AI TOOLS WIKI · MÉMO

Perplexity API Platform MCP Server *mémo*

L'essentiel pour utiliser **Perplexity MCP**, le serveur MCP officiel de Perplexity : recherche web, réponses sourcées, raisonnement et recherche approfondie depuis Claude Code ou Codex, facturés à l'usage sur l'API Perplexity.

Famille Recherche et veille · **Éditeur** Perplexity AI · **Licence** MIT
Note du wiki 11/15 · **Commandes vérifiées le** 15 novembre 2026
Fiche, notes et avis aitoolswiki.viellard.fr/outils/recherche-et-veille/perplexity-mcp

Installation

`claude mcp add --transport http perplexity https://api.perplexity.ai/mcp`	Serveur hébergé par Perplexity (recommandé) : connexion OAuth au premier /mcp
`claude mcp add --transport http perplexity https://api.perplexity.ai/mcp --header "Authorization: Bearer YOUR_API_KEY"`	Même serveur hébergé, avec une clé d'API au lieu d'OAuth
`claude mcp add perplexity --env 'PERPLEXITY_API_KEY=${PERPLEXITY_API_KEY}' -- npx -y @perplexity-ai/mcp-server`	Serveur local (stdio) : la clé est lue dans l'environnement qui lance Claude Code
`codex mcp add perplexity --env PERPLEXITY_API_KEY="your_key_here" -- npx -y @perplexity-ai/mcp-server`	Codex, serveur local
`claude plugin marketplace add perplexityai/modelcontextprotocol` `claude plugin install perplexity@perplexity-mcp-server`	Ou en plugin Claude Code

La clé d'API se crée dans le portail de l'API, qui demande un moyen de paiement. La connexion OAuth demande d'être administrateur d'une organisation API qui peut payer ; elle ne permet ni de créer des clés, ni de voir le solde.
Autres clients (Cursor, Claude Desktop, VS Code, Windsurf, Kiro) : même bloc `mcpServers` avec `"command": "npx"`, `"args": ["-y", "@perplexity-ai/mcp-server"]` et la clé dans `"env"`.

Outils MCP

L'agent les appelle lui-même ; les nommer dans la demande aide à choisir le bon.

Outil	Ce qu'il fait	À demander par exemple
`perplexity_search`	Résultats classés (titre, URL, extrait, dates) via la Search API.	« Avec Perplexity, trouve les sources officielles sur … »

Outils MCP (suite)

Outil	Ce qu'il fait	À demander par exemple
`perplexity_ask`	Réponse rédigée et sourcée (preset `fast` de l'Agent API).	« Demande à Perplexity quelle est la dernière version de … »
`perplexity_reason`	Raisonnement en plusieurs étapes (preset `medium`).	« Compare les deux approches avec Perplexity. »
`perplexity_research`	Recherche approfondie avec citations (preset `high`), plusieurs minutes.	« Lance une recherche approfondie Perplexity sur … »

Paramètres utiles

Paramètre	Outil	Effet
`search_recency_filter`	`perplexity_search`	Limite aux sources récentes.
`search_domain_filter`	`perplexity_search`	Restreint à certains domaines.
`search_type: "fast"`	`perplexity_search`	Recherche plus rapide et moins chère (serveur local 1.3.0 ou plus).
`max_tokens`	`perplexity_ask`, `perplexity_reason`	Budget total de contexte de recherche, jusqu'à 120 000.
`max_tokens_per_page`	`perplexity_ask`, `perplexity_reason`	Budget par page de résultats, jusqu'à 20 000.

Ces budgets règlent le contenu récupéré, pas la longueur de la réponse. L'ancien paramètre `search_context_size` (`low`, `medium`, `high`) est encore accepté et converti en budgets.

Variables d'environnement (serveur local)

Variable	Rôle	Défaut
`PERPLEXITY_API_KEY`	Clé d'API.	Obligatoire
`PERPLEXITY_TIMEOUT_MS`	Délai maximal d'une requête.	300 000 (5 minutes)
`PERPLEXITY_BASE_URL`	Adresse de l'API.	`https://api.perplexity.ai`
`PERPLEXITY_LOG_LEVEL`	`DEBUG`, `INFO`, `WARN` ou `ERROR`.	`ERROR`

Variables d'environnement (serveur local) (suite)

Variable	Rôle	Défaut
`PERPLEXITY_PROXY`	Proxy d'entreprise, puis `HTTPS_PROXY` et `HTTP_PROXY`.	Connexion directe

Auto-hébergement en HTTP

```
docker build -t perplexity-mcp-server .
docker run -p 8080:8080 -e PERPLEXITY_API_KEY=your_key_here perplexity-mcp-server
```

Ou, depuis le dépôt cloné : `npm install && npm run build && npm run start:http`. Le serveur répond sur `http://localhost:8080/mcp`. Il n'écoute que l'interface locale (`BIND_ADDRESS=127.0.0.1`) ; pour l'ouvrir, régler aussi `ALLOWED_HOSTS` et `ALLOWED_ORIGINS`.

Coûts (au 15 novembre 2026)

Appel	Tarif
`perplexity_search`	5 $ pour 1 000 requêtes ; 1 $ en mode `fast`.
`perplexity_ask`, `perplexity_reason`, `perplexity_research`	Agent API : jetons au tarif du modèle, plus des frais d'outils (recherche web : 2,50 $ pour 1 000 ; 1 $ en mode rapide).

Conseils d'utilisation

- **Choisir l'outil le moins cher qui suffit** : `perplexity_search` pour des sources, `perplexity_ask` pour un fait, `perplexity_reason` pour une analyse, `perplexity_research` seulement pour un vrai dossier.
- **Relever les délais pour la recherche approfondie** : `PERPLEXITY_TIMEOUT_MS` côté serveur, et le délai du client MCP, souvent de 60 secondes.
- **Erreurs EOF ou d'initialisation** avec un client strict : remplacer `npx -y` par `npx -yq`.
- **Ne jamais écrire la clé dans un fichier versionné** : utiliser une variable d'environnement, comme dans `'PERPLEXITY_API_KEY=${PERPLEXITY_API_KEY}'`.
- **Vérifier les sources primaires** : `perplexity_ask`, `perplexity_reason` et `perplexity_research` renvoient une synthèse.
- **Traiter le contenu web comme non fiable** : une page peut contenir des consignes destinées à l'agent.

Sources (documentation officielle)

- README du serveur officiel
- Documentation : serveur MCP
- Tarifs de l'API