

Le contexte dans les IA génératives

Ingénierie du contexte ■ Compréhension ■ Fenêtre et RAG

1. QU'EST-CE QUE LE CONTEXTE ?

Le contexte est l'ensemble des informations disponibles pour une IA au moment où elle doit comprendre une demande et produire une réponse. Il lui permet de ne pas considérer une phrase isolément, mais de l'interpréter en fonction d'éléments complémentaires.

- **Demande isolée** : Une phrase seule contient peu d'informations ; l'IA doit déterminer elle-même le ton ou l'objectif.
- **Demande contextualisée** : L'ajout d'informations réduit les ambiguïtés et oriente le résultat.
- **Portée élargie** : Le contexte ne se limite pas au prompt : il comprend l'historique, les fichiers, les consignes et les outils.

2. DE QUOI EST CONSTITUÉ LE CONTEXTE ?

Le contexte combine plusieurs dimensions d'importance variable :

- **La question / tâche** : Expliquer, rédiger, résumer, analyser, créer...
- **L'objectif et le destinataire** : Précise l'usage final et adapte le niveau de connaissance visé.
- **Les contraintes et données** : Longueur, ton, format, chiffres, tableaux et documents sources (PDF, images).
- **L'historique** : Les échanges précédents qui structurent la conversation.

3. FONCTIONNEMENT ET CONTEXTE CONVERSATIONNEL

- **Traitement par l'IA** : Le modèle relie les mots entre eux, désambiguïse les expressions et influence la génération sans « comprendre » au sens humain.
- **Contexte cumulatif** : La conversation s'enrichit au fil des échanges (préférences, corrections, références implicites comme « *modifie le deuxième paragraphe* »).

4. FENÊTRE DE CONTEXTE ET ARCHITECTURE RAG

- **Fenêtre de contexte** : Exprimée en *tokens* (unités de texte), elle possède une limite maximale d'entrée et de sortie. Un contexte trop long dégrade la prise en compte de l'information.
- **Le RAG (Retrieval-Augmented Generation)** : Recherche d'informations pertinentes dans une source documentaire externe avant de les injecter comme contexte pour la génération :

1. Question → 2. Recherche documentaire → 3. Sélection des passages → 4. Génération ciblée

5. QUALITÉ, HALLUCINATIONS ET BONNES PRATIQUES

- **Contexte ≠ vérité** : Une information fournie n'est pas automatiquement vérifiée ; une donnée fausse ou contradictoire induit l'IA en erreur.
- **Méthode de construction (Context Engineering)** : Sélectionner, structurer, hiérarchiser et actualiser les informations utiles.

1. Définir l'objectif	4. Préciser le destinataire
2. Décrire la situation	5. Définir les contraintes
3. Fournir les données	6. Spécifier le format

6. EXEMPLE COMPARATIF : SANS CONTEXTE VS AVEC CONTEXTE

Pour illustrer l'impact de l'ingénierie du contexte, voici comment une même requête générale change du tout au tout selon les informations fournies :

- **Sans contexte** : « *Rédige un article sur la sécurité informatique.* »
 - **Résultat** : Une réponse générique, superficielle, brassant des généralités sur les mots de passe et les pare-feu, sans public cible défini.
- **Avec contexte** : « *Rédige une note de synthèse de 300 mots (format Markdown) à destination d'un comité de direction, expliquant les risques du phishing ciblé et proposant 3 actions prioritaires pour le Q4.* »
 - **Résultat** : Une production ultra-ciblée, directement exploitable, respectant les contraintes de format, de longueur et de ton.

💡 L'IDÉE ESSENTIELLE

« Ne demandez pas seulement à l'IA quoi faire. Donnez-lui les informations dont elle a besoin pour comprendre dans quel contexte elle doit le faire. »

Maîtriser le contexte (Context Engineering) est une compétence fondamentale pour structurer et fiabiliser l'utilisation des IA génératives.