



# **Think Tank – Do Tank**

## **sur l'IA Générative appliquée au domaine de l'Ingénierie Logicielle**

**3ème édition – Mercredi 28 janvier 2026**

**Bilan & vision sur les tendances technologiques 2025-26+.**

# Réunion Générale - 3ème édition

**Table Ronde : Bilan & vision sur les tendances technologiques 2025-26+ .**



**Pascal PACCAUD** : Consultant / Expert IA Gen

**amadeus**



**Sylvain CAILLIAU** : Directeur Technique Opérations France

**CAST**



**25 minutes**

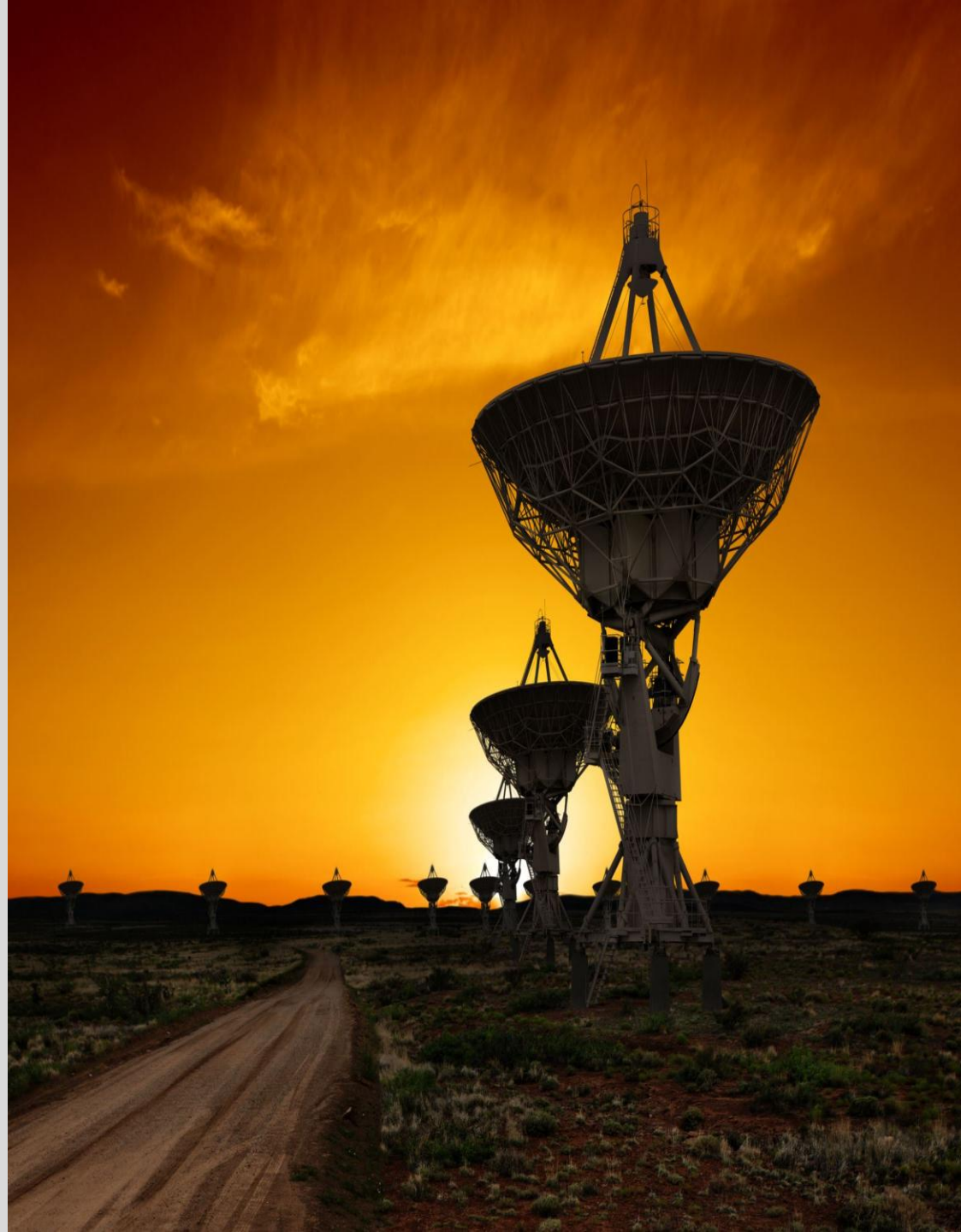
# Bilan et Perspectives

Janvier 2026

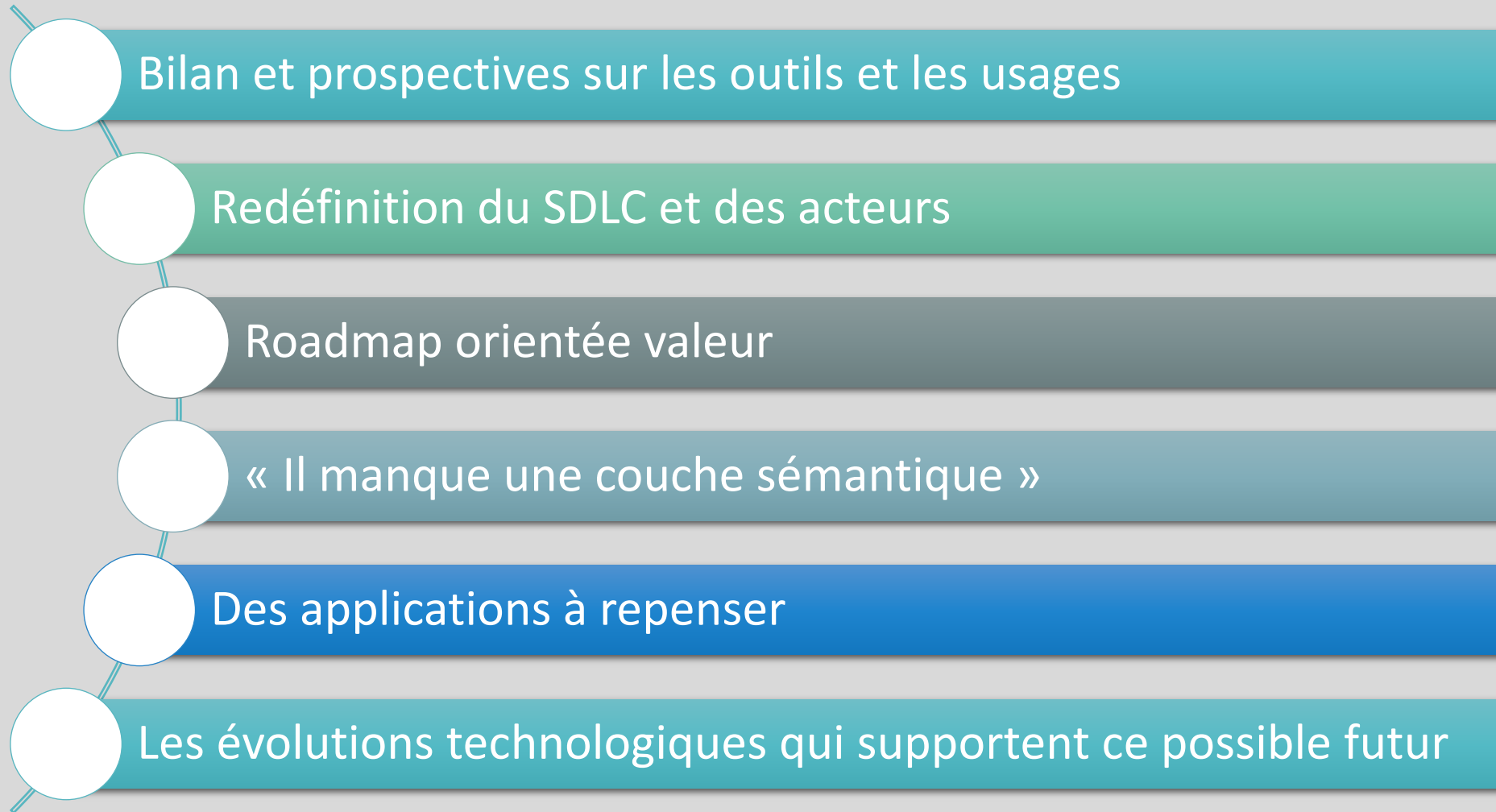
FR GenAI DO Tank

Pascal PACCAUD (Amadeus / Sopra Steria Group)

Sylvain CAILLIAU (Cast Software)



# Aspects principaux



# Des outils qui deviennent des commodités

## Fin 2024

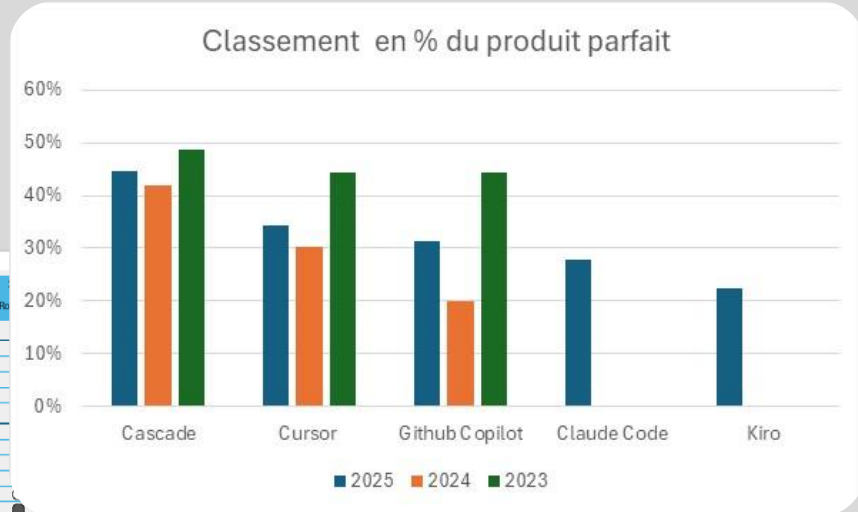
	2018	2021	2021	2022	2023	2023	2023	2023	2023	2023	2023
	Tabnine	Github Copilot	Open AI Codex	Replit	Aider	Codemium / Cascade	Continue	Cursor	Duet AI or code assists	JetBrainAssitant	Sourcegraph Cody
2023											
In-editor assistance (explain, fix...)	●	●	●	●	●	●	●	●	●	●	●
chat assistance and copy paste	○	○	○	○	○	○	○	○	○	○	○
Automated context management	○	○	○	○	○	○	○	○	○	○	○
Edge model support	○	○	○	○	○	○	○	○	○	○	○

## Fin 2025

	2018	2021	2021	2022	2023	2023	2023	2023	2023	2023	2024	2024	2024	2024	
	Tabnine	Github Copilot	Open AI Codex	Replit	Aider	Codemium / Cascade	Continue	Cursor	Duet AI or code assists	JetBrainAssitant	Sourcegraph Cody	Claude Code	Cline	Lovable	Q dev
2023															
In-editor assistance (explain, fix...)	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
chat assistance and copy paste	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Automated context management	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Edge model support	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
2024															
Terminal Usage	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Editing of 1 or more files	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Multi models	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Codebase indexation on PC	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Multiple IDE support	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Agent in IDE	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Slash (/) commands	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Human context definition	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
MCP server support	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System prompt modification	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
API/SDK	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Role based mode	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

## Fin 2026

	2018	2021	2021	2022	2023	2023	2023	2023	2023	2023	2023	2024	2024	2024	2024	2024	2024	2024	2025	2025	2025
	Tabnine	Github Copilot	Open AI Codex	Replit	Aider	Codemium / Cascade	Continue	Cursor	Duet AI or code assists	JetBrainAssitant	Sourcegraph Cody	WARP	Bolt	Claude Code	Cline	Lovable	Q dev	Roocode	Agent42	Kiro	Open Code
2023																					
In-editor assistance (explain, fix...)	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
chat assistance and copy paste	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Automated context management	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Edge model support	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
2024																					
Terminal Usage	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Editing of 1 or more files	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Multi models	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Codebase indexation on PC	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Multiple IDE support	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Agent in IDE	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Slash (/) commands	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Human context definition	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
MCP server support	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System prompt modification	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
API/SDK	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Role based mode	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Bring your own lang support	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
2025																					
Full fledged and context integrated terminal	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Multiple parallel agents	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
CI	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
background Agent	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Remote Agent	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Hook	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Agents_MPD instruction	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Skills	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Plugins (runtime driven)	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
MCP v2 support	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Spec Driven Dev	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○



Socle Commun

Differentiant

Multi Agents

Skills

Agents.md

MCP

Spec Driven Development

IDE/CLI/Code Cloud Agent

Gestion de contexte et Memoire

# Fonctionnalités anticipées en 2026

2026	
★	Support for user coded agent (e.g., AWS Strands)
★	Support for local potent model
★	LSP or advanced AST support and usage exposed
★	Git Workfree (under the hood)
★	Agents' dashboard
★	Extended unattended agents (Local or Cloud)
★	Context intent driven injection
★	Codebase Semantic layering and retrieval
☆	Automated pre commit code review and refactoring
☆	Exploratory evaluation
★	Competency delivery (e.g., Claude plugins, gh app modernization)
★	ACP (Agent Client Protocol)
★	AI-SDLC Asssitance in tools (SDD and Project Management)
★	Curated Marketplace for Skills, Powers, Agents

- Strands Agents - AWS Prescriptive Guidance
- LSP Servers | OpenCode
- Using Git Worktrees for Multi-Feature Development with AI Agents · Nick Mitchinson
- Introducing Agent HQ: Any agent, any way you work - The GitHub Blog
- Custom - AWS Transform
- Powers – Kiro
- MCP server/AI's new map

Find-tool Tool

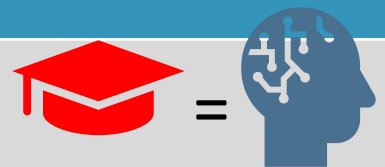
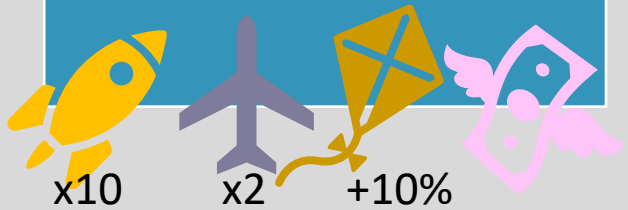
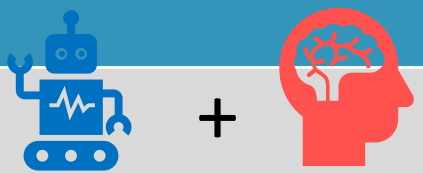
...

Il n'y a pas de facteur 'X' entre les solutions

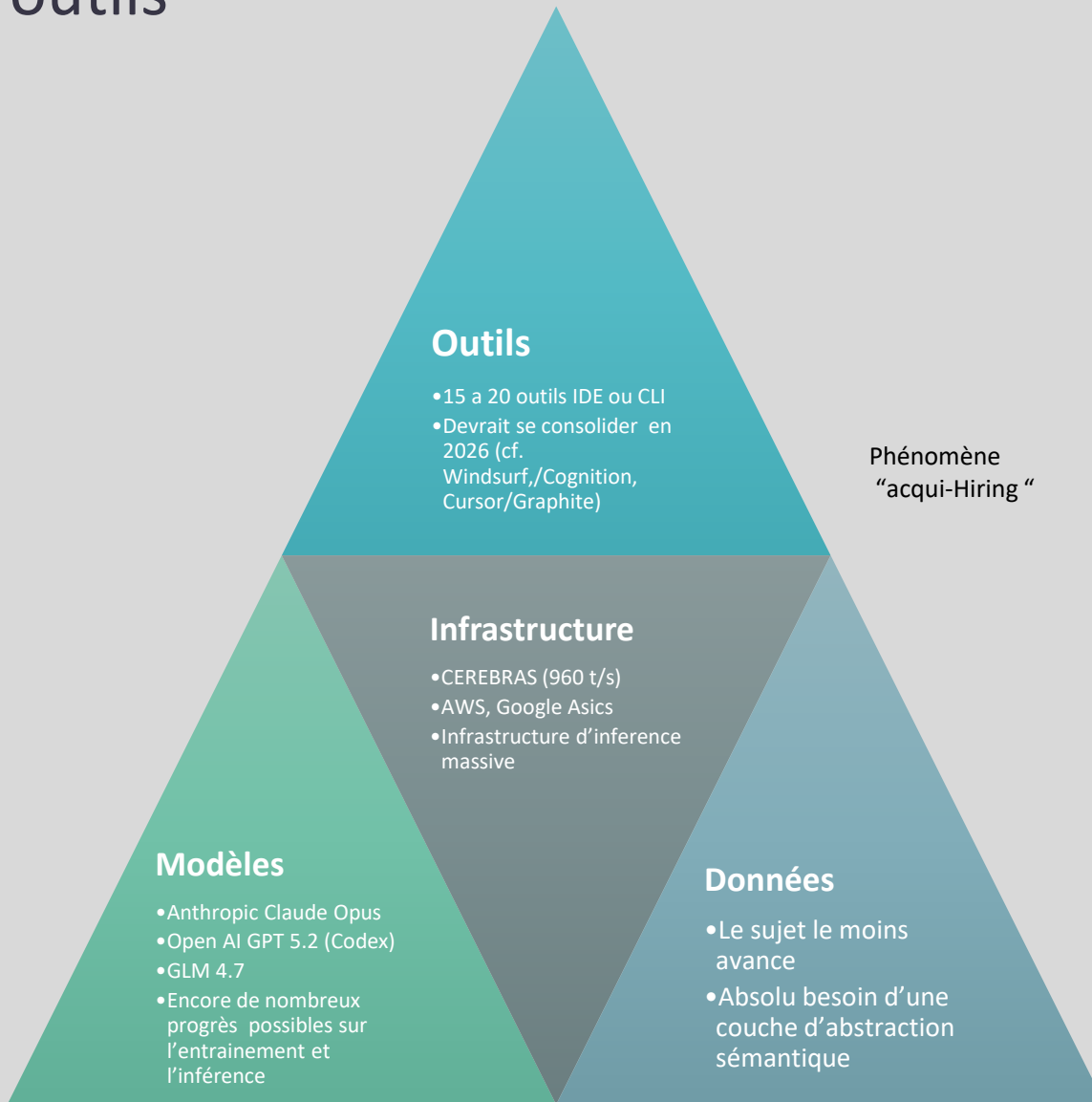
4 Catégories d'utilisation d'outils

Les plus productifs utilisent plusieurs outils

En 2027, les premières promos d'ingénieurs formés à l'IA sortiront



# 2026, année de la consolidation, l'important sera davantage les pratiques que les outils



Il faut changer de mode de travail .. mais tout le monde n'y arrivera pas ( cf. [Boris Cherny](#))

Le cout de telles sessions de travail reste énorme (cf. [Peter Steinberger 51 kUSD](#))  
L'anglais devient le nouveau langage de programmation (cf. Andrej Karpathy – jan 2023)

**LE CODE DEVIENT JETABLE, LE SAVOIR DOIT RESTER**

**Nous allons probablement devoir repenser le travail de tous les jours.**

**AI-SLDC ou ses variantes seront-elles adaptées à toutes les entreprises ?**

# La Mc Kinsey Way de parler du sujet

How can you go from "10x engineers" to a "10x team"? → How can you scale a "10x team" to a "10x company"?

McKinsey & Company 3

Vs.

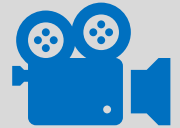
Impact of AI in individual developer tasks is large

<b>Speed of code reviews</b> <b>7x ↑</b> <i>GitHub Copilot</i>	<b>Speed of code refactors and prototyping</b> <b>2-5x ↑</b> <i>Cursor at Monday.com</i>
<b>Speed of ETL platform migration</b> <b>12x ↑</b> <i>Devin at Nubank</i>	<b>Speed of unit test development</b> <b>7x ↑</b> <i>Cursor at Salesforce</i>

McKinsey & Company 4

**MOVING AWAY FROM AGILE: WHAT'S NEXT?**

**MOVING AWAY FROM AGILE: WHAT'S NEXT?**



Les entreprises\* sont coincées dans un gain de +10% à +30% ...

Most organizations are at productivity of.. **1.1-1.3x<sup>1</sup>**

The current agile dev working model was developed with a set of constraints that are changing

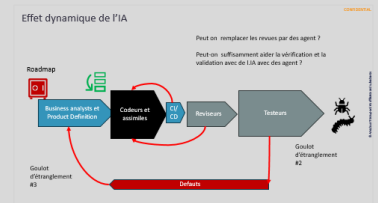
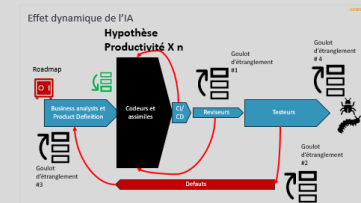
... but top performers can achieve productivity of... **2-5x**

By re-wiring the PDLC, operating model, and roles

1. 16-30% productivity improvements (2025 McKinsey survey), 39% increase in merges/PR (2025 UChicago study about Cursor impact on enterprise productivity)

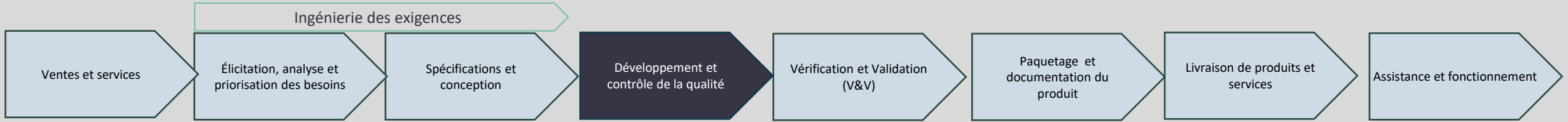
McKinsey & Company 7

... mais le potentiel est de x2 a x5

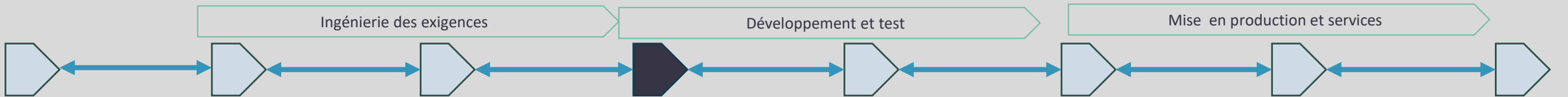


# Un SDLC qui doit et va évoluer

Vision "ingénieur" » **augmentée !!!**



Vision "Métier" **augmentée ???**



Le cycle de vie actuel du développement logiciel tente de pallier la... de form... d'exig... testable... développeurs et testeurs.

**Nous nous sommes concentrés sur les apports de la GenIA afin d'améliorer les 'actions'**  
**Même si nous gagnons 50% sur chaque poste d'actions, le Métier ne percevra pas complètement l'apport**

Stratégical Direct Classification;

opérationnels inutiles.

# Un SDLC qui doit et va évoluer vers du shift-up

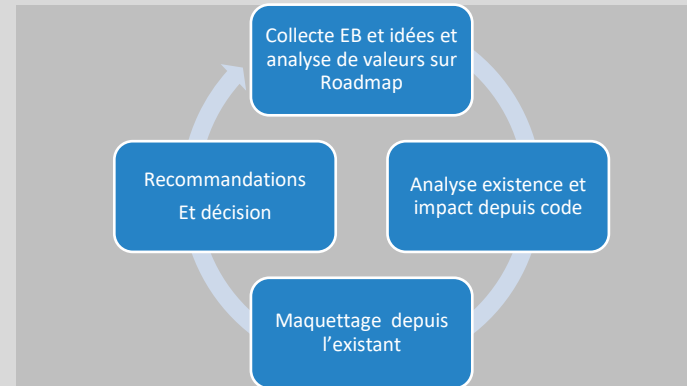
Explorations automatisées plutôt que planification

Maquettage direct avec puis par les métiers plutôt que des briefs et spécifications

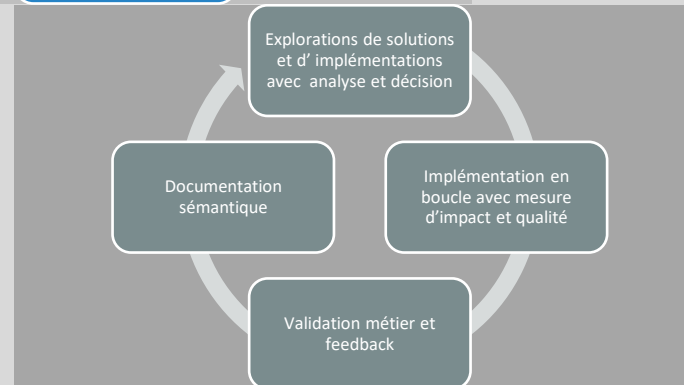
Usine à produire plutôt que gestion de projet et points de décisions réguliers

Dashboard d'avancement bases sur l'analyse des artefacts produits plutôt que de multiples réunions

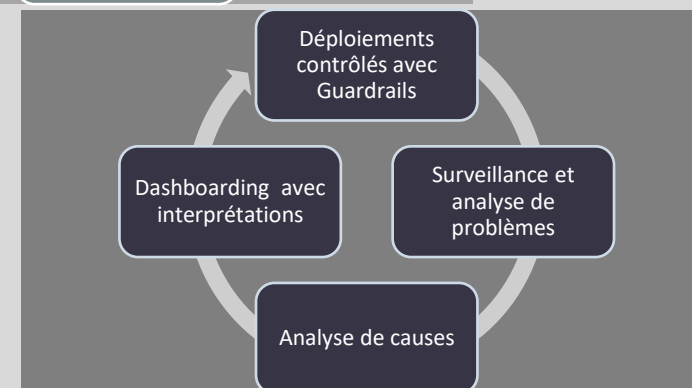
Analystes/Pilotes plutôt que codeurs/reviseurs/vérificateurs



... ce qui suppose un ROADMAP métier pilotée par la valeur (cf. [Agentic AI strategy | Deloitte Insights](#))



... ce qui impose la disponibilité d'un référentiel sémantique documentaire et technique ( see [MCP server/AI's new map](#))



... ce qui demande un GitOps efficient

# Nouvelles métriques ? Nouvelles unités ?

Time-to-first-Peek

Spec-as-code coverage

Ears-spec-to-code-automation ratio

AI-to human-reviews-ratio

Daily-meeting-reduction-trend

Des “story-points” aux “agent-runs”

Des “sprints” aux “agent-day”

Des “features et user-story” aux “transaction-flow”

# De « l'IA nous aide à livrer plus vite » à « l'IA nous aide à mieux décider »

**Hypothèse : la genAI permet de réduire le cycle de développement**



Exigence: choisir les éléments apportant le plus de valeurs aux clients, aux actionnaires/sociétaires aux employés, a la société



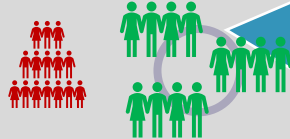
Conséquence : on "dépile" la roadmap technico fonctionnelle



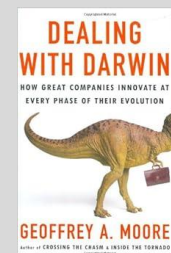
Question : La roadmap est-elle infinie ?



Note: « si la roadmap métier ne diminue pas alors la GenAI est probablement inutile »



Impacts : Redéploiement, re-intégration des sous-traitances, décommissionnements accélérés, renouvellement des équipes  
**R&D , innovation**

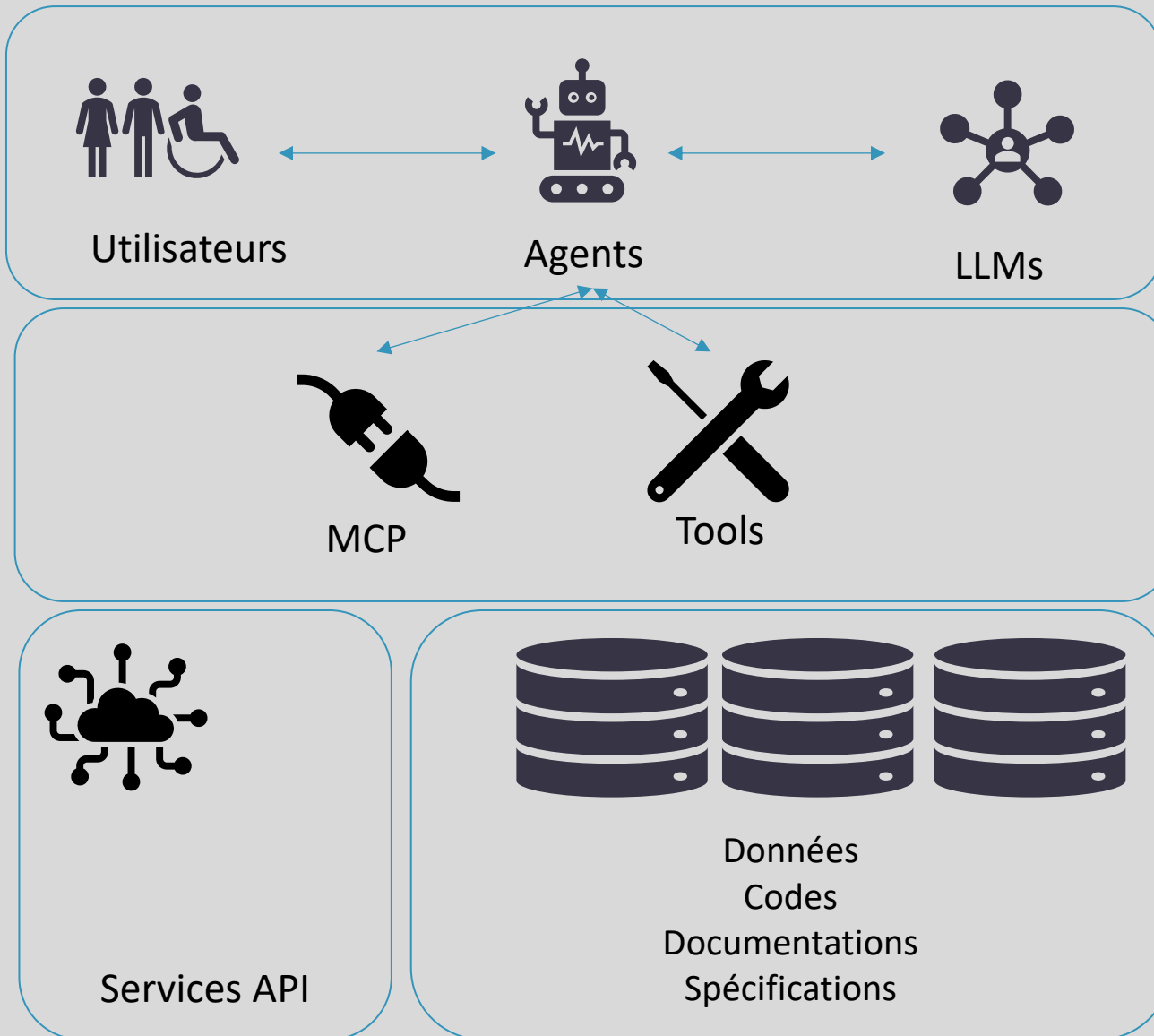


L'opportunité disruptive avec la GenAI et les systèmes agentiques n'est pas uniquement d'aider les équipes à « livrer la roadmap plus vite », mais de \*\*remettre continuellement en question si la roadmap métier devrait exister telle qu'elle est écrite\*\*

## Gouvernance Agentique des Roadmaps (GAR) ?

- Roadmap = hypothèses à prouver, pas promesses à tenir
- Preview rapide avant débat : montrer, mesurer, décider
- Des agents IA challengent chaque pari (usage, coût, risque, concurrence)
- Pas de preuve = on tue, pas "on espère"

# Techno-système en 2025



La gestion de contexte des LLMs reste le point dur.

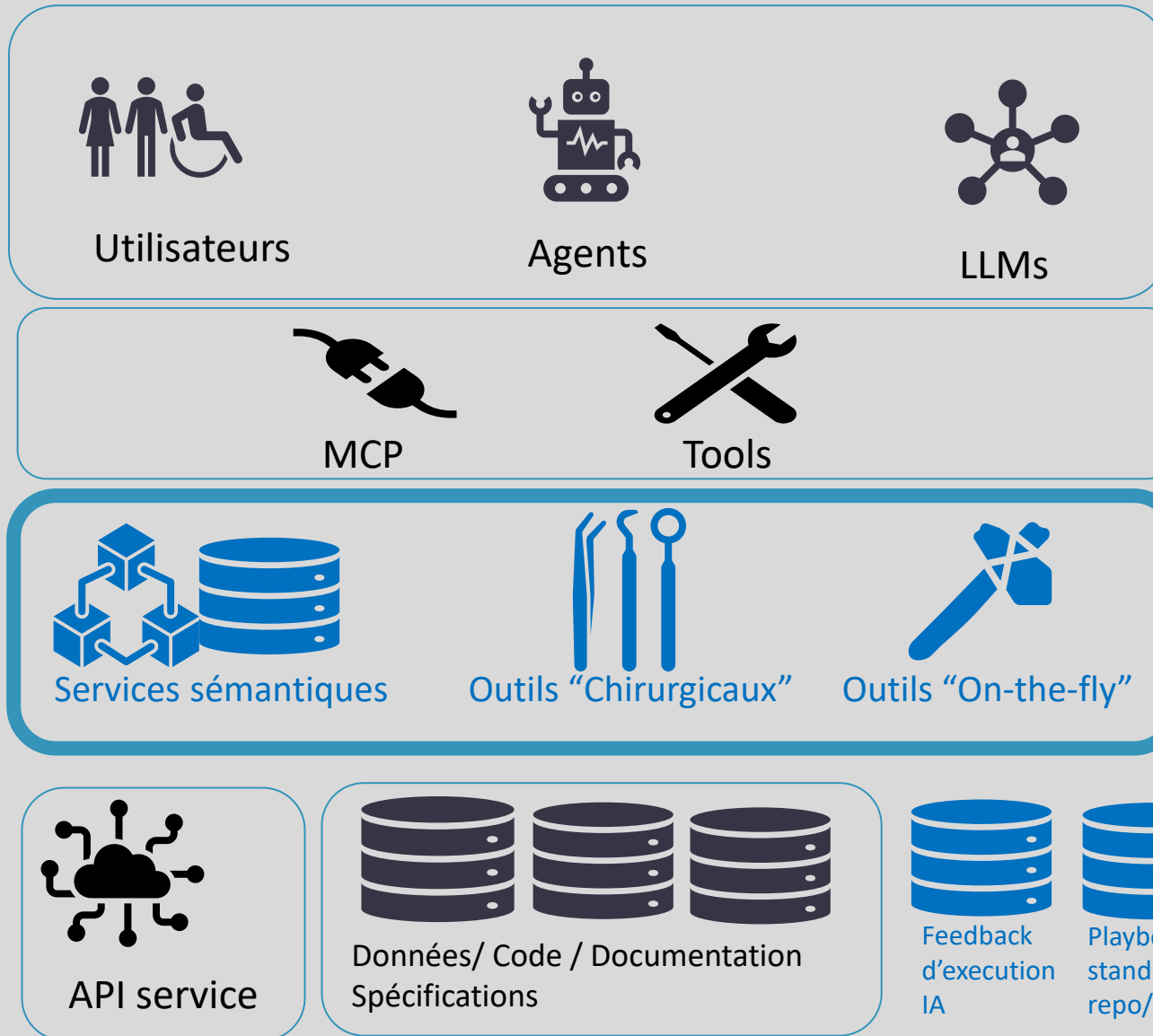
Les grosses applications ( >10 repos, >1 mLOC, >20 spécifications) ne sont pas gérables simplement.

La solution actuelle:

- Indexation
- « MCPisation »
- Outils spécifiques (ex : Code Mapper, « Hotspotizer », TreeSiter, LSP...)

**Il faut progresser sur l'identification et la collection en « quelques secondes » de l'ensemble de données « nécessaires et suffisantes »**

# Techno système cible en fin 2026



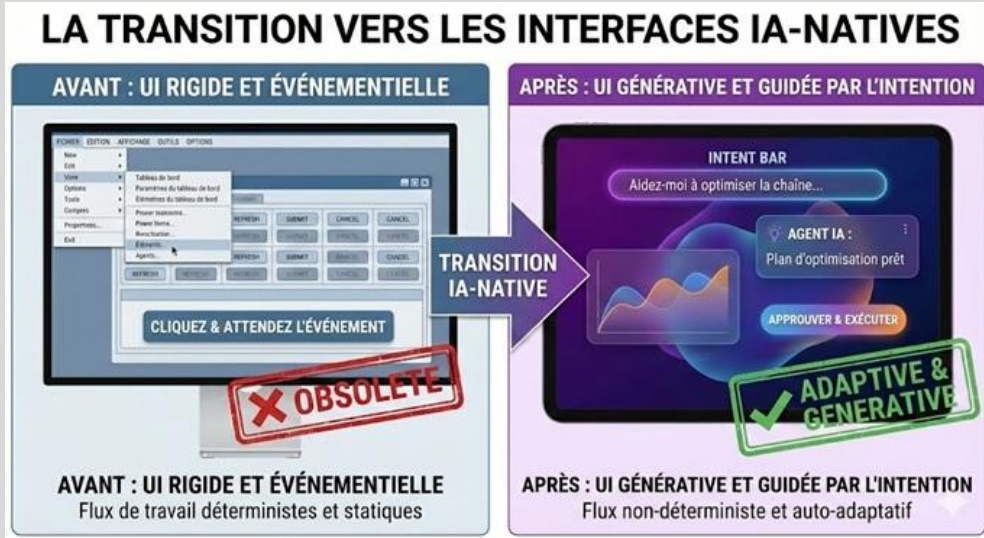
3 approches principales qui peuvent se compléter:

- Services sémantiques (ex. Cast Software)
- Boîtes à outils préparée, livrée dans des « skills ou agents » (cf. Olaf Skills)
- Génération « a la demande » par le LLM (cf. [MIT RLM - papier de recherche](#))

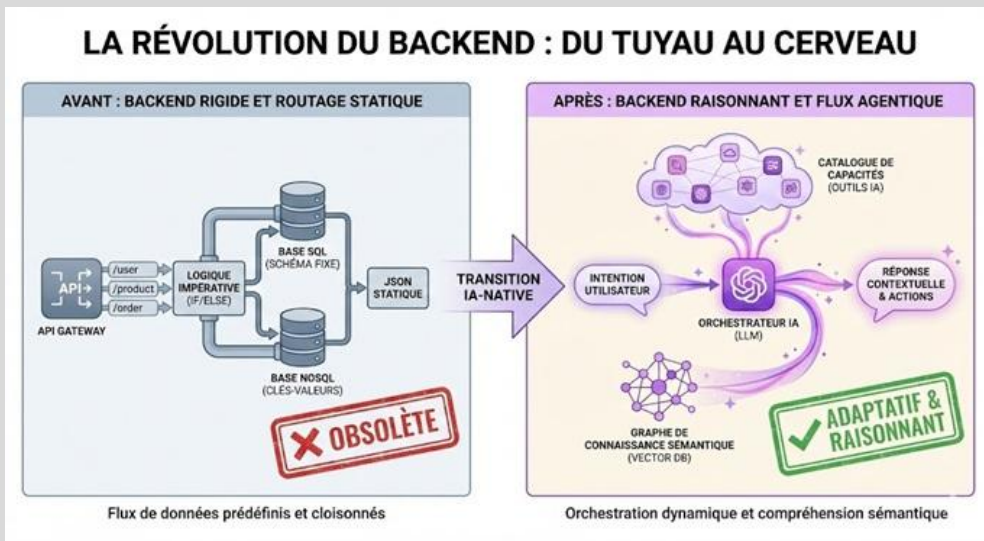
Les avantages seront très importants:

- Précision et complétude du contenu du contexte
- Réduction drastique du nombre d'échanges entre agents et LLMs
- Boucle de feedback en quelques minutes sur l'impact d'une modification de code
- Analyse d'impact et exploration en amont facilite et automatisable

# Repenser les applications avec l'IA : La fin du logiciel "Outil", l'avènement du logiciel "Partenaire"



<i>Concept Classique (Dépassé)</i>	<i>Concept IA-Native (Futur)</i>
Menu hiérarchique	Barre d'intention (Prompt/Voice)
Workflow rigide (Séquence)	Flux adaptatif (Agentique)
Pixels fixes (CSS)	Composants génératifs (Dynamic UI)
Erreur = Crash/Bug	Erreur = "Auto-Healing"



<i>Concept Classique (Dépassé)</i>	<i>Concept IA-Native (Futur)</i>
Routage statique (if/else/switch)	Chaînes de raisonnement Agentique
Endpoints : Swagger/OpenAPI	Capabilities/Tools/MCP
SQL/No-SQL = Schémas rigides	Vector et Graph database = auto-adaptatives
Performance optimisée pour la Vitesse brute	Performance optimisée pour la pertinence du contexte

# Repenser les applications avec IA : Les défis et danger du développement de ces nouvelles architectures applicatives

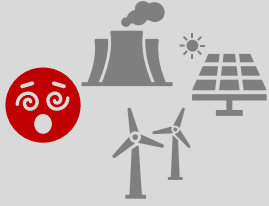
	Défi & Danger	Solution de Contrôle	Technologie / Outils
Qualité	Variabilité : Fin du Pass/Fail.	Scoring de confiance	Evals (LangSmith, Promptfoo)
Fiabilité	Hallucinations : Dérives métier.	Supervision par "Juge"	Multi-Agents (Deepchecks)
Sécurité	Prompt Injections : Vol d'intentions.	Filtres de contenu	Guardrails (AWS, NeMo)
Opérations	Coûts/Latence : Facture GPU.	Routage intelligent	Observabilité (Langfuse, Portkey)

Le **Modèle Juge** est une IA experte qui audite et score les réponses de votre application en temps réel pour garantir qu'elles respectent vos règles de sécurité et de métier.

Le modèle de "**Consensus d'Agents**" (type Minority Report) utilise plusieurs IA aux profils variés pour voter sur la validité d'une réponse, bloquant l'action si une majorité ou un membre détecte une anomalie

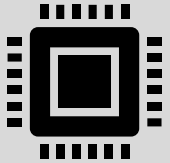


# Technologies pour supporter ce possible futur



« La contrainte principale c'est la consommation et la disponibilité d'Énergie ».

Les exigences majeures pour une agentique efficace ce sont la **vitesse d'inférence**, la réduction du « **context rot** », la capacité **d'utiliser de multiples modèles** spécialisés ou généraux suivant les cas d'utilisation.



[Cerebras Wafers](#) vs [Groq LPU](#)  
(cf. [licencing de groq par NVIDIA](#), )  
e.g., Windsurf SWE or GLM.4.x  
on Cerebras 960 t/s vs 50 t/s)

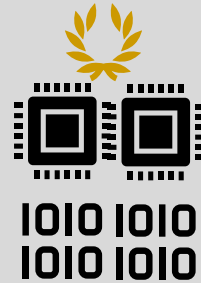
$O(n^2) \rightarrow O(n)$



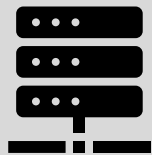
Amélioration des algorithmes d'attention ou variantes du modèle Transformer (cf. [Efficient Transformer Variants](#))



Modèles spécialisés (ex: [optim SFT](#))



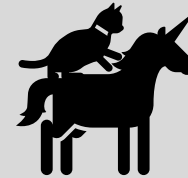
GPU vs LPU vs Asics (cf. [TPU v5 de Google](#) and [AWS Trainium](#))



Connections optique jusqu'à la PU (cf. [COUPE de TSMC](#) et [Lab IMEC](#), or [LightMatter](#))



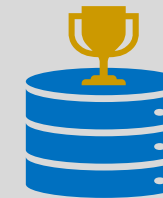
Conception des data centers plus économiques (cf. AWS 2.4 GW en Indiana)



Réduire les usages sans valeur ajoutée.  
Le marché de l'IT reste une très grosse niche.



Placer les agents sur la CPU existante - PC ou Cloud



**Réduire les Aller/Retour et augmenter la qualité du contexte**  
(cf. **sémantique des données, code, documentation**)

# Conclusions

Les propositions d'outillage et de services se sont multipliées et ont **fortement progressées en fonctionnalités**

Les gains de productivité potentiels se concentrent sur **une partie limitée des ingénieurs et analystes business...** mais restent une « **constante magique** » (+ 10%, x2 , x10 ?)

L'effort actuel porte sur les **pratiques d'utilisation** et sur **l'industrialisation des systèmes d'agents**

**La gestion de roadmap métier ET le SDLC vont devoir changer** afin de collecter les gains

Les technologies et les capacités **attaquent les exigences d'efficacité énergétique et de performance agentique**

2026 , l'année ou nous allons livrer et opérer des services d'agents en production ?

**DEPLOY and PROVE**