Eigenflow

GUIDE · HERMES AGENT

Hermes Agent et Gemini Notebook : ta veille en audio

Installer Hermes, le brancher à Telegram et à Gemini Notebook (ex-NotebookLM), la consigne de la veille du matin, et ce que l'outil non officiel fait de ton compte Google.

PAR

Abyane Nabeebaccus

INSTAGRAM

@abyane_nb

ÉDITION

novembre 2026

Tu as commenté HERMES, voici le guide. Hermes Agent est un agent IA open source qui tourne sur un serveur et te répond sur Telegram. Branché sur Gemini Notebook, le nouveau nom de NotebookLM, il peut chercher sur le web, ranger ce qu'il trouve dans un carnet et te préparer un résumé audio chaque matin.

Docs de Hermes Agent (version 0.23.2 du 16 novembre), dépôt et docs de notebooklm-mcp-cli (version 0.16.5 du 14 novembre), pages d'aide et pages développeurs de Google relus le 18 novembre 2026. On n'a aucun lien avec leurs auteurs. Rien n'a été installé ni lancé pour ce guide : tout vient des docs et du code lu sans l'exécuter.

À savoir avant de lire

- NotebookLM s'appelle Gemini Notebook depuis le 16 juillet 2026. Même produit, mêmes carnets, nouvelle adresse : notebook.google.com.
- Le branchement passe par un outil non officiel, notebooklm-mcp-cli. Il se connecte avec les cookies de ton compte Google, et son auteur écrit lui-même qu'il n'est peut-être pas conforme aux conditions de Google.
- Le connecteur officiel de Google sait créer un carnet et y ajouter des sources. Il ne sait pas encore répondre à une question ni créer un résumé audio.
- Hermes est gratuit. Le modèle qui le fait réfléchir se paie, et il faut une machine allumée jour et nuit.
- Gemini Notebook est gratuit avec des limites : 50 sources par carnet, un quota rechargé toutes les cinq heures. Google AI Pro (21,99 € par mois, prix affiché aux particuliers) monte à 300 sources.
- Prends un compte Google dédié : l'outil garde la connexion à ce compte sur la machine de l'agent.

---

1. Ce que tu construis

Trois pièces.

Hermes Agent, de Nous Research, sous licence MIT. Il garde une mémoire, se fabrique des skills, lance des tâches à heure fixe et parle sur Telegram, Discord, Slack, WhatsApp ou Signal. Son README propose de le faire tourner sur un serveur à 5 dollars par mois.

Gemini Notebook, l'assistant de recherche de Google. Tu y crées des carnets, tu ajoutes des sources (pages web, PDF, vidéos YouTube), tu poses des questions et il répond en citant tes sources. Il sait aussi chercher sur le web, en rapide ou en approfondi, et transformer un carnet en résumé audio.

Le pont : `nlm`, la commande de notebooklm-mcp-cli (Jacob Ben-David, licence MIT, 6 531 étoiles). Elle vient avec une skill, `nlm-skill`, qui apprend à Hermes à s'en servir. Hermes tape les commandes `nlm` dans son terminal, comme tu le ferais.



La FAQ de Gemini Notebook : « Depuis juillet 2026, NotebookLM s'appelle Gemini Notebook » (capture du 18 novembre)

---

2. Ce qu'il te faut

- Une machine allumée en permanence : un petit serveur Linux loué (un VPS) ou un Mac qui ne s'éteint pas. Sans elle, pas de veille le matin.
- Un modèle pour Hermes. D'après sa doc, un abonnement ChatGPT marche (ce qu'il consomme sur ton forfait n'est pas documenté), Claude Max seulement avec des crédits d'usage achetés en plus, Claude Pro et l'abonnement Gemini grand public ne marchent pas. Sinon, une clé API (OpenRouter, OpenAI, Anthropic) ou l'abonnement Nous Portal.
- Un bot Telegram créé chez @BotFather, et ton identifiant Telegram (un nombre, que @userinfobot te donne).
- Un compte Google dédié, avec le français comme langue du compte. Gemini Notebook s'en sert comme langue des résumés audio par défaut.
- uv, l'installeur Python d'Astral (ou pipx). L'outil demande Python 3.11 au moins, uv s'en occupe.
- Un navigateur sur la machine pour la connexion Google. Sur un serveur sans écran, on passe par un fichier de cookies (partie 4).

---

3. Installer Hermes et le brancher à Telegram

Sur ton serveur ou ton Mac, la commande officielle :

À COLLER

```
curl -fsSL https://hermes-agent.nousresearch.com/install.sh | bash
```

Recharge ton terminal (`source ~/.zshrc` sur Mac) et choisis ton modèle :

À COLLER

```
source ~/.bashrc
hermes model
```

Ensuite, Telegram :

À COLLER

```
hermes gateway setup
```

Choisis Telegram, colle le jeton donné par BotFather, puis ton identifiant. Seuls les identifiants listés pourront parler à ton agent. Le jeton donne le contrôle du bot : s'il fuite, révoque-le avec /revoke chez BotFather.

Pour que Hermes tourne même terminal fermé, installe-le comme service :

À COLLER

```
hermes gateway install
sudo loginctl enable-linger $USER
```

La deuxième ligne ne sert que sur un serveur Linux : le service reste en route après ta déconnexion et repart au redémarrage. Sur Mac, la première suffit.

Dernière étape, dans Telegram : écris à ton bot, puis tape /sethome. Les tâches planifiées livreront leurs résultats dans cette conversation.

---

4. Brancher Gemini Notebook en trois commandes

Fais-le après Hermes. L'installation de la skill cherche le dossier de Hermes et refuse de le créer s'il n'existe pas.

Si uv manque :

À COLLER

```
curl -LsSf https://astral.sh/uv/install.sh | sh
```

Puis les trois commandes :

```
À COLLER

uv tool install notebooklm-mcp-cli
nlm login
nlm skill install hermes
```

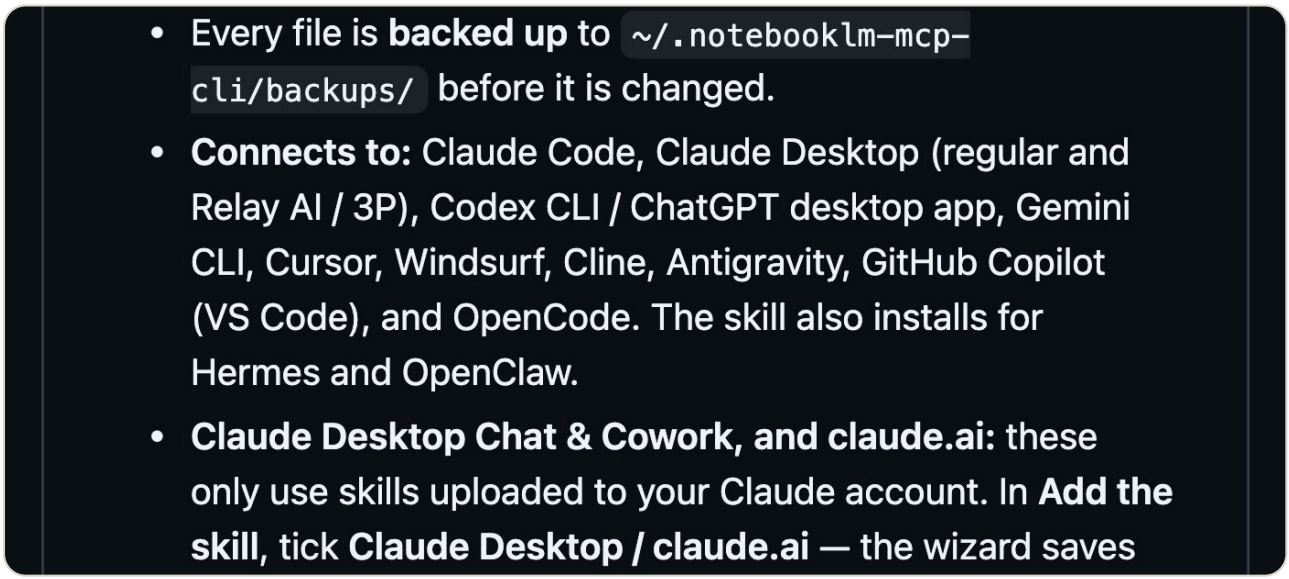
La première installe `nlm`, et un serveur MCP facultatif ici : la skill sait passer par la commande. La deuxième ouvre un navigateur : ferme d'abord Chrome complètement, puis connecte-toi avec le compte dédié. Avant d'ouvrir la fenêtre, `nlm` te demande où garder la connexion. Sur un Mac, prends « Protected » (chiffrée, clé dans le trousseau). La troisième copie la skill dans `~/.hermes/skills/nlm-skill`.

Pour vérifier :

```
À COLLER

nlm login --check
nlm notebook list
```

Sur Mac, le service Hermes garde les chemins connus le jour de son installation. Relance `hermes gateway install` après avoir installé `nlm`, sinon il risque de ne pas trouver la commande. La doc ne dit rien de pareil pour Linux : si Hermes ne trouve pas `nlm`, c'est la première piste.



Le README de notebooklm-mcp-cli : « The skill also installs for Hermes and OpenClaw » (capture du 18 novembre)

Sur un serveur sans écran

Pas de navigateur, donc pas de connexion automatique. L'outil prévoit l'import d'un fichier de cookies :

1. Sur ton ordinateur, ouvre Chrome sur notebook.google.com, connecté au compte dédié.

2. Ouvre les outils de développement (F12), onglet Network, filtre `batchexecute`, puis clique sur un carnet.

3. Dans une requête `batchexecute`, copie la valeur de la ligne `cookie:` des en-têtes et colle-la dans un fichier `cookies.txt`.

4. Copie ce fichier sur le serveur et importe-le.

À COLLER

```
nlm login --manual --file cookies.txt
```

Garde ce profil en fichier simple : le chiffrement par trousseau ne marche pas sur un serveur sans session graphique. Efface `cookies.txt` une fois importé. Ces cookies sont figés : le jour où l'outil dit qu'ils sont refusés, refais l'export.

---

5. Ta première demande

Écris à Hermes dans Telegram. Quelques demandes à copier :

À COLLER

```
Avec la skill nlm-skill, liste mes carnets Gemini Notebook.
```

À COLLER

```
Crée un carnet « Veille » et ajoute cette page comme source : [colle le lien]
```

À COLLER

```
Dans le carnet « Veille », réponds à cette question en citant tes sources :
quels outils reviennent le plus souvent ?
```

À COLLER

```
Lance une recherche web rapide sur « facturation électronique PME » dans un
nouveau carnet, importe les sources trouvées, puis résume-les en cinq points.
```

À COLLER

```
Lance une recherche approfondie sur « aides publiques à l'IA pour les PME » dans
le carnet « Veille » et importe les sources.
```

À COLLER

```
Génère un résumé audio au format brief, en français, du carnet « Veille »,
télécharge-le et envoie-le-moi ici. Tu as mon accord pour lancer la génération.
```

La skill demande ton accord avant toute génération ou suppression, d'où la dernière phrase. La recherche rapide prend environ 30 secondes et rapporte une dizaine de sources. L'approfondie prend environ 5 minutes, parfois jusqu'à 15, pour plus de 40 sources, sur le web seulement. Un résumé audio prend 1 à 5 minutes d'après l'outil, « quelques minutes » d'après Google. Le format brief, c'est une seule voix et moins de deux minutes. Le format par défaut fait discuter deux voix. Le réglage de la longueur n'existe qu'en anglais.

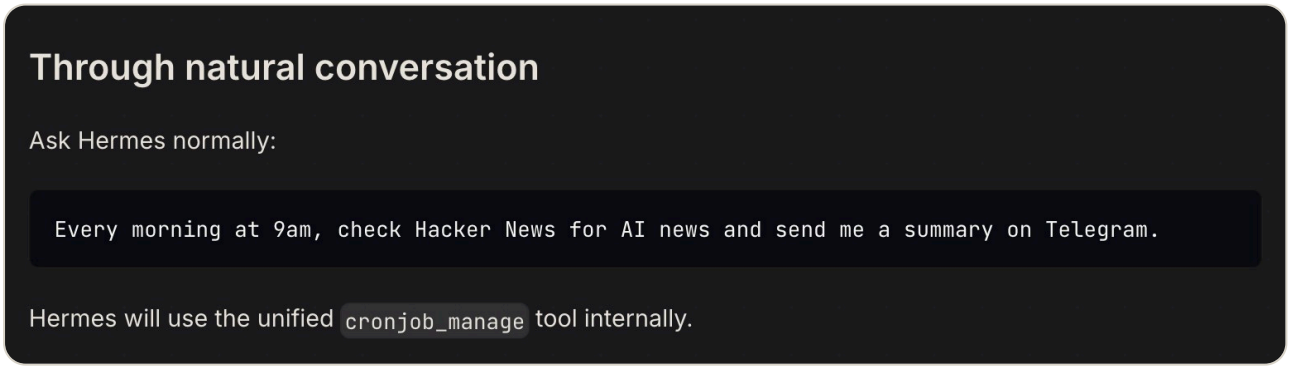
---

6. La veille de chaque matin

Une phrase suffit : Hermes crée la tâche lui-même. Sa doc donne l'exemple « Every morning at 9am, check Hacker News for AI news and send me a summary on Telegram. » La consigne à copier, ton secteur à la place des crochets :

À COLLER

```
Tous les jours à 6 h, avec la skill nlm-skill : lance une recherche web rapide
sur « [ton secteur] : nouveautés des dernières 24 heures » dans un nouveau
carnet nommé Veille suivi de la date du jour, importe les sources trouvées,
génère un résumé audio au format brief en français, télécharge-le en .m4a et
termine ta réponse par le chemin complet du fichier. Tu as mon accord pour créer
le carnet et lancer la génération sans me redemander.
```



La doc des tâches planifiées de Hermes : une tâche se crée en langage naturel (capture du 18 novembre)

Pourquoi ces choix :

- 6 h : la recherche et l'audio prennent quelques minutes. Mets l'heure un peu avant ton réveil.
- Un carnet par jour : en gratuit, un carnet tient 50 sources et une recherche rapide en ajoute une dizaine. Un carnet unique serait plein en cinq jours. Tu as droit à 100 carnets, fais le ménage chaque mois dans Gemini Notebook.
- L'accord écrit dans la consigne : à 6 h, personne n'est là pour dire oui.

- Le chemin du fichier en fin de réponse : la réponse finale part sur Telegram, et la passerelle envoie en pièce jointe un fichier audio dont elle voit le chemin complet. Un M4A arrive comme morceau audio. La doc prévoit aussi `[[audio_as_voice]]` pour en faire un message vocal, à essayer avec un M4A.

Vérifie le fuseau : Hermes lit l'heure dans celui de son profil (`timezone:` dans `~/.hermes/config.yaml`). Sur un serveur réglé en UTC, 6 h tombe à 7 h à Paris en hiver. Mets `Europe/Paris`.

Puis contrôle la tâche, et lance-la une fois pour voir :

À COLLER

```
hermes cron list
hermes cron run <id de la tâche>
```

La liste montre l'horaire et les skills attachées. Si `nlm-skill` n'y figure pas, crée la tâche en ligne de commande :

À COLLER

```
hermes cron create "0 7 * * *" "la consigne ci-dessus" --skill nlm-skill
```

Par défaut, une tâche planifiée refuse les commandes jugées dangereuses (réglage `cron_mode: deny`). Les commandes `nlm` ne figurent pas dans la liste de la doc : laisse ce réglage tel quel.

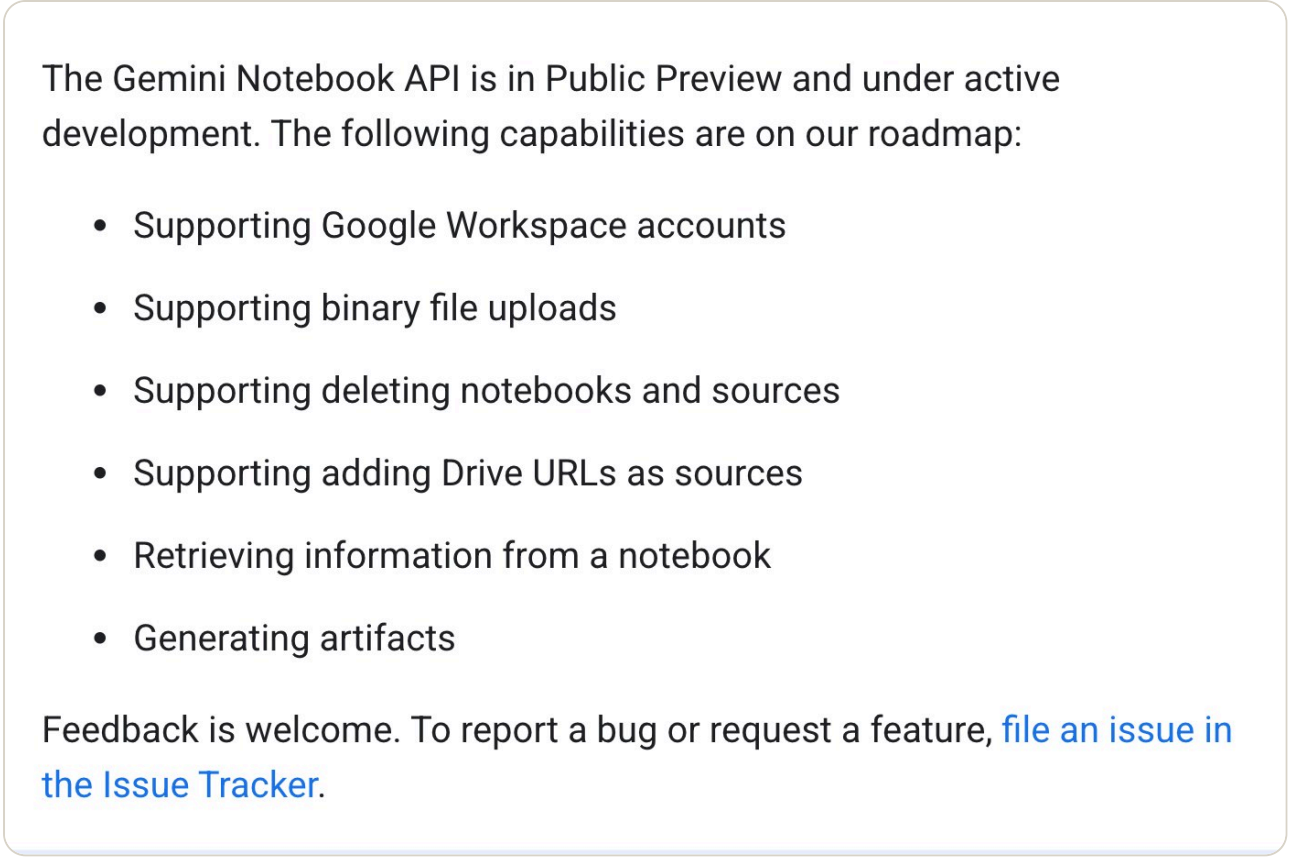
Ce qu'on ne sait pas, faute de l'avoir lancé : l'accent du résumé (l'outil a observé que le code de région choisit l'accent, `fr-FR` ou `fr-CA`), le temps réel d'un passage, ce qu'il coûte en modèle, la part du quota Gemini Notebook mangée chaque jour (`nlm usage` te le dit) et la durée de vie de la connexion Google sur un serveur.

---

7. Le connecteur officiel de Google

Google a ouvert une API Gemini Notebook et un serveur MCP (https://notebook.googleapis.com/mcp), en préversion publique, gratuits, pour les comptes Google personnels seulement. Sept outils : créer, lire, lister et modifier des carnets ; ajouter, lire et lister des sources (pages web, vidéos YouTube, texte, Markdown).

Sa feuille de route liste ce qui manque : l'envoi de fichiers (donc les PDF), la suppression, les liens Drive, l'interrogation d'un carnet et la génération d'artefacts, résumés audio compris. Google précise que cette liste est indicative et peut changer. Aucune date.



La feuille de route de l'API Gemini Notebook : « Retrieving information from a notebook » et « Generating artifacts » encore à venir (capture du 18 novembre)

Le brancher sur Hermes demande un projet Google Cloud, l'API activée et ton propre client OAuth. La doc MCP de Hermes décrit ce cas (`auth: oauth` avec `client_id` et `client_secret`). Les jetons sortis de gcloud expirent au bout d'une heure, inutilisables pour une tâche du matin. On ne l'a pas testé.

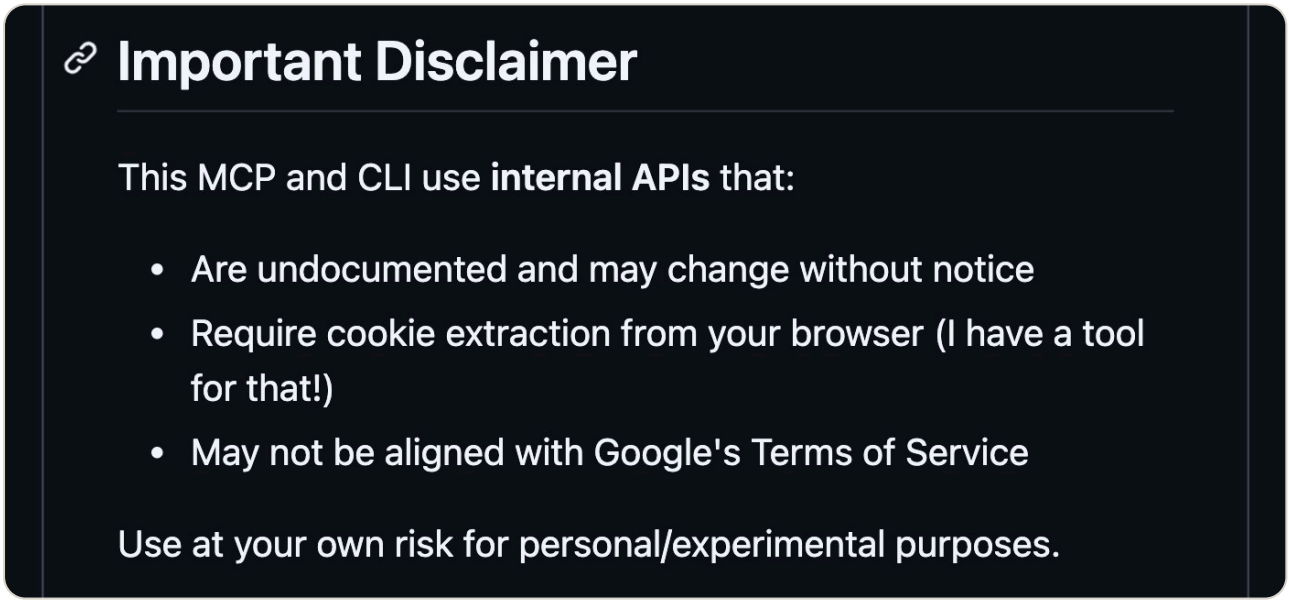
Il permet déjà une voie officielle à moitié automatique : Hermes trouve les liens avec sa propre recherche web et les ajoute à ton carnet, puis tu lances le résumé audio toi-même dans Gemini Notebook. Si c'est un compte pro sur lequel tu ne veux prendre aucun risque, c'est la voie à choisir pour l'instant. Le connecteur officiel remplacera l'outil non officiel le jour où l'interrogation des carnets et la génération d'artefacts sortiront de la feuille de route. D'ici là, ne branche pas deux connecteurs Gemini Notebook à la fois : l'auteur de notebooklm-mcp-cli en fait la première cause de « Hermes a pris le mauvais outil ».

---

8. Sécurité et limites

- Les cookies valent ton compte. La doc de sécurité de notebooklm-py, l'autre outil connu, l'écrit : « account-equivalent ». Chez notebooklm-mcp-cli, ils sont dans `~/.notebooklm-mcp-cli/profiles/<profil>/`, en fichier simple ou chiffrés.
- Le chiffrement ne te protège pas de Hermes. L'outil prévient que le mode protégé ne bloque pas un autre programme qui tourne sous ton compte utilisateur. Hermes tourne sous ton compte, avec un terminal. Et le profil de navigateur gardé pour les reconnexions n'est pas chiffré.
- D'où le compte dédié : rien d'autre dedans que tes carnets de veille. Ses limites sont celles de son forfait à lui. Pour lire un carnet depuis ton compte principal, partage-le (`nlm share invite`).

- Les conditions de Google interdisent la rétro-ingénierie de ses services et certains accès automatisés. Aucune page de Google ne nomme ces outils. On n'a trouvé aucun cas de compte suspendu, sans l'avoir cherché à fond. L'outil se dit lui-même réservé à un usage personnel ou expérimental.
- Ça peut casser. Le renommage du 16 juillet a cassé la connexion des comptes migrés jusqu'au correctif du 25 juillet (version 0.9.3). Mets l'outil à jour avec `uv tool upgrade notebooklm-mcp-cli`, puis `nlm skill update`.
- La connexion Google expire. Les cookies durent souvent plusieurs semaines et l'outil se reconnecte seul tant que la session vit. `nlm auth refresh`, lancé à intervalle régulier, aide sur une machine qui a un navigateur, mais il ne ranime pas une connexion morte. Sur un serveur à fichier de cookies, il faudra refaire l'export.
- Le quota de Gemini Notebook se recharge toutes les cinq heures, jusqu'à un plafond par semaine. `nlm usage` te dit ce qu'il reste.
- Tes données. Sources et questions partent chez Google, qui dit ne pas s'en servir pour entraîner Gemini Notebook sauf si tu lui envoies un avis. Elles passent aussi par le fournisseur du modèle de Hermes.
- Les pages piégées. Une page lue pendant une recherche peut cacher des ordres pour l'agent. Laisse `cron_mode: deny` et garde la liste des identifiants Telegram autorisés à toi seul.



L'avertissement du README de notebooklm-mcp-cli : API internes, cookies du navigateur, « May not be aligned with Google's Terms of Service » (capture du 18 novembre)

Sources

- Hermes Agent : dépôt, installation, passerelle, Telegram, tâches planifiées, skills, MCP, modèles et abonnements, sécurité.
- notebooklm-mcp-cli : dépôt, premiers pas, connexion, commandes.
- Google : renommage, Gemini Notebook, forfaits et limites, quotas, résumés audio, API et MCP, feuille de route, prix en France, conditions d'utilisation.
- Sécurité de notebooklm-py, installer uv.