

# J'ai mis des agents IA en production. Voici ce qu'ils ne font pas.

Un architecte expérimenté peut aujourd'hui piloter seul un portefeuille logiciel important avec des agents IA, à condition de remplacer le réflexe « je demande, il code » par un système d'ingénierie complet : mémoire, règles, tests, relectures et traçabilité.

|              |              |                               |                          |                    |                    |                 |
|--------------|--------------|-------------------------------|--------------------------|--------------------|--------------------|-----------------|
| <b>2 048</b> | <b>121</b>   | <b>226</b>                    | <b>60+</b>               | <b>~40</b>         | <b>20+</b>         | <b>40 %</b>     |
| commits      | jours actifs | demandes triées en un passage | décisions d'architecture | bibliothèques .NET | sources de données | de code de test |

Mesures sur quatre des cinq dépôts, sur 121 jours d'activité.

## QUI DÉCIDE, QUI EXÉCUTE ?

### À moi

- La décision métier et l'arbitrage
- La responsabilité et l'architecture
- La validation finale, surtout avant toute écriture en production

### À l'agent

- Explorer le code et les données, implémenter, tester
- Relire, documenter, déployer
- La surveillance répétitive, chaque nuit

L'agent peut proposer, implémenter, relire et documenter. La responsabilité reste humaine et ne se délègue pas.

## DE L'ISSUE À LA PRODUCTION

Décision humaine, issue GitHub, agent, tests, relecture, pull request, validation humaine, production, observation, nouvelle décision. Chaque demande devient d'abord une issue, notée par un score qui mesure le risque financier et le nombre d'enregistrements touchés :

$$\text{Score} = (\text{Valeur} \times \text{Impact} \times \text{Risque} \times \text{Environnement}) / \text{Effort}$$

## LE SYSTÈME

Cinq dépôts et une mémoire commune : un projet de données pour une PME, un moteur de recherche par références croisées, un framework d'environ 40 bibliothèques .NET, les vitrines web et un dépôt de pilotage (décisions, ADR, journal, relevés d'audit).

- **Garde-fous** : cahier des charges avant le code, auto-relecture avec preuve par mutation, suite de tests complète, relecture indépendante avant fusion.
- **Aucune erreur silencieuse, rien de deviné.** Mon processus vise à n'écrire en production qu'après validation sur Test.
- **Automates** : une routine du matin qui lit les journaux d'exécution des environnements de test et de production, rédige le rapport de nuit et propose la prochaine demande (son unique écriture : une pull request), un agent en lecture seule, des workflows et des hooks. Un automate observe et rapporte avant d'obtenir le droit d'agir.

## LES LIMITES

- Environ un commit sur cinq est une correction : l'agent va vite et se trompe parfois.
- Lors d'un triage, il a proposé 21 fermetures : 12 ont tenu à la vérification, 9 ont été réfutées.
- Le système a un coût d'entretien et suppose de l'expérience.

**L'IA n'a pas réduit la responsabilité. Elle a augmenté la capacité de pilotage.**

Un problème de PME est rarement seulement technique ou seulement humain, il est les deux, vus sous deux angles. Plus de 30 ans de terrain en architecture de données et en IT au service de PME.