

Statut d'implémentation

État réel du projet Hallucide au 4 juillet 2026, vérifié dans le code source.

Ce qui est implémenté et testé

Bloc	État	Preuve
Pipeline complet (décomposition, récupération, vérification, triage, validation, audit)	Complet	220 tests automatisés (make test)
Vérification verbatim avec re-contrôle du cycle de vie (texte abrogé jamais authentifié)	Complet	Tests dédiés + scénario réel
Plancher de risque incontournable	Complet	Tests de non-contournement
Validation humaine des résultats à risque élevé	Complet	Registre + journalisation testés
Journal de conformité rejouable, sans question ni identité	Complet	Garde-fou vérifié par assertion
Routes structurées : article de code, question parlementaire, commissions d'un député (avec rôle et dates), donnée tabulaire, recherche libre	Complet	Tests par route
Trois fournisseurs LLM interchangeables : Claude (défaut), Mistral, Gemini	Complet	Testés en direct
Front de chat : prose annotée, indice de confiance, lignes datées triées, liens vers les documents officiels	Complet	Déployé sur http://141.11.165.40:8770
Mode document (notes et synthèses vérifiées affirmation par affirmation)	Complet	16 tests dédiés

Vérifié en conditions réelles

Deux scénarios joués en direct contre les sources officielles :

- Prémisse fausse** : question sur une teneur inexistante d'une question orale. Le système a récupéré le vrai texte officiel et n'a rien inventé (NO_ANSWER), avec demande de validation humaine.
- Passage authentique mais hors sujet** : un article exact et opposable présenté en réponse à une question à laquelle il ne répond pas. Chaque

élément était vrai isolément ; le plancher de risque a bloqué la publication automatique et rendu la main à l'humain.

Limites connues

- La mesure de qualité inter-annotateurs est codée et testée, mais il manque de vraies annotations humaines pour l'exercer.
- La détection d'une pertinence trompeuse (un passage vrai qui semble répondre mais ne répond pas) est structurellement hors de portée d'un contrôle déterministe seul : c'est précisément le cas que la validation humaine couvre.
- La couche source dépend du serveur MCP communautaire pendant la démo ; en cas de coupure, l'interface affiche « source indisponible » plutôt qu'un résultat simulé.