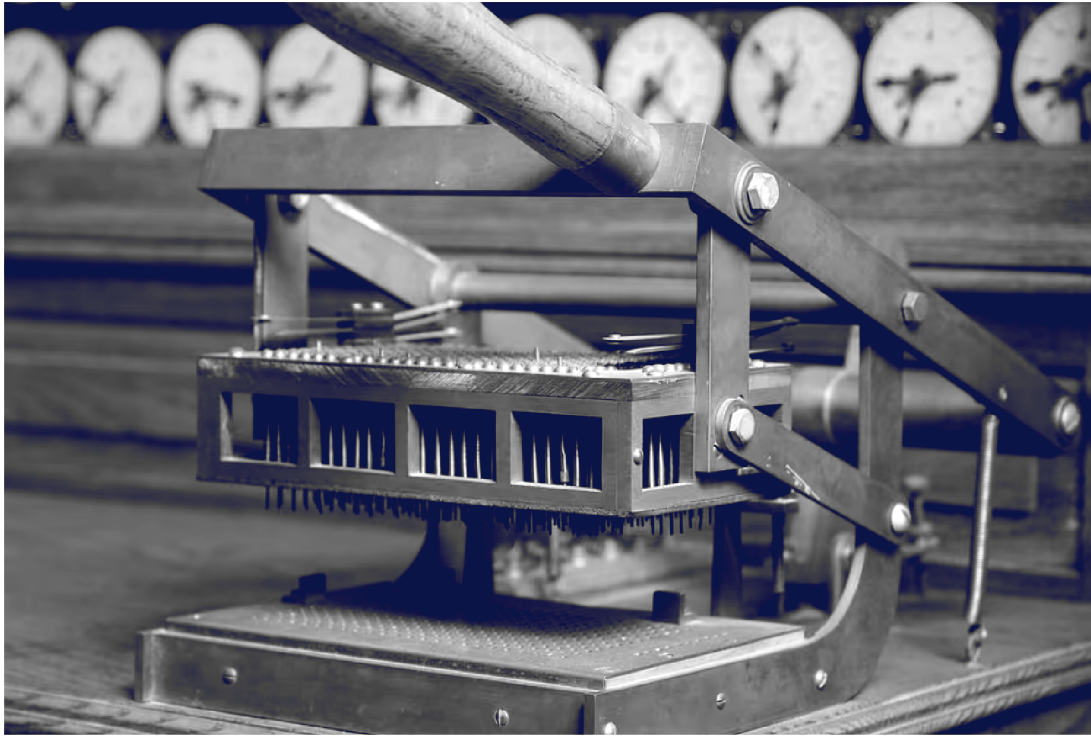
- ✓ source et ses métadonnées (date, auteur, droits), et retrouve vite les plus proches d'une question.

Recherche : la question est elle aussi transformée en liste de nombres, puis comparée aux

- ✓ passages. On l'associe souvent à une recherche par mots-clés, plus sûre pour une référence produit ou un code d'erreur.

Reclassement : un second tri, facultatif, réordonne les passages trouvés pour ne garder que les plus pertinents.

- ✓ **Génération** : le modèle reçoit la question, les passages retenus et une consigne (s'appuyer sur les sources, les citer, dire quand l'information manque), puis rédige la réponse.



LE LECTEUR DE CARTES D'UNE TABULATRICE HOLLERITH

MARCIN WICHARY · CC BY 2.0

04

RAG, fine-tuning ou long contexte : quelle différence ?

Le RAG apporte de l'information au modèle au moment de la question, le fine-tuning modifie le modèle lui-même, et le long contexte consiste à lui transmettre directement des documents entiers. Le fine-tuning, ou réentraînement ciblé, ajuste un modèle existant sur des exemples pour changer sa manière de répondre : un ton, un format, une classification. Il convient mal pour des faits qui changent, car chaque mise à jour demande un nouvel entraînement, et il ne permet pas de montrer d'où vient une affirmation.

Le long contexte est devenu une vraie option, car les modèles récents acceptent de très longs textes en une seule demande. En septembre 2024, Anthropic indiquait qu'une base de connaissances de moins de 200 000 tokens, soit environ 500 pages, pouvait être placée en entier dans la consigne, sans RAG. Un token est un morceau de mot, l'unité dans laquelle les modèles comptent le texte. Au-delà, ou quand chaque question ne concerne qu'une petite partie du fonds, la recherche préalable reste plus économe et plus traçable.

Reste la recherche classique, par mots-clés, dans votre gestion documentaire ou votre intranet. Elle suffit souvent quand les personnes connaissent le nom du document et veulent l'ouvrir. Le RAG devient intéressant quand elles posent des questions en langage courant, ne connaissent pas le vocabulaire des documents ou doivent rapprocher plusieurs passages.

Approche	Ce qu'elle fait	Quand elle convient	Point de vigilance
Recherche classique	Retrouve des documents à partir de mots-clés	La personne sait ce qu'elle cherche et veut le fichier d'origine	Ne rédige pas de réponse ; dépend des bons mots
RAG	Retrouve des passages, puis rédige une réponse sourcée	Questions en langage courant sur un fonds documentaire qui évolue	Limité par la qualité de la recherche et des documents
Long contexte	Transmet des documents entiers au modèle à chaque demande	Petit corpus, analyse complète d'un dossier	Volume transmis, donc coût, à chaque question
Fine-tuning	Ajuste le modèle sur des exemples	Changer un style, un format, une classification répétitive	Ne met pas les faits à jour ; aucune source affichée

Travailler avec Claude sur vos documents ↗

05

Quels cas d'usage du RAG en entreprise ?

Le RAG est utile partout où des personnes passent du temps à chercher une réponse qui existe déjà par écrit. Le critère n'est pas la taille de l'entreprise mais la répétition : les mêmes questions reviennent, les réponses sont dispersées dans plusieurs fichiers, et une erreur de lecture a un coût. Voici les familles d'usage les plus courantes, de la plus simple à la plus exigeante.

Dans chaque cas, commencez par une famille de questions précise plutôt que par « tous nos documents ». Un périmètre restreint permet de repérer les manques et de vérifier les réponses. Le RAG peut aussi devenir une brique d'un système plus large : un agent IA qui cherche une information, puis prépare une action, comme un brouillon de réponse ou une fiche complétée, soumise à validation.

- ✓ Base documentaire interne : retrouver une règle RH, une note de service, la marche à suivre dans un logiciel.
- ✓ Support client : préparer une réponse à partir des conditions générales, des notices et des réponses déjà validées, relue avant envoi.
- ✓ Juridique et contrats : localiser une clause, comparer deux versions, signaler ce qui manque. La décision reste à un juriste.
- ✓ Qualité et conformité : interroger un référentiel, des procédures et des fiches d'écart en citant la version applicable.

- ✓ Procédures techniques : retrouver la consigne d'une machine ou d'un logiciel à partir de la description d'un symptôme.
- ✓ Accueil des nouveaux arrivants : répondre aux questions de démarrage sans solliciter toujours les mêmes collègues.

Exemple illustratif : un service après-vente

Un technicien décrit un symptôme : « l'écran affiche E42 au démarrage ». La recherche par mots-clés retrouve le code dans la notice ; la recherche par le sens retrouve une note interne qui décrit le même défaut avec d'autres mots. La réponse propose la procédure, cite les deux passages et signale que la note interne concerne une version antérieure du produit. Le technicien décide de la suite.

Des agents IA connectés à vos sources ↗

Cinq cas d'usage concrets d'agents IA ↗

06

Quels documents donner à un RAG et comment les tenir à jour ?

Donnez à un RAG les documents qui font référence pour la famille de questions choisie, avec pour chacun un propriétaire, une date et une version. Ajouter des fichiers sans responsable augmente le bruit au lieu d'améliorer les réponses : un brouillon périmé ou une ancienne grille tarifaire seront retrouvés aussi bien que la version en vigueur. Écartez les doublons, identifiez clairement les archives et décidez quel document l'emporte quand deux sources se contredisent.

La forme compte autant que le fond. Un PDF lisible à l'œil n'est pas toujours exploitable : la documentation d'Anthropic précise par exemple qu'un PDF scanné sans texte extractible ne peut pas être cité par sa fonction de citations. Les tableaux, les colonnes, les schémas et les notes de bas de page se perdent facilement à l'extraction. Un titre explicite, une date visible et des sections nettes aident à la fois la recherche et la personne qui vérifiera la source.

La mise à jour suit la vie du document : quand une procédure change, la nouvelle version doit remplacer l'ancienne dans l'index, pas s'y ajouter. Prévoyez un déclencheur, comme le dépôt dans un dossier, et une personne chargée de dire ce qui est à jour. Sans ce responsable, l'index vieillit en silence.

- ✓ Une source de référence pour chaque sujet.
- ✓ Un responsable chargé des mises à jour.
- ✓ Une règle pour retirer les anciennes versions de l'index.
- ✓ Des droits d'accès connus avant l'import.

Préparer des données propres avant d'automatiser ↗

Une citation n'est pas une preuve suffisante

Afficher la source sous chaque réponse est indispensable, mais ne suffit pas : une réponse peut citer le bon document et en tirer une conclusion fausse. Microsoft distingue d'ailleurs, dans ses guides d'évaluation, la fidélité aux sources (la réponse s'appuie-t-elle seulement sur les passages fournis ?) et l'exactitude (la réponse est-elle juste ?). Un système peut être fidèle et inexact, par exemple en rapprochant deux phrases qui n'avaient aucun lien entre elles.

Les fonctions de citation des éditeurs rendent la vérification plus rapide. La documentation d'Anthropic décrit des citations qui renvoient au passage exact, avec le numéro de page pour un PDF ; OpenAI joint à la réponse de son outil de recherche des références aux fichiers utilisés. Côté interface, le bon réflexe est de pouvoir ouvrir le passage cité dans sa version, et de distinguer ce que dit le document de ce que le modèle en déduit.

Demandez enfin au système de savoir se taire. Une réponse « je n'ai pas trouvé cette information dans les documents disponibles » est plus utile qu'une supposition bien rédigée. Cette consigne se teste avec des questions dont la réponse ne figure pas dans le corpus : le système doit le reconnaître au lieu de combler le vide.

Exemple illustratif : une procédure de retour

La documentation autorise un retour sous conditions et prévoit une exception pour les produits personnalisés. Une réponse qui ne cite que le délai de retour est exacte en apparence, mais incomplète. Le test doit vérifier que l'exception apparaît quand la question porte sur un produit personnalisé, avec un lien vers la version de référence de la procédure.

Comment évaluer un RAG avant de le déployer ?

On évalue un RAG avec un jeu de questions de test dont on connaît la réponse attendue et le passage qui la justifie. Rassemblez des questions réellement posées par les équipes, plutôt que des questions inventées par ceux qui ont conçu le système. Pour chacune, notez la réponse correcte, le document de référence et les personnes autorisées à la recevoir. Ajoutez des questions ambiguës et des questions sans réponse dans le corpus.

Évaluez ensuite séparément la recherche et la réponse. La recherche a-t-elle remonté le bon passage ? La réponse s'appuie-t-elle sur lui, est-elle complète, est-elle juste ? Les réponses d'un modèle varient d'une exécution à l'autre, et Microsoft conseille de raisonner sur une fourchette plutôt que sur un score unique. Rejouez le même jeu après chaque changement de document, de réglage ou de modèle, avec une personne du métier capable de juger si une formulation exacte reste trompeuse en pratique.

Critère	Question à se poser	Piste si le résultat est mauvais
Recherche	Le bon passage figure-t-il parmi ceux transmis au modèle ?	Revoir le découpage, les métadonnées, la recherche par mots-clés
Fidélité	La réponse s'appuie-t-elle seulement sur les passages fournis ?	Renforcer la consigne, exiger des citations
Exactitude	La réponse est-elle juste et complète ?	Vérifier le document source, ajouter l'exception manquante
Abstention	Le système dit-il quand l'information manque ?	Ajouter des questions hors corpus au jeu de test
Droits	La réponse n'utilise-t-elle que des documents autorisés pour cette personne ?	Corriger le filtrage avant la recherche

Pourquoi un RAG peut-il encore halluciner ?

Un RAG réduit les inventions sans les supprimer, parce que le modèle reste un générateur de texte qui peut mal lire, mal relier ou combler un silence. Une hallucination désigne une affirmation fausse présentée avec assurance. Dans un RAG, elle vient le plus souvent d'une recherche qui a manqué le bon passage, d'un document lui-même faux ou périmé, ou d'une question à laquelle le corpus ne répond pas et que le modèle traite quand même.

Ces limites ne disqualifient pas la méthode ; elles dessinent les garde-fous. Restreindre le corpus aux sources de référence, imposer les citations, autoriser l'abstention et garder une validation humaine sur ce qui engage l'entreprise : une réponse envoyée à un client, une position juridique, une décision qualifiée. Plus la conséquence d'une erreur est lourde, plus la relecture doit être systématique.

- ✓ Questions de synthèse globale (« quels sont tous les contrats qui... ») : quelques passages ne suffisent pas à couvrir l'ensemble du fonds.
- ✓ Calculs et totaux chiffrés : mieux confiés à une requête dans le logiciel qui détient les données qu'à la lecture de passages.
- ✓ Documents contradictoires : sans règle de priorité, le modèle choisit l'un des deux ou mélange les deux.
- ✓ Injection de consignes : un document peut contenir un texte rédigé pour détourner le modèle. Microsoft classe ce risque, dit d'injection indirecte, parmi les menaces propres aux RAG.

Protéger les droits jusque dans la réponse

Un RAG doit respecter les droits d'accès d'origine : une personne ne doit recevoir que des passages issus de documents qu'elle a déjà le droit d'ouvrir. Indexer un dossier confidentiel dans la même base qu'une procédure générale ne doit pas le rendre lisible par tous. Le filtrage intervient avant que les passages soient transmis au modèle ; Microsoft recommande de l'appliquer à chaque recherche, y compris quand c'est un agent qui interroge la base.

Pensez aussi aux copies : index, base vectorielle, historiques de conversations, journaux techniques et sauvegardes. Retirer un fichier du dossier partagé ne le retire pas forcément de tous ces composants. Définissez qui peut lire les journaux, combien de temps ils sont conservés et comment un document est supprimé partout. Ces choix sont plus simples à poser au départ qu'à rattraper une fois le système en service.

Quand les documents contiennent des données personnelles, le RGPD s'applique. Dans ses questions-réponses sur l'IA générative, la CNIL précise que l'organisme qui connecte un

système à sa propre base de connaissances est responsable de ce traitement lorsqu'elle contient des données personnelles. Pour des données sensibles, elle juge généralement préférable un déploiement sur site. Les lieux d'hébergement et les services autorisés se décident au cadrage ; les conditions d'utilisation des données par chaque service IA se vérifient pour la configuration retenue.

IA et RGPD : les règles à connaître ↗

Sécurité et protection de vos données ↗

11

Combien coûte un RAG : quels postes prévoir ?

Le coût d'un RAG dépend du volume de documents, du nombre de questions et du niveau de contrôle attendu, et il se décompose en postes que vous pouvez estimer séparément. Aucun prix unique n'a de sens : un assistant sur une centaine de procédures et un système branché sur plusieurs logiciels, avec des droits fins, n'ont presque rien en commun. Méfiez-vous d'un chiffrage qui ne parle que de licence ou d'abonnement à un outil.

Pour juger l'intérêt, rapportez ces postes au temps passé aujourd'hui à chercher et à vérifier, sans oublier le temps de relecture qui restera nécessaire. Posez cette estimation comme une hypothèse explicite, à confirmer sur un premier périmètre, et non comme une promesse. Le calculateur de Blink aide à rendre ce raisonnement visible, avec vos propres volumes.

- ✓ Préparation documentaire : tri, mise à jour et conversion des fichiers, souvent du temps interne à prévoir.
- ✓ Conception et développement : connexion aux sources, découpage, recherche, consigne, interface et gestion des droits.
- ✓ Indexation : calcul des embeddings à l'import puis à chaque mise à jour, facturé à l'usage par les fournisseurs qui le proposent.
- ✓ Stockage et hébergement : base vectorielle, fichiers d'origine, journaux.
- ✓ Usage du modèle : chaque question envoie la question et les passages retenus ; la facturation se fait généralement au volume de texte traité.
- ✓ Évaluation et maintenance : jeu de test rejoué, documents suivis, réglages revus quand un modèle change.

Estimer le temps passé à chercher ↗

12

Qu'est-ce que le RAG agentique ?

Le RAG agentique désigne un RAG piloté par un agent IA : au lieu de suivre un enchaînement fixe, une recherche puis une réponse, le modèle décide lui-même s'il doit

chercher, où, et combien de fois. Le terme n'a pas de définition normalisée. C'est le sens que retient par exemple la documentation d'architecture de Microsoft, qui décrit la recherche comme un outil que l'agent appelle à la demande, puis dont il évalue le résultat avant de continuer.

Cette approche sert les questions en plusieurs étapes : retrouver une liste dans un catalogue, puis vérifier chaque élément dans une autre source, ou chercher puis agir, comme préparer un retour après avoir lu l'historique d'une commande. Elle a un prix : chaque étape ajoute du délai, de la consommation et des occasions d'erreur. Microsoft conseille le RAG classique quand une seule recherche suffit. Dès qu'un agent peut agir, la question des validations humaines devient centrale.

Comprendre les agents IA ↗

13

Le choix Blink : une réponse reliée à une source et à une action

Notre position : un RAG utile en entreprise montre d'où vient chaque réponse et s'arrête là où une décision humaine est attendue. Une synthèse sans document consultable fait gagner du temps à la lecture et en fait perdre au moment de vérifier. Nous préférons une réponse courte, citée, qui sait dire ce qu'elle ignore.

Le cadrage part d'un trio : une famille de questions, un ensemble de sources de référence et une personne capable de juger les réponses. Nous écrivons ensuite du code sur mesure, branché sur les outils où vos documents vivent déjà, plutôt que de les déplacer dans une plateforme de plus. Nous cherchons à reprendre les droits d'accès existants plutôt qu'à en recréer. Les lieux d'hébergement et les services IA autorisés sont définis avec vous.

Quand le RAG alimente une automatisation, le portail livré à chaque client permet de voir ce qui tourne, de trier ce que l'automatisation a trouvé et de valider avant ce qui part. La validation humaine porte sur les actions et droits convenus au cadrage. Après la mise en place, l'abonnement mensuel couvre la maintenance et un point de suivi : le moment de rejouer les questions de test, de signaler les documents à mettre à jour et de regarder ce que les évolutions des modèles changent.

Le portail Blink pour voir et valider ↗

Automatisations sur mesure branchées sur vos outils ↗

Cadrer le projet par un audit de processus ↗

Sources et repères

arXiv : Lewis et al., Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks (2020)

Anthropic : Contextual Retrieval, découpage et recherche dans un RAG

Anthropic : fonction de citations de l'API Claude

OpenAI : outil de recherche dans des fichiers (file search)

Microsoft Learn : évaluer une solution RAG de bout en bout

Microsoft Learn : RAG agentique

Microsoft Learn : droits d'accès et filtrage dans un RAG

CNIL : questions-réponses sur l'utilisation d'un système d'IA générative

QUESTIONS

Questions fréquentes

Votre question n'y est pas ? Posez-la pendant l'échange gratuit de trente minutes.

Poser ma question →

01 Que veut dire RAG en intelligence artificielle ?

RAG est l'abréviation de retrieval augmented generation, que l'on traduit par génération augmentée par récupération. Avant de répondre, le système récupère des extraits pertinents dans une collection de documents ; le modèle de langage rédige ensuite sa réponse à partir de ces extraits. L'expression vient d'un article de recherche de 2020 et désigne aujourd'hui la plupart des assistants qui répondent à partir de documents d'entreprise.

02 Le RAG entraîne-t-il l'IA sur nos documents ?

Non, pas en lui-même. Le RAG transmet des extraits au modèle au moment de chaque question ; il ne modifie pas le modèle et ne lui fait pas apprendre durablement votre bibliothèque. Ce que le fournisseur fait ensuite des données reçues dépend de son contrat et de vos paramètres. Ces conditions se vérifient pour le service et l'offre retenus, avant d'envoyer le moindre document interne.

03 **Le RAG supprime-t-il les hallucinations ?**

Non. Il les réduit en donnant au modèle des sources à suivre, mais le modèle peut encore choisir un mauvais passage, mal l'interpréter ou combler un manque. Des citations visibles, la consigne de signaler l'absence d'information et un jeu de questions de test rejoué régulièrement limitent ce risque. Pour les réponses qui engagent l'entreprise, une relecture humaine reste nécessaire.

04 **Faut-il une base vectorielle pour faire du RAG ?**

Pas obligatoirement. Une base vectorielle retrouve des passages par le sens, même formulés autrement que la question. Mais une recherche par mots-clés, ou une base de données interrogée directement, peut suffire pour des références, des codes ou des fiches structurées. Beaucoup de systèmes combinent les deux. Le bon choix se constate sur vos propres questions, pas sur une fiche technique.

05 **Comment créer un chatbot sur ses propres documents ?**

Trois voies existent. Les assistants professionnels permettent d'ajouter des fichiers à un espace de travail, ce qui suffit souvent pour un usage individuel. Les API des éditeurs proposent des outils de recherche dans des fichiers et de citation pour bâtir sa propre interface. Un développement sur mesure devient utile pour des droits d'accès fins, plusieurs logiciels à brancher ou des actions à déclencher. Dans tous les cas, commencez par un corpus restreint et des questions de test.

06 **Peut-on utiliser des PDF scannés, des tableaux ou des images ?**

Oui, avec précautions. Un PDF scanné doit passer par une reconnaissance de caractères pour devenir du texte exploitable. Les tableaux perdent souvent leurs colonnes à l'extraction, et l'information portée par un schéma ou une image n'est pas toujours lue. Vérifiez sur quelques documents représentatifs que le texte extrait est fidèle avant d'importer tout un fonds documentaire.

07 **À quelle fréquence mettre à jour les documents d'un RAG ?**

Au rythme où ils changent dans la réalité. Une procédure modifiée doit remplacer l'ancienne version dans l'index dès son entrée en vigueur, et non s'y ajouter. Le plus sûr est une mise à jour déclenchée par le dépôt ou la validation du document, avec une personne responsable du contenu. Un contrôle périodique des documents les plus consultés complète le dispositif.

08 **Le RAG est-il adapté à une petite entreprise ?**

Oui, si les mêmes questions reviennent souvent et que les réponses existent par écrit. La taille compte moins que la répétition et le coût d'une erreur. Pour quelques dizaines de documents, déposer les fichiers dans un assistant peut suffire. Un RAG construit se justifie quand le fonds grossit, change souvent ou doit respecter des droits d'accès différents selon les personnes.

09 **Le RAG est-il compatible avec le RGPD ?**

Il peut l'être, mais la conformité ne vient pas de la technique. Si les documents contiennent des données personnelles, l'organisme qui connecte l'IA à sa base de connaissances est responsable de ce traitement : finalité, minimisation, droits d'accès, durée de conservation, information des personnes et choix des prestataires. Notre guide sur l'IA et le RGPD détaille ces points un par un.

ET MAINTENANT ?

Ce guide pose les bases. Votre cas mérite trente minutes.

Le premier échange avec Blink est gratuit et sans engagement : trente minutes en visio pour examiner une tâche de votre équipe et ce qu'elle demanderait vraiment.

fr-blink.com/contact
contact@fr-blink.com

BLINK · PAUL ROBERT ET BASTIEN DOUBLET · AUTOMATISATIONS ET AGENTS IA SUR MESURE, PARTOUT EN FRANCE