AI TOOLS WIKI · MÉMO

Claude-Mem *mémo*

L'essentiel pour utiliser **Claude-Mem**, le plugin de mémoire persistante de Claude Code. Il capture automatiquement ce que fait Claude pendant une session, le résume, et réinjecte le contexte utile dans les sessions suivantes. Au quotidien, il n'y a presque rien à taper : il suffit de poser des questions sur l'historique.

Famille Mémoire et contexte · **Éditeur** thedotmack · **Licence** Apache-2.0
Note du wiki 10/15 · **Commandes vérifiées le** 25 septembre 2026
Fiche, notes et avis aitoolswiki.viellard.fr/outils/memoire-et-contexte/claude-mem

Installation

`npx claude-mem install`	Installation recommandée (hooks + service worker)
`/plugin marketplace add thedotmack/claude-mem` `/plugin install claude-mem`	Ou depuis Claude Code
`npx claude-mem install --ide opencode` `npx claude-mem install --ide antigravity`	Autres agents

Redémarrer ensuite Claude Code. Attention : `npm install -g claude-mem` n'installe que la bibliothèque, sans les hooks ni le service.
L'installeur propose de créer un compte claude-mem (observateur hébergé, gratuit 30 jours). Pour l'éviter, passer `--provider` explicitement ou définir `CLAUDE_MEM_ONLINE_OPTIN=false`.

Utilisation au quotidien

Action	Comment
Retrouver un travail passé	Poser la question en langage naturel : « Quels bugs a-t-on corrigés la dernière fois ? », « Comment avions-nous implémenté l'authentification ? ». La skill `mem-search` se déclenche seule.

Utilisation au quotidien (suite)

Action	Comment
Exclure une information de la mémoire	L'entourer de balises `<private>…</private>` dans le prompt. Claude la voit pendant la session, mais elle n'est jamais stockée.
Voir la mémoire en direct	Ouvrir la Web Viewer UI à l'URL du worker affichée au démarrage.
Repartir avec un contexte propre	`/clear` : la session redémarre et Claude-Mem réinjecte le contexte récent.
Diagnostiquer un problème	Décrire le souci à Claude : la skill de dépannage se déclenche automatiquement.

Outils MCP de recherche (workflow en 3 couches)

Claude les utilise lui-même, mais les connaître aide à formuler les demandes :

Outil	Ce qu'il fait	Coût
`search`	Renvoie un index compact de résultats avec leurs IDs, filtrable par type, date, projet.	~50–100 tokens par résultat
`timeline`	Montre ce qui se passait autour d'une observation ou d'une requête.	faible
`get_observations`	Charge le détail complet, uniquement pour les IDs retenus.	~500–1 000 tokens par résultat

Filtrer avec `search` avant de charger le détail économise environ 10 fois plus de tokens que de tout lire.

Service worker

Depuis le dossier du plugin (`~/.claude/plugins/marketplaces/thedotmack`) :

`npm run worker:status`	État du service.
`npm run worker:restart`	Redémarre le service.

Service worker (suite)

`npm run worker:logs`	Affiche les logs.
`npm run worker:stop` / `worker:start`	Arrête ou démarre le service (il démarre seul en temps normal).
`npm run bug-report`	Génère un rapport de bug complet.

Configuration

Fichier `~/.claude-mem/settings.json` : modèle IA, port du worker, dossier de données, niveau de logs, contexte injecté.

- **Langue des observations** : `"CLAUDE_MEM_MODE": "code"` (anglais, par défaut). Les variantes suivent le motif `code--<langue>` (ex. `code--zh`, `code--ja`). Redémarrer Claude Code après modification.
- **Base de données** : `~/.claude-mem/claude-mem.db` (SQLite), consultable avec `sqlite3`.

Conseils d'utilisation

- **Surveiller la consommation de tokens** : plusieurs utilisateurs signalent un contexte déjà bien entamé au démarrage de session (issues #618 et #1848). Réduire le contexte injecté dans `settings.json` si c'est le cas.
- **Protéger les secrets** : tout ce qui passe dans les outils est capturé. Utiliser `<private>` pour les clés, données clients ou informations sensibles.
- **Demander des recherches ciblées** (type, période, projet) plutôt que « tout ce qu'on a fait » : c'est moins coûteux.
- **Choisir le fournisseur de mémoire** : l'observateur hébergé, sa propre clé OpenRouter ou Gemini, ou le forfait Anthropic. Après l'essai gratuit, la mémoire bascule sur le forfait Anthropic et consomme donc ton quota.

Sources (documentation officielle)

- README Claude-Mem
- Guide de démarrage
- Balises private
- Outils de recherche
- Issue #618 – consommation de tokens
- Issue #1848 – consommation de tokens