

Poser une question (flux classique)

Poser une question (flux classique)

« Dernière mise à jour : 2026-05-10

Étapes

1. **Va sur** `/play` depuis le Home en cliquant sur un jeu.
2. **Coche les extensions actives** dans le header (les non cochées sont ignorées par le retrieval).
3. **Tape ta question** dans le composer en bas.
4. **Envoie** (Entrée, ou bouton). L'oracle commence à streamer la réponse.
5. **Lis et clique sur les citations** pour vérifier la source dans le PDF.
6. **Vote pouce** ↑ ↓ + commentaire si la réponse t'a aidé ou pas — tout est enregistré pour `/admin/feedback`.

Ce que tu vois pendant la génération

- **Phases verbalisées** dans une bulle "thinking" : « L'Oracle pèse votre question », « plonge dans le grimoire », « rédige sa réponse »

- **Tokens streamés** dans la bulle finale au fur et à mesure
- **Fallback hors-ligne** : si la connexion SSE casse, l'animation passe à « L'Oracle finalise hors-ligne... » et l'app récupère la réponse via polling sur `GET /api/ask/:questionId` (jusqu'à 45s).

Que faire si la réponse semble fausse

1. Vote pouce bas + commentaire détaillé.
2. Va sur `/admin/feedback`, filtre par ton jeu + vote down, ouvre le détail.
3. Clique "Copier (md)" pour avoir un dump markdown des diagnostics (chunks retrouvés, scores, HyDE, timings).
4. Si c'est un problème de retrieval (mauvais chunks), vérifie que le PDF est bien indexé (count chunks > 0) et que la question n'est pas trop floue.
5. Si c'est un problème de réponse Claude malgré bons chunks, l'Oracle a hallu — pas grand-chose à faire à part voter pour qu'il apprenne (le feedback alimente ton banc d'éval RAG).

Limites

- 500 caractères max par question
- 5 turns d'historique pris en compte
- 1 image PNG injectée au max (pour limiter la latence vision)
- Heartbeat SSE toutes les 8s pour ne pas se faire timeout par NPM

Revision #1

Created 2026-05-10 15:19:47 UTC by thymon

Updated 2026-05-10 15:19:47 UTC by thymon