

La structure de coûts de ton système IA

Tous les postes de coût de ton système (tokens IA, dictée, outils, serveur, Obsidian Sync) et comment les maîtriser : quels abonnements prendre, comment router tes tâches, et 3 scénarios chiffrés selon ton usage.

Source : <https://prisme.one/plateforme/ia-juillet-2026/semaine-1-structure-couts>

La structure de coûts de ton système IA

C'est la question qui revient le plus souvent, et c'est le piège n°1 de la promo précédente : des coûts IA non maîtrisés, des factures qui explosent d'un mois sur l'autre. Cette leçon existe pour que ça ne t'arrive pas.

La bonne nouvelle : ton système ne coûte pas un prix fixe. Il coûte ce que **ton** usage décide. Et une fois que tu connais chaque poste, tu pilotes ta facture au lieu de la subir.

Ce qui fait varier ta facture

Ton coût mensuel dépend de 4 facteurs, tous liés à toi :

Facteur 1

Ton intensité

Combien tu consommes. Quelques requêtes par jour ou une IA qui tourne en continu, ce n'est pas la même facture.

Facteur 2

La complexité de tes tâches

Une tâche lourde (long contexte, raisonnement) mange bien plus de tokens qu'une reformulation rapide.

Facteur 3

L'intelligence dont tu as besoin

Un modèle puissant coûte plus cher qu'un modèle léger. Tout l'enjeu est de ne le sortir que quand la tâche l'exige.

Facteur 4

Ton niveau de vie privée requis

Plus tu as besoin de confidentialité (tokens privés, ZDR), plus tu t'orientes vers des abonnements dédiés.

Le principe qui gouverne tout

Ne prends un outil QUE s'il t'est utile

Chaque ligne d'abonnement ci-dessous est optionnelle par défaut. Tu ne l'ajoutes que si un usage réel le justifie. Répète-toi cette phrase à chaque fois que tu es tenté de cliquer sur "s'abonner". C'est ta meilleure protection contre la facture qui gonfle.

1. Les tokens IA (le poste principal)

C'est le gros morceau. Tout ce que ton IA lit et écrit se paie en tokens. **L'objectif : trouver des abonnements qui t'évitent de payer l'IA à la consommation**, parce que la consommation, c'est ce qui explose sans prévenir.

La reco de base : deux abonnements complémentaires

Tâches performantes, non privées

Abonnement ChatGPT

20 €/mois (Plus), ou à partir de 100 €/mois (Pro) selon ton usage. Pour tout ce qui demande de la performance et qui n'est pas confidentiel.

Tout ce qui est confidentiel

Abonnement Synthetic AI

30 \$/mois, modèle GLM 5.2, tokens **PRIVÉS (ZDR)**. Pour tout ce qui touche à des données sensibles. synthetic.new

Avec ces deux abonnements bien utilisés, ton poste tokens tient dans une fourchette prévisible : **entre 20 €/mois (ChatGPT seul) et 150 à 200 \$/mois** selon ta consommation. Pas de mauvaise surprise.

Pourquoi des abonnements plutôt que la conso

En ce moment, le marché est dans une phase où les abonnements te donnent **beaucoup plus de tokens** que si tu les payais à la consommation. Autant en profiter.

- **Synthetic** est excellent pour ça : tu paies un forfait fixe et tu obtiens des tokens privés (ZDR) en quantité, pour à peu près le prix d'un ChatGPT.
- **ChatGPT** est top parce qu'il est utilisable dans **Codex** (la meilleure boucle agentique du moment) ET dans **OpenCode** (si tu veux garder une seule interface pour tout).

Bien router : le nerf de la guerre

Router, c'est envoyer chaque tâche vers le bon modèle : **un modèle puissant seulement si la tâche l'exige, sinon un modèle moins cher**. C'est le geste qui fait la différence entre une facture de 20 \$ et une facture de 200 \$ pour le même travail.

Si tu paies au token via **OpenRouter**, applique deux règles :

- Prends les **modèles les moins chers** pour les tâches les moins complexes.
- Réserve les **modèles chers** aux abonnements (pas à la conso au token).
- Diminue la place d'OpenRouter au profit de **Synthetic** dès que tu peux : tu passes d'un coût variable à un forfait maîtrisé.

Le défaut open source pas cher

Quand tu paies au token via OpenRouter, **DeepSeek V4 Flash** reste ton modèle par défaut : très bon marché, ZDR, et il couvre l'immense majorité des tâches courantes.

Note sur Claude

Claude ne se connecte pas via abonnement dans OpenCode : uniquement via le CLI Claude Code. Pas grave, ça ne change rien à ta structure de coûts, c'est juste bon à savoir.

Pour le détail du choix des modèles, va voir [Choisir son modèle IA](#).

2. La dictée vocale (voice-to-text)

Parler à ton IA au lieu de tout taper change ton rythme de travail. Et ça peut coûter **0 €**.

- **Gratuit** : Murmure (open source, ZDR) ou d'autres logiciels gratuits font très bien le job.
- **Reco** : SuperWhisper. Pro à 8,49 \$/mois, ou 84,99 \$/an, ou Lifetime 249,99 \$ une seule fois.
- **Alternative** : Wispr Flow (plan gratuit + environ 12 à 15 \$/mois).

C'est **optionnel mais fortement recommandé** : avoir un bon outil pour parler à l'IA fait vraiment gagner du temps au quotidien. Détails côté setup dans [Prérequis et bonnes](#)

pratiques.

3. Les autres abonnements outils (selon tes process)

Ceux-là, tu ne les prends **que si un process précis en a besoin**. Sinon tu les laisses de côté.

- **Recherche web** : Firecrawl (gratuit jusqu'à 1000 crédits/mois, puis environ 16 à 19 \$/mois) ou Apify (gratuit avec environ 5 \$ de crédit/mois, Starter à 29 \$/mois). Tu as des crédits gratuits pour démarrer, donc pas besoin de payer au début.
 - **Le reste** : uniquement si un workflow précis le réclame. Pas d'abonnement "au cas où".
-

4. Le serveur (semaine 3)

Le serveur arrive en semaine 3. Tu ne le prends que si tu veux **héberger ton site, auto-héberger tes outils et économiser** dessus.

- **VPS** : environ 5 à 10 €/mois (par exemple Hetzner CX22, autour de 5,5 € HT).
- **Deux options de gestion** : tu le gères toi-même, OU tu prends **Cloudron** qui le manage pour toi (+15 €/mois).
- **Plafond** : jusqu'à environ 25 €/mois max, tout compris.

Détails dans [Ton serveur IA](#).

5. Obsidian Sync

Vraiment recommandé. Tu pourrais bricoler avec Google Drive, mais c'est galère et source de conflits. Sync règle le problème proprement.

- **Standard** : 4 \$/mois.
 - **Plus** : 8 \$/mois, selon tes besoins.
-

Trois scénarios chiffrés

Voici trois profils types, poste par poste, pour te donner un ordre de grandeur concret. Les montants mélangent USD et EUR (les outils sont facturés dans leur devise d'origine), et **les prix évoluent** : vérifie toujours au moment où tu t'abonnes.

Scénario A : petit utilisateur (perso, léger)

Usage personnel, quelques tâches par jour, rien de confidentiel de niveau entreprise.

Poste	Prix	Optionnel ?
ChatGPT Plus	20 €/mois	Non (socle)
Obsidian Sync (Standard)	4 \$/mois	Non (recommandé)
SuperWhisper (Pro)	8,49 \$/mois	Oui
Total	~24 à 32 \$/mois	

Scénario B : utilisateur boîte / intensif

Usage pro quotidien, données clients à protéger, dictée vocale au quotidien, un serveur pour auto-héberger.

Poste	Prix	Optionnel ?
ChatGPT Plus	20 €/mois	Non (socle)
Synthetic AI (GLM 5.2, ZDR)	30 \$/mois	Non (données privées)
Obsidian Sync (Plus)	8 \$/mois	Non (recommandé)
SuperWhisper (Pro)	8,49 \$/mois	Oui
Serveur (VPS)	~10 €/mois	Oui
Total	~75 à 90 \$/mois	

Scénario C : gros utilisateur / power

Usage intensif, IA en mode agent une bonne partie de la journée, recherche web automatisée, serveur managé.

Poste	Prix	Optionnel ?
ChatGPT Pro	100 €/mois	Non (socle)
Synthetic AI (GLM 5.2, ZDR)	30 \$/mois	Non (données privées)

Poste	Prix	Optionnel ?
Obsidian Sync (Plus)	8 \$/mois	Non (recommandé)
SuperWhisper	8,49 \$/mois	Oui
Firecrawl	19 \$/mois	Oui
Serveur + Cloudron	~25 €/mois	Oui
Total	~180 à 200 \$/mois	

À retenir

Le passage d'un scénario à l'autre ne se fait pas par plaisir : il se fait quand ton usage réel le justifie. Beaucoup de gens vivent très bien en scénario A pendant des mois. Ne monte de palier que quand tu sens la limite, pas avant.

Pour finir

Retiens une seule chose : **ne prends que ce qui t'est utile**. Chaque poste optionnel reste optionnel tant qu'un usage réel ne le rend pas indispensable. C'est comme ça que tu gardes une facture nette et prévisible au lieu de la voir gonfler.

Pour aller plus loin sur chaque brique :

- Choisir son modèle IA : bien router chaque tâche vers le bon modèle.
 - Comment utiliser OpenCode Desktop : ton interface unique.
 - Installer tes outils IA : le setup complet.
 - Ton serveur IA : VPS, Cloudron, auto-hébergement.
 - Prérequis et bonnes pratiques : dont l'installation de SuperWhisper.
-

- ☐ J'ai choisi mes abonnements selon mon usage réel
- ☐ Je sais router chaque tâche vers le bon modèle (puissant seulement si nécessaire)
- ☐ J'ai séparé mes tâches performantes (ChatGPT) et confidentielles (Synthetic, ZDR)
- ☐ Je ne paie que les outils qui me sont utiles
- ☐ J'ai identifié dans quel scénario (A, B ou C) je me situe
- ☐ J'ai un ordre de grandeur clair de mon coût mensuel

Extraction verbatim depuis Prisme One · <https://prisme.one/plateforme/ia-juillet-2026/semaine-1-structure-couts>