

# Comment l'agent LLM de la Partie 2 a réellement été construit

Compagnon technique de la [Partie 2 de la trilogie SHORA](#). Pour les lecteurs curieux du pipeline [pji-agent](#) qui a produit le taux de navigation de 9 % sur 10,5 millions de pages produit.

Ce post compagnon documente comment l'agent LLM de la Partie 2 a été construit, à destination des ingénieurs, des chercheurs et de la communauté open-source. Chaque référence de code est un lien profond vers le dépôt public sous licence MIT :

[github.com/crospector/pji-agent](https://github.com/crospector/pji-agent).

## Ce qu'il couvre

- **L'agent d'extraction** — explore le catalogue de 1 056 enseignes françaises et extrait des relevés produit structurés (nom, prix, URL, marque, hiérarchie de catégories, avis).
- **L'agent de parcours** — pour chaque produit, exécute un parcours d'achat complet : naviguer vers la PDP, capturer le prix PDP, cliquer sur « ajouter au panier », naviguer vers le panier, capturer le prix du panier, classifier le résultat.
- **Le modèle dans le chemin de données** — `eu.anthropic.claude-opus-4-6-v1` sur Amazon Bedrock, en région `eu-west-3`, sur des instances EC2 `c6i.8xlarge`, pilotant 20 à 40 navigateurs Camoufox concurrents via Playwright.
- **La classification des résultats** — `JOURNEY_COMPLETE`, `JOURNEY_BLOCKED`, `PRICE_MISMATCH` — et la disposition des preuves dans S3 (capture d'écran, HTML signé, arbre d'accessibilité, HAR réseau, journal de console, vidéo, trace Playwright).
- **Le contexte des benchmarks tiers** — WebArena, Online-Mind2Web, WAREX, OSWorld — qui situent les 9 % dans la littérature publiée sur les agents web.

La version intégrale — avec chaque chemin de fichier, chaque prompt, la disposition exacte des preuves et le schéma DynamoDB — est en anglais, parce qu'elle est truffée d'identifiants de code que la traduction ne devrait pas altérer.

[Lire le compagnon technique complet \(anglais\) →](#)