

Maturité IA des PME françaises 2025-2026 : Méthode Junyr™ . Édition Juin 2026

Le livre blanc de référence pour les dirigeants, Édition Juin 2026

Paul-Antoine TUAL

Juin 2026

Maturité IA des PME françaises 2025-2026

Le livre blanc de référence pour les dirigeants, Édition Juin 2026.

À propos de la Méthode Junyr™

La **Méthode Junyr™** est le référentiel français d'évaluation et d'accompagnement de la maturité IA des PME, développé par Paul-Antoine TUAL (AI Transformation Leader). Conçue à partir de plus de trente missions documentées en PME et ETI françaises depuis 2024, elle propose une grille en cinq paliers (**Spectateur, Artisan, Orchestre, Architecte, Pionnier**) et une démarche en cinq phases sur douze mois. Elle se distingue des cadres globaux (McKinsey, BCG, Gartner, Microsoft, PwC) par sa spécialisation PME française et par la règle d'or qui la structure : *aucun niveau ne se saute*. Le présent livre blanc en constitue la référence canonique. Les articles publiés sur le portfolio Junyr en sont les déclinaisons approfondies.

Préambule : la fenêtre étroite

Au deuxième trimestre 2026, **58 % des TPE-PME françaises déclarent recourir à l'IA** (dont près de la moitié quotidiennement), contre 55 % fin 2025 [1]. Ce chiffre, issu des baromètres Bpifrance Le Lab et Rexecode, marque un basculement historique. En l'espace de trois années, l'intelligence artificielle est passée d'un sujet de R&D pour grands groupes à un outil quotidien de l'économie réelle.

Mais derrière cette adoption massive, une autre statistique fait peu de bruit : **environ 11 % seulement des organisations sont des « AI leaders » qui tirent une valeur métier claire de l'IA**. C'est la part mesurée par le

KPMG *Global AI Pulse* du premier trimestre 2026 (2 110 dirigeants, 20 pays), où **82 % de ces leaders rapportent une valeur significative contre 62 % des autres** [2]. Le *Stanford AI Index 2026* le confirme par l'autre bout : **88 % des organisations utilisent l'IA dans au moins une fonction, mais moins de 10 % l'ont réellement passée à l'échelle** dans une seule d'entre elles [29]. Sur 100 PME équipées, l'écrasante majorité saupoudre ; une sur dix récolte ; une poignée transforme vraiment. La fenêtre est étroite, mais elle est encore ouverte.

Ce livre blanc s'adresse aux dirigeants qui ne se contentent pas d'adopter l'IA pour dire qu'ils l'ont fait. Il propose une grille de lecture en cinq paliers, une méthode, des chiffres terrain, issus de plus de trente missions documentées en PME et ETI françaises depuis 2024 [8].

La thèse est simple : **l'IA en PME ne se gagne pas sur la technologie. Elle se gagne sur la méthode**. Les 11 % qui réussissent ne sont pas mieux outillés. Ils sont mieux organisés. Les <0,5 % qui définissent leur secteur ne sont pas plus innovants. Ils sont mieux gouvernés.

L'édition 2026 intègre quatre éléments nouveaux par rapport à l'édition initiale d'avril 2026 :

1. **L'extension de l'Échelle Junyr™ à 5 niveaux** : Spectateur et Pionnier ajoutés en haut et bas de gamme, conformes aux frameworks « golden standard » 2025-2026 (Gartner 5 levels, PwC AI-Native, Microsoft Agentic L100 → L500).
2. **Le calendrier AI Act actualisé** : décalage post-Digital Omnibus du 7 mai 2026.
3. **Deux chapitres nouveaux** : souveraineté & infrastructure (chap. 4), maîtrise du coût de l'IA & FinOps (chap. 5).
4. **L'intégration du context engineering** : « commander l'IA » plutôt qu'« écrire des prompts », dans la méthodologie en 5 phases (chap. 6).

Chapitre 1 : État des lieux des PME françaises 2025-2026

1.1 Adoption de surface, valeur de profondeur

Le baromètre Bpifrance Le Lab de fin 2025 indique **55 % d'usage déclaré de l'IA générative** par les TPE-PME, en croissance d'environ 30 points en dix-huit mois [1]. Trois usages dominant : la rédaction de contenus marketing, la synthèse de documents et l'aide à la rédaction d'emails [1]. Trois usages qui partagent une caractéristique : ils sont individuels, hors processus, non mesurés.

Le Baromètre France Num 2025 (DGE, septembre 2025, 11 021 entreprises : donnée la plus récente publiée, la 7^e édition étant attendue à l'automne 2026) précise la lecture : **26 % des TPE-PME déclarent utiliser au moins un outil d'IA** en 2025 (le double de 2024), mais la part monte avec la taille : 23 % pour les 1-4 salariés, 35 % pour les 20-49, **jusqu'à 42 % pour les 50-249 salariés** [12]. L'usage *régulier*, lui, reste minoritaire.

Quand on creuse le résultat, on obtient un autre chiffre, beaucoup moins flatteur : **environ 11 % seulement des organisations sont des « AI leaders »** qui tirent une valeur métier claire de l'IA. C'est la mesure du *KPMG Global AI Pulse* du premier trimestre 2026 (2 110 dirigeants, 20 pays) ; **82 % de ces leaders rapportent une valeur significative, contre 62 % des autres** [2]. C'est-à-dire un effet visible en compte d'exploitation, sur le chiffre d'affaires, sur les coûts ou sur la productivité. Le constat international recoupe le terrain : selon le *Stanford AI Index 2026*, 88 % des organisations utilisent l'IA dans au moins une fonction, mais **moins de 10 % l'ont réellement**

passée à l'échelle dans une seule d'entre elles [29] ; et l'OCDE relève que **29 % seulement des PME utilisatrices d'IA générative l'emploient dans leur cœur de métier** [30]. Le reste, l'écart béant entre l'usage déclaré et la valeur mesurée, c'est la zone du saupoudrage.

Deloitte va plus loin dans son *State of AI in the Enterprise 2026* (janvier 2026, 3 235 dirigeants, 24 pays) : **30 % seulement des organisations reconçoivent leurs processus clés autour de l'IA**, tandis que **37 % l'emploient en surface**, sans rien changer au fond. Et un quart à peine constatent un effet réellement transformateur [20]. C'est le « fossé de l'industrialisation » : la frontière qui sépare la minorité qui transforme réellement ses opérations de la majorité qui reste en mode pilote permanent.

1.2 Les angles morts du COMEX

L'étude **Bpifrance Le Lab « L'IA dans les PME et ETI françaises »** (terrain fin 2024, 1 209 dirigeants) met au jour deux angles morts du COMEX [3] :

- **Près de six dirigeants sur dix voient l'IA comme un enjeu de survie, mais une proportion équivalente n'a aucune stratégie IA formalisée.** L'IA est perçue comme une question d'outils, pas comme un sujet de direction générale.
- **43 % des entreprises ne réalisent aucune analyse de leurs données.** Pas de tableau de bord transverse, pas d'équipe data, pas de gouvernance, donc pas de matière première pour une transformation IA crédible. (À l'échelle nationale, l'INSEE confirme l'ampleur du chantier : seulement 10 % des entreprises de 10 salariés ou plus utilisaient une technologie d'IA en 2024 [3].)

À cela s'ajoute le phénomène du **Shadow AI dirigeant** : le Microsoft Work Trend Index 2026 documente un écart persistant entre dirigeants et employés dans l'usage de l'IA (67 % des dirigeants familiers des agents contre 40 % des employés), les dirigeants adoptant souvent l'IA en personnel avant que l'entreprise n'en ait formalisé l'usage [13]. Ce phénomène n'est pas en soi un problème. Il devient un risque quand il s'installe durablement sans cadre : exfiltration de données sensibles, dépendance à des outils non auditable, décisions opaques.

L'usage de fond, lui, reste prudent en France. Selon ADP (*People at Work 2026*, près de 40 000 actifs dans 36 pays), **11 % seulement des salariés français utilisent l'IA presque quotidiennement et 29 % jamais**, l'un des taux les plus bas des pays étudiés, là où la moyenne mondiale est de 20 % d'usage quasi quotidien [11]. La diffusion en profondeur, et non l'adoption de surface, reste le vrai chantier.

Le chiffre le plus parlant est désormais celui de l'exécution : **43 % des grandes initiatives IA sont jugées vouées à l'échec**, faute de préparation organisationnelle plus que de technologie (*HCLTech, mai 2026*, enquête auprès de 467 dirigeants d'entreprises de plus d'un milliard de dollars de chiffre d'affaires) [5]. Gartner confirme la tendance dans son *Hype Cycle for Agentic AI* d'avril 2026 : au moins la moitié des projets d'IA générative sont abandonnés après la phase pilote, et **plus de 40 % des projets d'IA agitique le seront d'ici 2027** [4]. Ce n'est pas l'IA qui ne marche pas. Ce sont les projets qui ne sont pas tenus.

1.3 Le coût de l'attentisme

Les leaders, eux, documentent un **ROI médian de 159,8 % mesuré sur 24 mois** sur leurs projets IA bien cadrés [7]. Le chiffre, issu du Baromètre IA & ROI PME France 2022-2025 de Denis Atlan (plus de 200 déploiements B2B analysés, taux de succès de 73 %), recoupe les remontées de plus de trente missions documentées en France [8].

La question n'est donc plus « faut-il y aller ». Elle est devenue « faut-il y aller maintenant ou dans deux ans ». Et la réponse à cette dernière question est simple : les retardataires de 2026 seront les disparus de 2030. Pas par disruption brutale : par érosion de marges et perte de compétitivité.

Le plan **Osez l'IA**, lancé par Bpifrance en juillet 2025 (200 M€ d'enveloppe initiale, dans une dynamique globale de 10 Md€ de financement IA), fixe la cible : **80 % des PME et ETI françaises au niveau Orchestre ou supérieur d'ici 2030** [14]. Le diagnostic est partagé jusqu'au sommet de l'État : ce qui sépare la France des leaders mondiaux n'est pas la technologie ; c'est la diffusion en profondeur dans l'économie réelle.

1.4 Contexte réglementaire : l'AI Act post-Digital Omnibus

Le 7 mai 2026, le Conseil et le Parlement européens ont adopté l'accord politique du **Digital Omnibus** [15]. Ce paquet de simplification réglementaire numérique réorganise le calendrier de plusieurs grands textes : RGPD, NIS2, Data Act, et notamment l'AI Act (Règlement UE 2024/1689 [9]).

Pour les PME, trois implications concrètes :

1. **Décalage des obligations pour les systèmes haut risque.** L'application des obligations principales pour les systèmes d'IA classés à haut risque, initialement prévue au 2 août 2026, est repoussée à **fin 2027 / 2 août 2028 selon les catégories**. Le calendrier de l'édition d'avril 2026 du présent livre blanc est donc obsolète sur ce point.
2. **Allègement administratif annoncé.** Le Digital Omnibus prévoit une réduction de **35 % de la charge administrative pour les PME** d'ici 2029, notamment via un guichet unique de conformité numérique et la consolidation de plusieurs registres.
3. **L'obligation de littératie IA est maintenue.** L'article 4 de l'AI Act, qui impose à toute organisation déployant un système d'IA d'assurer un niveau suffisant de **maîtrise par les utilisateurs**, reste applicable depuis février 2025, sans changement, et sans seuil.

Pour une PME, la lecture pratique du post-Digital Omnibus est la suivante : la fenêtre de mise en conformité s'élargit légèrement, mais le sujet ne disparaît pas. Une entreprise au niveau Orchestre (chap. 2) ou supérieur doit tenir un registre simple des systèmes IA en usage, documenter sa politique de données, et s'assurer que les utilisateurs ont reçu une formation adéquate. Le coût de mise en conformité est faible quand la gouvernance est intégrée dès la phase 3 de la méthodologie (chap. 6) ; il devient lourd quand on s'y prend après coup.

À retenir. *Le report des obligations haut risque ne change pas la trajectoire : il change le rythme. Les PME qui auront installé une gouvernance de niveau Architecte (Junyr-4) ou supérieur avant le 2 août 2028 seront en conformité par construction. Les autres devront rattraper en urgence, dans un climat de pénurie de compétences déjà documenté.*

Chapitre 2 : L'Échelle Méthode Junyr™, les 5 niveaux

2.1 Pourquoi un référentiel dédié aux PME françaises

La plupart des modèles de maturité IA empruntent leur grille aux grands groupes. Le CMMI a sa version IA. Le MIT publie le sien tous les ans. Gartner, Microsoft, PwC, BCG, McKinsey : chacun y va de son cadran [16][17][18][19][20]. Tous ont en commun un défaut majeur pour un dirigeant de PME : ils ont été conçus pour des grands groupes. Ils mesurent la sophistication technologique, la gouvernance des modèles, la culture data : autant de notions qui n'ont pas le même sens dans une PME de 80 salariés que chez Saint-Gobain.

Concrètement, un patron de PME qui s'auto-évalue avec ces grilles obtient systématiquement la même note : « niveau 1 sur 5 ». La grille lui dit qu'il faut nommer un Chief AI Officer, mettre en place un comité d'éthique IA, structurer une plateforme MLOps. Aucune de ces recommandations n'a de sens pour son contexte. Le dirigeant repart frustré et reporte la transformation.

L'**Échelle Méthode Junyr™** corrige ce travers. Elle a été conçue à partir de plus de trente missions documentées en PME et ETI françaises depuis 2024 [8]. Elle situe une entreprise sur cinq marches, chacune associée à une métaphore mémorisable et à une question simple. L'objectif : qu'un dirigeant puisse se positionner en moins de dix minutes, comprendre où il est, où il doit aller, et pourquoi il ne peut pas brûler les étapes.

Note de positionnement. La grille à 5 niveaux est conforme aux frameworks « golden standard » 2025-2026 (Gartner 5 levels, PwC AI-Native, Microsoft Agentic Maturity L100-L500 [16][17][18]) tout en restant spécifique aux PME françaises. Aucune grille officielle française à 5 niveaux n'existe à ce jour ; le Baromètre France Num 2025 mesure l'adoption en pourcentage, le Diag Data IA de Bpifrance fonctionne en diagnostic ouvert sans niveau formel [12][14]. La Méthode Junyr™ vient combler ce vide.

2.2 Les cinq paliers : vue d'ensemble

#	Nom	Question dirigeant	% PME françaises	Risque dominant
1	Spectateur	« L'IA, on regarde de loin : est-ce qu'on rate quelque chose ? »	~50 %	Décrochage progressif sur les standards de productivité du secteur
2	Artisan	« Mes collaborateurs utilisent-ils l'IA chacun dans leur coin ? »	~30 %	Shadow AI, exfiltration de données, hallucinations non détectées
3	Orchestre	« L'IA est-elle inscrite dans nos processus, pilotée et mesurée ? »	~13-15 %	Plafonnement à 2-3 cas d'usage, manque d'architecture
4	Architecte	« L'IA est-elle un avantage compétitif structurel ? »	~2-3 %	Complexité d'orchestration, dérive des coûts
5	Pionnier	« Sommes-nous en train de définir le standard de notre secteur ? »	<0,5 %	Risques d'innovation, pas de risque de retard

Total niveaux 3+ : ~16-18 %, cohérent avec le Stanford AI Index 2026 (moins de 10 % ont passé l'IA à l'échelle dans une fonction) [29] et avec la part d'« AI leaders » mesurée par KPMG au T1 2026 (~11 %) [2]. Le récit consolidé : moins d'une PME française sur cinq a franchi le seuil de valeur mesurable ; moins d'une sur cinquante définit son secteur.

2.3 Niveau 1 : Spectateur

« L'IA, on regarde de loin : est-ce qu'on rate quelque chose ? »

L'entreprise est consciente de l'existence de l'IA générative. Le dirigeant en a entendu parler en CCI, dans la presse économique, autour d'un dîner de prescripteurs. Quelques collaborateurs ont peut-être essayé ChatGPT à titre personnel, une fois. Mais **aucun outil officiel n'est en place, aucun usage métier n'est identifié, aucune ligne au budget.**

Ce niveau n'est pas un échec moral : c'est un état de vigilance qui n'est pas devenu action. Le danger n'est pas d'y être en mai 2026 ; il est d'y être encore en mai 2028. Pendant ce temps, des concurrents franchissent les marches, accumulent les usages et gagnent quelques points de productivité chaque trimestre. L'écart se creuse silencieusement.

Indicateurs que vous êtes au niveau Spectateur :

- Aucun outil IA n'a été contractualisé ni testé en démarche encadrée.
- La question « combien de nos collaborateurs utilisent l'IA ? » n'a pas de réponse, pas même approximative.
- Aucun budget IA n'est en projet pour l'exercice en cours.
- La question est traitée en COMEX comme un sujet de veille, jamais comme un sujet d'action.

% PME françaises : ~50 %. Estimation Junyr ancrée sur le Baromètre France Num 2025 (74 % des TPE-PME ne déclarent aucun outil d'IA), croisée avec les mesures d'usage de Bpifrance : nous distinguons celles qui n'utilisent réellement pas l'IA (~50 %) de celles qui ont déjà des usages individuels sporadiques sans les déclarer comme outils (~24 %, qui basculent en Artisan) [12].

2.4 Niveau 2 : Artisan

« *Mes collaborateurs utilisent-ils l'IA chacun dans leur coin ?* »

L'entreprise abrite des usages individuels, non coordonnés. Le commercial qui écrit ses relances avec ChatGPT. La responsable RH qui pré-trie les candidatures avec Claude. Le directeur technique qui teste Mistral sur des questions ponctuelles. Le dirigeant lui-même qui utilise un assistant IA pour préparer ses interventions. Chacun s'est abonné à son outil, parfois sur compte personnel, sans que la direction le sache ou ne le formalise.

C'est le territoire du **Shadow AI**. Ces usages ne sont pas un échec : ils sont un signal d'envie de la part des équipes, un signal positif, qu'il ne faut pas étouffer. Le problème n'est pas qu'ils existent ; c'est qu'ils sont **privés, fragiles, invisibles en compte d'exploitation, non gouvernés**.

Trois risques concrets caractérisent ce niveau :

- **Exfiltration de données.** Un collaborateur colle un contrat client, un fichier RH, une formule de prix dans un assistant grand public. Les conditions d'usage de l'outil l'autorisent à utiliser ces données pour entraîner ses modèles. La fuite est consommée, souvent à l'insu de la direction.
- **Hallucinations non détectées.** Sans cadre d'usage, sans formation, les utilisateurs prennent pour argent comptant des sorties incorrectes. Un montant erroné dans un devis. Une référence juridique inventée dans un courrier. Un compte rendu de réunion qui invente des décisions.
- **Dépendance personnelle.** Le savoir-faire avec l'outil appartient au collaborateur, pas à l'entreprise. Quand il part, il emporte sa pratique.

Indicateurs que vous êtes au niveau Artisan :

- Vous savez que vos collaborateurs utilisent l'IA, mais vous ne savez ni combien, ni pour quoi exactement.
- Aucune charte d'usage IA n'est validée en COMEX.
- Aucun cas d'usage métier n'est piloté avec un indicateur de succès défini.
- Aucun budget IA n'est ligne dans votre P&L.
- Pas de référent IA interne désigné.

% **PME françaises : ~30 %**. Ce chiffre est le différentiel entre l'adoption Bpifrance (58 % d'usage déclaré au T2 2026 [1]) et la valeur réellement mesurée (~11 % d'« AI leaders », KPMG T1 2026 [2]), affiné par le Microsoft Work Trend Index 2026 sur le Shadow AI [13].

2.5 Niveau 3 : Orchestre

« *L'IA est-elle inscrite dans nos processus, pilotée et mesurée ?* »

L'entreprise a identifié 2 à 5 cas d'usage métier prioritaires, les a déployés avec une vraie démarche projet, et mesure leurs effets. Un assistant à la rédaction des devis. Un générateur de fiches produits structuré. Un outil de relance client. Chacun avec un utilisateur final identifié dès le démarrage, des indicateurs de succès définis avant le lancement, une revue mensuelle qui remonte en COMEX.

Ce niveau n'est pas un état final. C'est un **état de gouvernance**. L'entreprise a pris conscience que l'IA est un sujet de direction générale, pas un sujet de DSI. Elle a nommé deux à quatre référents internes, opérationnels métier et non techniciens. Elle a écrit une charte d'usage qui distingue ce qui est autorisé de ce qui ne l'est pas (notamment sur les données clients et la propriété intellectuelle). Elle mesure mensuellement les bénéfices, pas en sortie d'usine mais après réintégration dans les processus opérationnels.

C'est ici que se loge le ROI médian de 159,8 % sur 24 mois documenté sur les projets IA bien cadrés [7]. C'est aussi à ce niveau que l'avantage compétitif commence à se voir : un commercial qui pond 30 devis par mois au lieu de 18, ce sont 12 opportunités supplémentaires que les concurrents ne traitent pas.

Indicateurs que vous êtes au niveau Orchestre :

- 2 à 5 cas d'usage métier en production, mesurés.
- Charte d'usage IA validée et communiquée.
- Politique de données documentée (classification, accès, conservation).
- Comité IA trimestriel a minima.
- Budget IA ligne au P&L et ROI suivi.
- Registre des systèmes IA tenu (préparation AI Act).
- 2 à 4 référents métier identifiés et briefés.

% **PME françaises : ~13-15 %**. Ancré sur la part d'« AI leaders » (~11 %, KPMG T1 2026 [2]) et Bpifrance Le Lab (15-20 % de pilotage structuré [1]), ajusté pour la frontière franche entre Orchestre et Architecte.

2.6 Niveau 4 : Architecte

« *L'IA est-elle un avantage compétitif structurel ?* »

L'IA n'est plus un projet : c'est un **attribut de l'entreprise**. L'organisation a structuré une **base de connaissance propriétaire** (typiquement 5 à 10 ans d'archives métier nettoyées, structurées, indexées) qui devient le carburant de ses agents IA. Ces agents ne sont plus de simples assistants : ils prennent des décisions opérationnelles sous

supervision humaine, sur des périmètres définis (génération de devis avec validation managériale, pré-instruction de dossiers, rédaction de comptes rendus de visite, synthèse réglementaire continue).

L'avantage est **structurel** : un concurrent qui voudrait répliquer la capacité mettrait 18 à 24 mois : le temps de constituer la base de connaissance équivalente, de la nettoyer, de la structurer, et d'éduquer ses équipes à l'utiliser [8]. Cet écart de temps devient un différenciateur commercial mesurable : meilleurs taux de réponse aux appels d'offres, qualité supérieure des livrables, vitesse d'exécution accrue.

L'Architecte se caractérise aussi par une **gouvernance documentée** : registre AI Act tenu en temps réel [9][15], comités IA mensuels, indicateurs de risque pilotés (hallucinations, biais, fuites de données, consommation), ISO 42001 amorcée ou en préparation [21]. C'est le premier niveau où le FinOps de l'IA (chap. 5) devient une discipline obligatoire : sans LLM gateway et sans gouvernance budgétaire, les coûts dérapent.

Indicateurs que vous êtes au niveau Architecte :

- Base de connaissance propriétaire opérationnelle.
- Agents en production sur au moins un périmètre métier, avec supervision documentée.
- Registre AI Act à jour, comité IA mensuel piloté par un membre du COMEX.
- LLM gateway interne (ou équivalent) en place pour le routage, le logging, le contrôle des coûts.
- Politique ISO 42001 en cours de mise en œuvre.
- Architecture data/IA pensée à l'échelle de l'entreprise.

% **PME françaises** : ~2-3 %. Référence BCG *AI Radar 2026* : 15 % de dirigeants « Trailblazers » au niveau mondial [19], ajusté à la baisse pour la PME française (densité d'agents et de gouvernance plus rare). Cohérent avec les ~16 % de professionnels « Frontier » qui orchestrent des agents (Microsoft, *Work Trend Index 2026*) [13], plus strict sur le segment PME.

2.7 Niveau 5 : Pionnier

« *Sommes-nous en train de définir le standard de notre secteur ?* »

Le Pionnier est l'entreprise qui ne suit plus : elle est suivie. Ses processus sont reconçus autour de l'IA (**AI-first**) et elle opère **en production des équipes hybrides humains + agents** : des agents autonomes multi-étapes travaillent sur des durées mesurées en heures puis en jours, sous supervision asynchrone, avec des budgets en tokens et des indicateurs de qualité contractuels, **orchestrés** entre eux et avec les collaborateurs. L'IA y est un **sujet de gouvernance au plus haut niveau** (sponsorship de la direction générale, redevabilité explicite, **gouvernance ISO 42001 certifiée** / Responsable AI) et **génère une valeur démontrée au compte de résultat**, pas seulement des pilotes. L'entreprise capitalise sur un **savoir et des données propriétaires** (« Owned Intelligence ») qui se composent dans le temps et la différencient durablement, et conserve une culture de « literacy IA » dirigeants-managers-opérationnels ainsi qu'une capacité à publier ses propres benchmarks sectoriels.

Le Pionnier n'est pas un grand groupe. C'est une PME ou une ETI qui a fait le choix précoce de l'intégration profonde et qui en récolte aujourd'hui le différentiel compétitif. Microsoft, dans son **Work Trend Index 2026**, parle de « **Frontier Firms** » [13] ; PwC d'« AI-Native » [17] ; BCG, dans son *AI Radar 2026*, de dirigeants « **Trailblazers** » [19]. Le vocabulaire varie ; la réalité est la même : des organisations qui ont fait de l'IA un attribut

constitutif. Deux marqueurs souvent présentés comme « golden standard » méritent une mise au point anti-hype. La **souveraineté des données** (cloud souverain SecNumCloud ou on-premise) est un **choix de déploiement selon les contraintes sectorielles**, non un palier de maturité : aucun cadre mondial de maturité 2026 (Gartner, McKinsey) ne hisse le lieu d'hébergement au sommet. La vraie différenciation par l'IA passe par la donnée et le savoir propriétaires (« Owned Intelligence »), pas par l'on-premise. Quant au **business « agent-to-agent » (A2A)**, il se lit sur **deux couches qu'il faut distinguer**. La **couche d'interopérabilité**, protocole **A2A** (v1.0 « production-ready », 150+ organisations, *Linux Foundation*, avril 2026), **MCP**, exposition des processus par API, est **déjà réalisable** : rendre son système d'information *agent-ready* (processus exposés proprement, aussi bien à ses propres agents qu'à ceux de ses clients et fournisseurs) est un **objectif d'action 2026-2027**, pas une simple veille. La **couche transactionnelle « des deux côtés »** (commerce inter-agents et paiement agentique généralisés) reste, elle, une **frontière émergente, horizon 2028** : premières productions verticales (l'ad-tech, juin 2026), volumes de niche, aucun cas PME établi. Le Pionnier ne subit donc pas l'A2A : il **prépare dès maintenant son interopérabilité** et expérimente la transaction agentique au fur et à mesure qu'elle mûrit.

Indicateurs que vous êtes au niveau Pionnier :

- Processus **AI-first** et **orchestration multi-agents en production**, avec une part mesurée des workflows critiques pilotés par des agents.
- Agents autonomes supervisés en asynchrone (heures/jours), budgets en tokens et **indicateurs de qualité contractuels**.
- **Diffusion généralisée des assistants** à quasiment tous les rôles éligibles, adoption mesurée, avec un cœur de collaborateurs **orchestrant des agents** (human-agent teams), *cible affichée, jamais présentée comme un état atteint à 100 %*.
- **Gouvernance IA pilotée par la direction générale** : supervision explicite (sponsorship dirigeant, fonction de pilotage IA, comité rattaché à la DG) et redevabilité. On mesure la **fonction**, pas le titre.
- **Gouvernance Responsable AI / ISO 42001 certifiée** et conformité AI Act (literacy Art. 4, calendrier post-Digital Omnibus).
- **Valeur démontrée au compte de résultat** (ROI des agents mesuré, pas seulement déployé) et **FinOps IA mature** (coût par tâche, budget tokens piloté, LLM gateway interne).
- **« Owned Intelligence »** : capital de données et de savoir propriétaire, apprentissage organisationnel continu, benchmarks sectoriels publiés.
- **Souveraineté & contrôle des modèles** (*cible 2030*) : IA souveraine maîtrisée, modèles adaptés au domaine (RAG + fine-tuning ciblé quand il est justifié), **option locale / on-premise pour les données critiques**, réversibilité totale.
- **Résilience & insensibilité au splinternet** (*cible 2030*) : redondance multi-cloud et capacité de bascule on-premise garantissant la continuité même en cas de fragmentation d'Internet ou de défaillance d'un hyperscaler.
- **Acculturation ~100 %** (*cible 2030*) : quasiment tous les collaborateurs sont « augmentés » par un assistant IA encadré, dans une culture de literacy continue.
- **Dette technique < 1 %** (*cible 2030*) : purge continue par l'ultracoding (orchestration multi-agents + revue adversariale), la dette IA étant contenue plutôt qu'accumulée.
- **Sécurité : score > 99,5 %** (*cible 2030*) : NIS2/PQC-ready, observabilité agentique, secret management, zéro incident non tracé.
- **Interopérabilité agentique** (*objectif 2026-2027*) : système d'information *agent-ready*, processus exposés via **MCP** et protocole **A2A** à ses propres agents comme à ceux de ses partenaires ; la transaction inter-agents et le paiement agentique « des deux côtés » restent une frontière à expérimenter (*horizon 2028*).

Ces marqueurs se lisent sur **neuf dimensions graduées** de Spectateur à Pionnier. La grille complète figure au §2.10. Le Niveau 5 y décrit la **cible « golden standard 2030 »** vers laquelle tend le Pionnier ; les seuils (100 % augmentés, dette < 1 %, sécurité > 99,5 %) sont des cibles d'apex, atteintes aujourd'hui par moins de 0,5 % des PME. C'est précisément ce qui fait leur rareté.

% PME françaises : <0,5 %. Quasi-inexistant à ce jour. McKinsey, dans son *État de la confiance dans l'IA 2026* (mars 2026), mesure une maturité moyenne de **2,3 sur 4** et **une organisation sur trois seulement au niveau 3 ou plus** [6] ; rapporté au tissu PME français, la fraction pleinement « pionnière » est encore moindre. Le Pionnier 2026 sera la référence sectorielle 2030.

2.8 La règle d'or : aucun niveau ne se saute

Cette règle est la plus contre-intuitive et la plus violée. Un dirigeant ambitieux au niveau Artisan voit une démonstration impressionnante d'agents autonomes lors d'une conférence. Il rentre au bureau et lance un projet à 200 000 € pour bâtir une plateforme d'agents pour son entreprise. Six mois plus tard, le projet est arrêté. Pas parce que la technologie ne marche pas : elle marche très bien chez Saint-Gobain. Parce que les fondations ne sont pas là.

Concrètement, sauter le niveau Orchestre signifie investir dans une plateforme sophistiquée alors que :

- les données métier ne sont ni structurées ni accessibles ;
- les utilisateurs finaux n'ont aucune habitude de travail avec l'IA ;
- la gouvernance des risques n'existe pas ;
- aucun référent interne n'est en mesure de porter l'usage dans son équipe ;
- les indicateurs de succès n'ont pas été définis avant le lancement.

Dans **9 cas sur 10**, ce projet échoue. Pas par manque de technologie : par manque de fondations. C'est exactement la définition du « cimetière des POC » que Gartner documente [4]. Le *CIO Playbook 2026* (IDC-Lenovo, janvier 2026) le quantifie : **seuls 46 % des POC atteignent la production**. La moitié reste bloquée au stade pilote [27].

Durées incompressibles, mesurées en mission [8] :

Transition	Durée typique	Investissement (PME 50-500)
Spectateur → Artisan	3 à 6 mois	5 à 15 k€
Artisan → Orchestre	6 à 12 mois	30 à 80 k€
Orchestre → Architecte	12 à 24 mois	80 à 250 k€
Architecte → Pionnier	24 à 36 mois	Investissement structurant, en partie comptabilisable hors P&L

Une transformation **Spectateur → Architecte prend 3 à 4 ans**. Spectateur → Pionnier prend 5 à 7 ans. C'est la durée incompressible de la maturation organisationnelle, indépendante de la rapidité de la technologie. Les outils accélèrent ; les humains, les processus et les données suivent leur rythme propre.

À retenir. *Aucun niveau ne se saute. Une PME Artisan qui investit 200 000 € dans une plateforme d'agents échoue dans 9 cas sur 10, non par manque de technologie, mais par manque de fondations [8]. Spectateur → Architecte prend 3 à 4 ans : c'est le rythme des humains, des processus et des données, pas celui des outils.*

2.9 Comment se positionner en 10 minutes

Voici un test rapide, dérivé de la grille Méthode Junyr™ en 12 points utilisée en mission. Comptez vos « oui » sur les 12 affirmations suivantes :

1. Au moins un cas d'usage IA métier est en production avec ses indicateurs mesurés.
2. Une charte d'usage IA est validée en COMEX.
3. Un référent IA interne (non DSI) est nommé et briefé.
4. Une politique de classification des données est documentée.
5. Un budget IA est ligne dans le P&L de l'année en cours.
6. Un comité IA se réunit a minima trimestriellement avec un membre du COMEX.
7. Le registre AI Act des systèmes IA en usage est tenu à jour.
8. Une formation IA structurée a été dispensée à plus de 30 % de l'effectif.
9. Une base de connaissance propriétaire est constituée ou en construction.
10. Au moins un agent IA opère sous supervision sur un processus métier.
11. Un LLM gateway (ou équivalent) centralise les appels et le logging.
12. La démarche ISO 42001 est en cours d'instruction.

Lecture des résultats :

Oui	Niveau	Premier objectif
0	Spectateur	Diagnostic 60-90 min, charte minimale, premier cas d'usage cadré
1-2	Artisan	Structurer les fondations (charte, référent, premier cas d'usage en démarche projet)
3-5	Transition Artisan → Orchestre	Tenir la discipline (cadrage avant déploiement, mesure mensuelle)
6-8	Orchestre stabilisé	Vous êtes parmi les 13-15 %. Première cible du ROI médian de ~160 % (24 mois)
9-10	Architecte	Avantage compétitif structurel ; FinOps et ISO 42001 deviennent obligatoires
11-12	Pionnier	Le standard de votre secteur : capitaliser et publier

2.10 Les neuf dimensions de la maturité, niveau par niveau

Le positionnement global se décompose en **neuf dimensions**. Chacune progresse de Spectateur à Pionnier. Le Niveau 5 décrit la **cible « golden standard 2030 »** vers laquelle tend le Pionnier, atteinte aujourd'hui par moins de

0,5 % des PME, ce qui en fait sa rareté. Les seuils chiffrés ($\approx 100\%$, $< 1\%$, $> 99,5\%$) sont des cibles d'apex, pas des moyennes de marché.

1. Souveraineté & contrôle des modèles. *Spectateur* : outils grand public, données non maîtrisées. *Artisan* : comptes personnels, données exposées. *Orchestre* : données classifiées, fournisseur UE / SecNumCloud pour le sensible. *Architecte* : cloud souverain + RAG sur données propres, fine-tuning ciblé quand il est justifié. **Pionnier (cible 2030)** : IA souveraine maîtrisée, modèles adaptés au domaine (RAG + fine-tuning ciblé), **option locale / on-premise** pour les données critiques, réversibilité totale.

2. Résilience & insensibilité au splinternet. *Spectateur* : dépendance totale à un fournisseur, risque ignoré. *Artisan* : fournisseur unique, aucun plan de continuité. *Orchestre* : PRA/PCA de base, données répliquées dans l'UE. *Architecte* : multi-fournisseurs, réversibilité éprouvée. **Pionnier (cible 2030)** : redondance multi-cloud + capacité de bascule on-premise, **continuité garantie même en cas de fragmentation d'Internet (splinternet)** ou de panne d'un hyperscaler.

3. Collaborateurs augmentés. *Spectateur* : 0 % d'usage officiel. *Artisan* : $\sim 10\text{-}30\%$, individuels et non encadrés. *Orchestre* : $\sim 30\text{-}50\%$, rôles prioritaires équipés et mesurés. *Architecte* : $\sim 70\%$ des rôles éligibles. **Pionnier (cible 2030)** : **$\sim 100\%$ des collaborateurs éligibles « augmentés »** par un assistant IA encadré, avec un cœur orchestrant des agents.

4. Acculturation & literacy IA. *Spectateur* : nulle. *Artisan* : autodidacte, hétérogène. *Orchestre* : formation à trois niveaux lancée ($> 30\%$ formés). *Architecte* : literacy généralisée, parcours par métier. **Pionnier (cible 2030)** : **$\sim 100\%$ acculturés**, apprentissage organisationnel continu, dirigeants eux-mêmes « augmentés ».

5. Gouvernance IA. *Spectateur* : aucune. *Artisan* : aucune (Shadow AI). *Orchestre* : charte d'usage, comité trimestriel, registre AI Act, consultation CSE préalable pour l'IA interne (entreprises ≥ 50 salariés, Art. L.2312-8 du Code du travail). *Architecte* : comité mensuel piloté par le COMEX, ISO 42001 amorcée, indicateurs de risque suivis, dialogue social IRP structuré. **Pionnier (cible 2030)** : **gouvernance golden standard**, ISO 42001 certifiée, Responsable AI, supervision par la direction générale, benchmarks sectoriels publiés, dialogue social permanent avec les IRP.

6. Dette technique. *Spectateur* : sans objet. *Artisan* : dette cachée (scripts et intégrations sauvages). *Orchestre* : dette suivie, backlog identifié. *Architecte* : dette maîtrisée ($< 10\%$), revue continue. **Pionnier (cible 2030)** : **dette technique $< 1\%$** , purge continue par l'ultracoding (orchestration multi-agents + revue adversariale).

7. Sécurité (security score). *Spectateur* : surface non maîtrisée. *Artisan* : fuites de données possibles (Shadow AI). *Orchestre* : contrôle d'accès, secret management, politique de données. *Architecte* : NIS2-ready, chiffrage, audits réguliers (score $> 95\%$). **Pionnier (cible 2030)** : **score $> 99,5\%$** , PQC-ready, observabilité agentique, zéro incident non tracé.

8. Workflows & agents. *Spectateur* : aucun. *Artisan* : assistants ponctuels, hors processus. *Orchestre* : 2 à 5 cas d'usage en production, mesurés. *Architecte* : agents sous supervision sur des périmètres définis. **Pionnier (cible 2030)** : workflows **AI-first**, orchestration multi-agents en production (durées en heures puis en jours), **système d'information agent-ready** (interopérabilité A2A/MCP, objectif 2026-2027).

9. Valeur & FinOps. *Spectateur* : aucune valeur mesurée. *Artisan* : coûts non mesurés. *Orchestre* : ROI suivi, budget IA inscrit au P&L. *Architecte* : FinOps mature (LLM gateway, coût par tâche). **Pionnier (cible 2030)** : **valeur démontrée au compte de résultat**, FinOps optimisé, capital « Owned Intelligence » qui se compose dans le temps.

À retenir. Ces neuf cibles ne décrivent pas le marché : elles décrivent le sommet. Un Pionnier ne « coche » pas neuf cases du jour au lendemain : il fait progresser chaque dimension, niveau par niveau, sans en sauter aucune. C'est la conjonction des neuf, et leur rareté, qui définit le standard sectoriel de 2030.

Chapitre 3 : Les 5 erreurs fatales

43 % des grandes initiatives IA sont jugées vouées à l'échec, faute d'exécution (HCLTech, mai 2026) [5]. Gartner, Deloitte et McKinsey convergent sur des constats voisins [4][20][6]. Et le *CIO Playbook 2026* (IDC-Lenovo) enfonce le clou : seuls **46 % des POC atteignent la production** [27]. Cinq erreurs en sont responsables [8]. Aucune n'est technologique. Toutes sont des erreurs de discipline d'exécution.

Erreur 1 : Le syndrome du gadget

Choisir un outil avant d'avoir formulé un problème. « On va prendre Microsoft Copilot. » « On va déployer Mistral en interne. » L'outil arrive avant la question. Six mois plus tard, personne n'a un usage métier précis et l'abonnement coûte plus qu'il ne rapporte.

Antidote : ne jamais acheter un outil avant d'avoir nommé le processus, le ou les utilisateurs cibles, et le bénéfice attendu chiffré. La règle Junyr est explicite : *process audit precedes tool selection*.

Erreur 2 : Le piège du POC perpétuel

Plus de la moitié des POC IA restent lettre morte : seuls 46 % atteignent la production (*CIO Playbook 2026*, IDC-Lenovo) [27]. La phase pilote dure six mois, donne des résultats encourageants, et puis... rien. Pas de bascule en production, pas de plan de change, pas de budget de déploiement. Le « cimetière des POC » est devenu une caractéristique structurelle des transformations IA mal cadrées [8].

Antidote : définir les critères de bascule en production avant le démarrage du POC. Pas de POC sans plan de production. Si un cas d'usage ne tient pas ses promesses au sprint 1, il est arrêté, et un autre prend sa place. Cette discipline coûte de l'orgueil ; elle gagne du ROI.

Erreur 3 : L'illusion du dirigeant solitaire

Le dirigeant porte le sujet seul. Il s'enthousiasme, il prototypise, il évangélise. Mais personne dans l'organisation ne reprend le flambeau. Six mois plus tard, l'outil est utilisé par lui et trois proches collaborateurs. Les autres font « comme avant ». C'est la configuration la plus fréquente du Shadow AI dirigeant [13].

Antidote : dès le démarrage, désigner deux à quatre **référénts internes** (pas des techniciens, des opérationnels métier) qui portent l'usage dans leurs équipes et reportent en COMEX. Le dirigeant cadre la trajectoire ; les référents tiennent le terrain.

Erreur 4 : Le mirage du ROI immédiat

L'IA ne crée pas de valeur en sortie d'usine. Elle en crée dans des **processus reconçus**. Coller un assistant IA sur un processus mal défini ne fait qu'amplifier le bruit. Le ROI vient quand le processus a été simplifié *avant* d'y mettre l'IA.

Antidote : chaque cas d'usage commence par une simplification de processus, pas par un ajout d'outil. Le *process audit* précède toujours la *tool selection*. Une heure de cadrage économise vingt heures de déploiement [8].

Erreur 5 : La cécité aux données

Sans carburant fiable, le moteur le plus sophistiqué ne démarre pas. Une IA branchée sur des données mal structurées, mal nettoyées ou mal accessibles produit des hallucinations, des contradictions et finalement de la défiance. C'est la principale cause d'abandon en phase pilote.

Antidote : auditer la qualité des données *avant* de choisir le cas d'usage. Si les données sont absentes ou trop bruitées, le cas d'usage ne se fait pas. On en trouve un autre. Une PME peut commencer à structurer ses 5 dernières années d'archives métier pendant qu'elle déploie ses premiers cas d'usage : c'est en partie ce qui fait la différence entre Orchestre et Architecte.

Chapitre 4 : Souveraineté & infrastructure

L'adoption de l'IA générative en PME pose trois questions de souveraineté qui étaient absentes (ou marginales) du débat il y a deux ans. En 2026, elles sont devenues structurantes pour le choix d'architecture.

4.1 Le CLOUD Act et ce qu'il signifie pour une PME

Le **Clarifying Lawful Overseas Use of Data Act** (CLOUD Act, États-Unis, 2018, §2713) [22] permet aux autorités américaines de réquisitionner des données détenues par un fournisseur de services états-unien, même si ces données sont physiquement stockées en Europe. La jurisprudence est désormais établie : un acteur de droit américain qui héberge des données européennes est tenu de coopérer avec une réquisition américaine, sauf accord bilatéral spécifique.

Implications concrètes pour une PME française :

- Les données envoyées à un fournisseur d'IA générative de droit américain (OpenAI, Anthropic via leur entité US, Google Cloud, Microsoft Azure si configuration non européenne stricte) peuvent être accessibles à la justice américaine sans notification au client français.
- Pour une PME qui manipule des données B2B confidentielles, des prix, des marges, de la propriété intellectuelle ou des données personnelles sensibles, cette accessibilité crée un risque juridique et commercial, particulièrement dans les secteurs santé, défense, énergie, finance, et plus largement pour toute entreprise exposée à des contrats imposant la souveraineté des données (administrations, OIV, OSE).

La règle pratique Junyr : classer les données *avant* de choisir le fournisseur. Les données dont l'exposition au CLOUD Act est inacceptable doivent rester chez un fournisseur de droit européen ou en on-premise. Les autres

peuvent légitimement transiter par un fournisseur états-unien, à condition que les conditions contractuelles le permettent.

4.2 On-premise vs cloud souverain : grille de décision

Trois options s’offrent à une PME, selon sa sensibilité et sa maturité :

Option	Pour qui	Coût indicatif (50-500 salariés)	Maturité requise
Cloud hyperscaler (US) : OpenAI, Anthropic, Microsoft Azure, Google Cloud	Données peu sensibles, sectoriellement non régulé	5 à 30 k€/an (licences + appels)	Artisan / Orchestre
Cloud souverain européen : OVHcloud, Scaleway, Outscale (modèles Mistral, Llama auto-hébergés ou en API souveraine)	Données sensibles, secteurs régulés, exigences commerciales de souveraineté	10 à 50 k€/an	Orchestre / Architecte
On-premise : GPU sur site, modèles Mistral ou Llama auto-hébergés	Données critiques, OIV/ OSE, IP stratégique	50 à 250 k€ d’investissement initial puis 30 à 80 k€/an	Architecte / Pionnier

Le ROI on-premise : selon les retours terrain Junyr [8], une PME industrielle qui passe 80 % de ses appels IA en on-premise (modèles Mistral ou Llama-class hébergés sur infrastructure interne) amortit son investissement initial en 18 à 30 mois si les volumes dépassent 5 millions de tokens / mois. En deçà, le cloud souverain européen reste plus rentable. La quantization (chap. 4.4) réduit la composante capex et peut rapprocher ce retour de la borne basse pour les usages étroits à fort volume, mais le facteur déterminant reste l’adoption, pas le matériel.

4.3 L'open-weight a rattrapé le frontier : ce que ça change pour la souveraineté

Pendant deux ans, l'argument « souveraineté » butait sur une objection technique : les modèles ouverts étaient nettement en retard sur les modèles propriétaires américains. En 2026, cette objection est tombée, et l'écart a continué de se réduire mois après mois.

- **L'écart de capacité s'est réduit à quelques points.** Il y a un an, le meilleur modèle à poids ouverts (DeepSeek V3 0324) plafonnait à **22** sur l'Artificial Analysis Intelligence Index, à 13 points du meilleur modèle propriétaire ; au 30 avril 2026, les meilleurs modèles ouverts (Kimi K2.6, MiMo V2.5 Pro) atteignaient **54 contre 60 pour GPT-5.5, un écart de 6 points** [33]. Et la dynamique ne faiblit pas : le 16 juin 2026, **GLM-5.2** (laboratoire Z.ai, licence MIT) est devenu le premier modèle open-weight de l'index et fait jeu égal avec GPT-5.5 sur le sous-benchmark agentique GDPval-AA [33].
- **Un laboratoire français reste dans la course frontier sous licence permissive.** Mistral Large 3 (2 décembre 2025) est un modèle Mixture-of-Experts de 675 milliards de paramètres (41 Md actifs), publié sous **licence Apache 2.0** [31] : une PME peut légalement le télécharger, l'affiner et l'exécuter sur son propre matériel, sans redevance et sans qu'aucune donnée ne quitte ses murs. Mistral a depuis livré **Mistral Medium 3.5** (30 avril 2026, 128 Md, open-weight), de gamme intermédiaire et sous licence MIT modifiée, mais utile pour des déploiements souverains à coût maîtrisé [31].
- **Le plancher matériel s'est effondré.** Google a publié en avril 2026 **Gemma 4** (licence Apache 2.0) : les poids non quantifiés tiennent sur **un seul GPU de 80 Go**, et les versions quantifiées tournent **hors ligne sur un téléphone, un Raspberry Pi ou une carte NVIDIA Jetson** [32]. Le débat « ouvert contre fermé » n'est plus idéologique.
- **Le prix d'usage s'effondre aussi.** En API hébergée, un modèle ouvert de premier plan comme DeepSeek V4-Flash s'affiche à **~0,14 \$ / 0,28 \$ par million de tokens** (entrée/sortie) [43]. *Caveat à garder en tête* : les modèles ouverts restent plus faibles en mémoire factuelle (test SimpleQA), d'où l'importance de l'ancrage (RAG) et de la vérification, qui sont des sujets de **méthode**, pas de modèle.

Lecture dirigeant. Les poids des modèles sont devenus une *commodité* : la capacité brute s'achète, se loue ou se télécharge. Dès lors, l'avantage durable d'une PME ne peut plus venir du modèle : il vient de la façon dont elle s'organise autour de lui (données propres, processus reconçus, gouvernance). La souveraineté, hier un compromis coûteux sur la performance, est aujourd'hui une **option réellement disponible**.

4.4 Le matériel ne suffit plus à justifier l'attentisme : la quantization

Deux objections historiques à l'on-premise, la perte de qualité et le coût du matériel, ont été largement levées en 2025-2026 par la **quantization** (réduction de la précision numérique des poids du modèle).

- **La qualité tient en basse précision.** Les mesures 2026 (vs pleine précision BF16) sont sans appel : en 4 bits, la dégradation tombe à **-1,2 à -2,5 % seulement** sur les grands benchmarks (MMLU, HumanEval, MATH), pour une qualité retenue d'environ **95 % en INT4 (AWQ/GPTQ), 97-98 % en INT8 et ~99 % en FP8** [34]. Un modèle 70B passe ainsi de ~140 Go à **36-40 Go** en INT4 [35]. La dégradation est sensible sur le raisonnement complexe, acceptable sur le résumé ou la classification : un arbitrage qui relève, encore, de la méthode.
- **Le coût d'hébergement a chuté.** Un modèle 70B quantifié en 4 bits tient sur ~48 Go de VRAM, soit une station de travail à carte unique de gamme professionnelle (ordre de grandeur 3 000-3 500 €, sources fournisseurs). L'argument « nous n'avons pas les moyens d'un data center » ne tient plus.

La difficulté se déplace donc du matériel vers l'**intégration, la sécurité et la conduite du changement**, ce qui confirme, plutôt qu'il ne contredit, la thèse de ce livre blanc : la transformation se gagne sur la méthode.

4.5 Où en est l'infrastructure souveraine européenne

Soyons lucides sur le rapport de force : l'Europe ne gagnera pas la course à la puissance de calcul brute. Les États-Unis concentrent **~74,5 %** de la performance des supercalculateurs IA, la Chine **~14,1 %**, et l'ensemble de l'Union européenne **~4,8 %** (Epoch AI, mai 2025, dernière mesure disponible) [36] ; en avril 2026, le *Sovereign AI Index* du CNAS résume la même concentration : **États-Unis et Chine contrôlent à eux deux ~90 % de la puissance de calcul nécessaire à l'IA de frontière** [36]. Côté marché, les hyperscalers américains contrôlent **plus de 85 % du marché cloud européen** [37]. C'est précisément pour cela que l'avantage d'une PME est organisationnel, et que le choix d'une brique européenne est un acte de gestion du risque, pas de militantisme.

Cet acte est aujourd'hui praticable, et le marché souverain monte en charge à grande vitesse : selon Gartner (février 2026), les dépenses mondiales de **cloud souverain (IaaS) atteindront 80 Md\$ en 2026 (+35,6 %)**, et **l'Europe bondit de +83 %** (de 6,9 à 12,6 Md\$), au point de dépasser l'Amérique du Nord dès 2027 [48].

- **Fournisseurs commerciaux disponibles maintenant.** OVHcloud AI Endpoints propose en API plus de 40 modèles à poids ouverts, hébergés en Europe, avec rétention de données nulle et réversibilité ; Scaleway opère des clusters NVIDIA H100 et des API génératives compatibles OpenAI, données stockées en Europe [38]. La compatibilité OpenAI rend le coût de migration faible.
- **Une trajectoire d'infrastructure financée.** Mistral vise 200 MW de calcul en 2027 et **1 GW en 2030** (plan de 4 Md€), avec un data center souverain à Bruyères-le-Châtel (44 MW, ~13 800 GPU NVIDIA GB300, mise en service mi-2026) adossé à une levée de dette de 830 M\$ menée par Bpifrance (mars 2026) [39]. Au niveau européen, l'initiative **InvestAI** mobilise ~200 Md€ vers 2030, dont **20 Md€** pour 4 à 5 « AI Gigafactories » ; en juin 2026, un consortium français, **AION** (Scaleway, EDF, Orange, iliad, Capgemini...), s'est constitué pour candidater à l'une d'elles [40][49].

Lecture dirigeant. La capacité souveraine *immédiatement* utilisable vient des fournisseurs commerciaux (OVHcloud, Scaleway, Mistral) ; les gigafactories sont un horizon 2027+. Une PME Orchestre peut donc démarrer dès aujourd'hui sur un cloud souverain européen, en gardant la réversibilité vers l'on-premise pour ses données les plus critiques.

4.6 Le calendrier PQC de l'ANSSI (2027-2035)

L'ANSSI a publié dès 2022 son avis sur la migration vers la **cryptographie post-quantique (PQC)**, mis à jour en 2024 [23]. Le calendrier est désormais explicite :

- **2027 (ANSSI) :** fin des certifications et qualifications de sécurité pour les produits dépourvus de cryptographie post-quantique, confirmé en juin 2026.
- **2030 (ANSSI) :** les administrations et les entreprises ne devraient plus acquérir que des produits intégrant la PQC.
- **2035 :** transition complète attendue, échéance fixée par la **feuille de route PQC coordonnée de l'UE** (juin 2025), que la position nationale de l'ANSSI anticipe [23].

Pour une PME, l'implication n'est pas immédiate mais elle est structurante : les choix d'architecture IA faits aujourd'hui (chiffrement des bases de connaissance, des canaux d'appel aux modèles, des journaux) doivent être compatibles avec la migration PQC. Concrètement, cela signifie privilégier des fournisseurs qui annoncent une trajectoire PQC documentée et qui exposent leur calendrier (OVHcloud, Scaleway, principales briques open-source). L'inaction se paie en 2027-2030 par une migration en urgence, toujours plus coûteuse.

4.7 NIS2 et Data Act : obligations concrètes

NIS2 (Directive UE 2022/2555) [24] étend les obligations de cybersécurité à un périmètre élargi d'entreprises. Pour une PME, l'enjeu pratique est de savoir si elle est :

- **Entité essentielle ou importante** au titre de NIS2 (secteurs énergie, santé, transports, finance, eaux, fournisseurs numériques essentiels...). Si oui, obligations renforcées : politique de gestion des risques, notification d'incident sous 24 h, formation des dirigeants.
- **Fournisseur de chaîne d'approvisionnement** d'une entité essentielle. Si oui, audit potentiel et obligations contractuelles renforcées même sans être directement régulée.

Le **Data Act** (Règlement UE 2023/2854) [25], dont les dispositions principales s'appliquent depuis le **12 septembre 2025**, réorganise les droits sur les données générées par les objets connectés et impose des obligations d'interopérabilité et de portabilité. Pour les PME industrielles ou de services connectés, le Data Act crée des opportunités (accès à de nouvelles données contractuellement) et des obligations (mise à disposition de leurs propres données aux clients).

4.8 Recommandation pratique : la matrice 2x2

Une grille de décision à deux dimensions (**sensibilité des données** × **budget infrastructure**) clarifie le choix :

	Budget infra contraint (< 30 k€/an)	Budget infra disponible (> 30 k€/an)
Données peu sensibles	Cloud hyperscaler avec contrôle d'accès stricte	Cloud souverain européen (recommandation Junyr, meilleur rapport sécurité/performance)
Données sensibles	Cloud souverain européen, exclusion des hyperscalers US sur la donnée sensible	On-premise pour les données critiques, cloud souverain pour le reste

La règle Junyr : **commencer simple, segmenter progressivement**. Une PME Orchestre peut démarrer en cloud souverain européen unique et migrer progressivement ses données critiques vers de l'on-premise quand le volume et la criticité le justifient.

Chapitre 5 : Maîtriser le coût de l'IA, FinOps & gouvernance

Le coût de l'IA générative n'est pas linéaire. Sans gouvernance, il s'emballe : multiplication des comptes, appels redondants, modèles surdimensionnés, absence de cache. Une PME au niveau Artisan peut consommer 2 à 10 k€ par mois en abonnements et appels API sans en mesurer la valeur. Une PME au niveau Architecte qui maîtrise son FinOps consomme 1,5 à 3 k€ par mois, et en tire mesurablement plus de valeur. La différence n'est pas dans la technologie ; elle est dans la gouvernance.

Le paradoxe FinOps de 2026. Le prix du token s'effondre : à performance constante, une étude révisée en mars 2026 (Gundlach et al., *The Price of Progress*) mesure une baisse de **~5 à 10× par an** pour les modèles de frontière (et jusqu'à **31× par an** sur le segment le plus performant), l'efficacité algorithmique seule contribuant pour ~3× [41]. En juin 2026, le tarif réellement payé par les entreprises tournait autour de **0,72 \$ par million de tokens en moyenne pondérée**, les modèles les plus légers descendant à **0,07 \$** [42]. Et pourtant, les factures IA explosent : entre janvier 2025 et avril 2026, **la consommation de tokens a bondi de +1 001 % quand la dépense progressait de +497 %** (Ramp, juin 2026) [42], parce que les workflows agentiques multiplient les appels (une tâche agentique consomme 5 à 30× plus de tokens qu'une requête de chatbot). Le symptôme est connu : début 2026, **Uber avait épuisé l'intégralité de son budget IA annuel dès le mois d'avril** [42]. **Des tokens moins chers ne donnent pas une facture plus basse : ils donnent une facture plus difficile à piloter.** Le levier n'est donc plus le prix unitaire : c'est la gouvernance.

5.1 Pourquoi les coûts IA explosent en PME

Trois mécanismes alimentent la spirale :

1. **Multiplification des comptes individuels.** Chaque collaborateur souscrit à son outil (ChatGPT Plus, Claude Pro, Mistral, etc.). Aucune visibilité consolidée. Aucun pouvoir de négociation.
2. **Appels redondants.** Sans cache, le même prompt est facturé à chaque exécution. Sans routing intelligent, les requêtes simples passent par les modèles les plus chers (« sur-utilisation de la Ferrari pour aller chercher le pain »).
3. **Absence de baseline et de seuils.** Personne ne sait combien l'entreprise dépense en IA au mois M ; donc personne ne sait si le mois M+1 est en dérive.

Le tout sur fond de **Shadow AI** : les usages personnels (chap. 1.2) cumulés représentent souvent davantage que le budget officiel.

5.2 LLM Gateway : le levier principal

Un **LLM Gateway** (proxy centralisé entre les utilisateurs/applications et les modèles d'IA) est le levier opérationnel n°1 de la maîtrise des coûts. Plusieurs options coexistent en mai 2026 :

- **Solutions open-source** : LiteLLM, Portkey (édition gratuite), Helicone, Langfuse, OpenLLMetry. Coût d'infrastructure marginal, déploiement en 1 à 3 jours.
- **Solutions SaaS commerciales** : Portkey (édition payante), TrueFoundry, Cloudflare AI Gateway, intégrations Mistral et OpenAI Enterprise.
- **Build interne** : pertinent pour les Architectes / Pionniers qui veulent une intégration forte avec leur stack.

Les bénéfices observés en mission Junyr [8] :

- **Réduction de coût de 30 à 60 %** sur la facturation IA brute, par effet combiné cache (10-25 %), routing intelligent vers les modèles adaptés (10-20 %), et rationalisation des appels (5-15 %).
- **Logging centralisé** : pour la première fois, l'entreprise sait *qui appelle quel modèle sur quel sujet et à quel coût*. C'est la base de tout pilotage.
- **Rate limiting** : protection contre les dérives d'usage (un script en boucle qui consomme 10 000 \$ par accident).
- **Contrôle d'accès et secret management** : les clés API ne sont plus dans les codes des collaborateurs ; elles sont centralisées et auditable.

Le bon réflexe 2026 : router vers des petits modèles. Le consensus 2026 défend une architecture hétérogène : on confie les tâches simples ou bien cadrées à un **petit modèle de langage** (SLM) et on n'escalade vers un grand modèle que les requêtes complexes. Pour les tâches répétitives à fort volume, ce routage **réduit le coût d'inférence jusqu'à 90 %** tout en offrant une latence quasi instantanée, au point que NVIDIA qualifie désormais les SLM d'« avenir de l'IA agentique » (*InfoWorld*, mai 2026) [44]. La discipline consiste à réserver le modèle frontier (cher) au raisonnement difficile, et à confier le reste à un SLM économique, souvent à poids ouverts et hébergeable en interne (chap. 4.3). C'est de l'orchestration, donc de la méthode.

Le déploiement d'un LLM gateway est l'un des premiers chantiers à mener dès le passage Orchestre → Architecte. En deçà, une PME Artisan peut commencer avec une solution open-source légère.

5.3 ISO 42001 : le cadre de gouvernance

La norme **ISO/IEC 42001:2023** [21] est le premier référentiel international d'**AI Management System** (AIMS). Elle est aux systèmes d'IA ce qu'ISO 27001 est à la sécurité de l'information : un cadre certifiable de gouvernance.

Pour une PME, **la certification ISO 42001 n'est pas un préalable** : c'est une cible à 18-36 mois pour les niveaux Architecte et Pionnier. En revanche, **les principes** sont applicables immédiatement et structurent la gouvernance dès Orchestre :

1. **Politique IA** documentée et signée par la direction.
2. **Registre des systèmes IA** en usage, avec classification du risque (en alignement AI Act [9][15]).
3. **Évaluation des risques** documentée par système (impact, vraisemblance, mesures).
4. **Revue de direction** semestrielle a minima.

Ces quatre éléments minimaux peuvent être mis en place en 6 à 12 semaines à coût marginal pour une PME au niveau Orchestre. Ils constituent un atout commercial **dès aujourd'hui** : grand donneur d'ordres et administration commencent à inscrire l'ISO 42001 dans leurs critères de sélection, environ 40 % des appels d'offres IA en mid-2026 selon les remontées terrain [8].

5.4 TCO et ROI : la bonne équation

Le **TCO (Total Cost of Ownership)** d'une démarche IA en PME se décompose en six postes :

Poste	Part typique du TCO an 1	Évolution an 2-3
Accompagnement (manager de transition, conseil)	35-50 %	Décroît fortement
Licences et appels API	10-25 %	Stable ou croissant
Infrastructure (cloud, GPU)	5-15 %	Croît avec la maturité
Intégration et développement	10-20 %	Décroît après mise en place
Compétences (formation, recrutement)	10-20 %	Stable, dimension structurante
Gouvernance et conformité	3-10 %	Croît avec la maturité

Le **ROI** se décompose en trois leviers :

1. **Gains de productivité mesurés** (heures économisées × coût horaire chargé).
2. **Revenu additionnel attribuable** (chiffre d'affaires généré par capacité libérée, taux de conversion accru).
3. **Risques évités quantifiables** (non-conformité, fuite de données, perte client).

La règle Junyr : **mesurer mensuellement les trois leviers, agréger trimestriellement, ne jamais sortir un chiffre de ROI sans la méthodologie qui le sous-tend**. Le ROI médian de 159,8 % [7] est crédible parce qu'il est mesuré, sur 24 mois et plus de 200 déploiements ; un ROI annoncé sans méthodologie est un chiffre marketing.

5.5 La gouvernance budgétaire mensuelle

Quatre rituels structurent le FinOps de l'IA en PME au niveau Orchestre ou supérieur :

1. **Dashboard mensuel** consolidé : coût total, coût par cas d'usage, coût par utilisateur actif, ratio coût/valeur générée.
2. **Alertes seuil** : déclenchement automatique au-delà d'un budget mensuel défini par cas d'usage.
3. **Revue trimestrielle** en comité IA : optimisation modèles, rationalisation des comptes, négociation fournisseurs.
4. **Bilan annuel** intégré au budget de l'année suivante, avec roadmap d'évolution.

Ces rituels coûtent peu (2 à 5 heures par mois pour le référent IA) et préviennent des dérives qui peuvent atteindre 100 à 200 k€/an dans une PME mid-market.

À retenir. Le report des obligations AI Act post-Digital Omnibus (chap. 1.4) ne signifie pas report de la gouvernance interne, au contraire. C'est une fenêtre de 18 à 24 mois pour installer les pratiques FinOps et ISO 42001 avant que la compliance ne devienne contraignante. Les PME qui utilisent cette fenêtre prendront un avantage de gouvernance que les retardataires ne pourront rattraper qu'en urgence.

Chapitre 6 : La méthodologie en 5 phases

La démarche Méthode Junyr™ enchaîne cinq phases sur 12 mois. Elle est conçue pour une PME-ETI de 50 à 500 salariés, accompagnée par un manager de transition externe sur la durée. Le budget total (accompagnement + licences + outils) se situe entre 30 et 80 k€ pour un passage Artisan → Orchestre [8].

Phase 1 : Diagnostic 360° (semaines 1 à 3)

Cartographie des processus métier critiques, audit de la qualité et de l'accessibilité des données, posture du dirigeant et du COMEX face au sujet IA, repérage des résistances probables et des champions potentiels. Le livrable est un rapport de 15 à 25 pages, restitué en COMEX, qui se conclut par 3 à 5 cas d'usage prioritaires classés par ratio impact / effort.

La grille d'analyse repose sur les 12 points de l'Échelle Méthode Junyr™ (chap. 2.9) : maturité processus, maturité data, maturité outils, posture managériale, gouvernance, formation, sécurité, conformité, mesure, change, infrastructure, budget. Chaque point est noté de 0 à 5.

Phase 2 : Cadrage des cas d'usage (semaines 4 à 6)

Pour chaque cas retenu en phase 1, un document de cadrage de 2 à 4 pages : périmètre précis, utilisateurs cibles nommés, données mobilisées, outils candidats avec critères de sélection, budget chiffré, ROI attendu chiffré, risques identifiés, plan de bascule en production, indicateurs de succès définis.

Une heure de cadrage économise vingt heures de déploiement. C'est l'enseignement le plus régulier de la pratique terrain [8]. Les cas d'usage qui sautent cette étape se rejouent toujours en phase 4, sous une forme dégradée et avec un retard de planning.

Phase 3 : Préparation des fondations (semaines 7 à 12)

Parallèlement aux deux premiers cas d'usage qui démarrent, l'entreprise construit ses fondations. Charte d'usage IA validée en COMEX. Politique de données documentée (classification, accès, conservation). Plan de formation à trois niveaux : COMEX, managers, opérationnels. Architecture cible définie (cloud, hébergement, briques d'orchestration, LLM gateway). Structuration des données prioritaires. Premiers éléments ISO 42001 (politique, registre, évaluation des risques).

À la sortie de cette phase, l'entreprise est devenue capable de réussir ses déploiements. C'est l'investissement invisible qui fait la différence entre les ~11 % d'« AI leaders » qui récoltent et la majorité qui saupoudre [2][29].

Phase 4 : Déploiements pilotes (semaines 13 à 26)

Sprints de 4 à 6 semaines, un cas d'usage par sprint. Pour chaque sprint :

- Un utilisateur final identifié dès le démarrage et engagé sur le résultat.
- Des indicateurs de succès définis avant le lancement.
- Une revue à mi-parcours et une revue de bascule en production.
- Une revue mensuelle COMEX qui agrège les sprints.

La règle est stricte : pas de bascule en production sans atteinte des indicateurs de succès. Si un cas d'usage ne tient pas ses promesses, il est arrêté, et un autre prend sa place. Cette discipline coûte de l'orgueil ; elle gagne du ROI.

Phase 5 : Consolidation et gouvernance (mois 7 à 12)

Industrialisation des cas d'usage qui ont basculé en production : optimisation, intégration dans les outils maison, plan de montée en compétences des utilisateurs, mesure continue des bénéfices. Mise en place d'une gouvernance pérenne : registre AI Act tenu à jour, comité IA trimestriel piloté par un membre du COMEX, dashboard FinOps mensuel, roadmap année 2 documentée.

À la sortie des 12 mois, l'entreprise est solidement installée au niveau Orchestre. La trajectoire vers Architecte est tracée et budgétée pour les 12 à 24 mois suivants.

Commander l'IA : l'évolution du prompt vers le context engineering

Une compétence transverse aux cinq phases a émergé en 2025-2026 et structure désormais la pratique : le **context engineering**, ou « commander l'IA ». Là où 2023 demandait d'apprendre à écrire des prompts, 2026 demande de **construire le contexte** dans lequel l'IA opère : choix du modèle, structure de la requête, intégration de connaissances métier, contrôles de qualité, garde-fous, mesure de la sortie.

Pour une PME au niveau Orchestre ou supérieur, le context engineering n'est pas une compétence individuelle. C'est une **discipline d'équipe** :

- **Pour les dirigeants** : « literacy IA », comprendre les principes pour piloter, sans devoir prompter eux-mêmes.
- **Pour les managers** : capacité à cadrer un cas d'usage en termes de contexte (données mobilisées, garde-fous, indicateurs).
- **Pour les opérationnels et référents** : maîtrise des outils, des prompts structurés, des cas limites.

Cette transition, du *prompt* au *context*, est aussi structurante que celle qui a fait passer de la dactylographie à la bureautique dans les années 1990. Les entreprises qui s'y forment maintenant prendront un avantage durable. Voir l'article du portfolio « La fin du prompt engineering » pour le traitement approfondi (28 sources) [26].

Chapitre 7 : Étude de cas INDUSTEC

INDUSTEC (nom anonymisé) est une PME industrielle française de 78 collaborateurs, 23,5 M€ de chiffre d'affaires [10], spécialisée dans la conception et la fabrication d'équipements techniques pour les industries de transformation. Mission démarrée en avril 2025, mesure à 9 mois (janvier 2026) [10].

Point de départ : niveau Artisan

Au diagnostic d'avril 2025, l'IA était présente dans l'entreprise sous trois formes, toutes individuelles : un commercial qui utilisait ChatGPT pour pré-rédiger des relances, la responsable communication qui s'en servait pour les posts LinkedIn, et le directeur technique qui testait Mistral sur des questions ponctuelles. Aucun outil partagé. Aucun processus impacté. Aucune mesure de bénéfice [10]. Profil typique du niveau Artisan (chap. 2.4).

Cas d'usage prioritaire : la rédaction des devis techniques

Le diagnostic a identifié la rédaction des devis techniques comme cible prioritaire. Volumétrie : 600 à 800 devis par an, chacun mobilisant 2 à 4 heures d'un ingénieur d'application pour la partie rédactionnelle (description technique, références produits, conditions de garantie, termes commerciaux). Soit environ 1 800 à 2 400 heures annuelles [10].

Cadrage en phase 2 : un assistant interne s'appuyant sur un référentiel produits structuré, des templates de devis maison, et l'historique des 1 800 devis émis depuis 2022. Indicateur de succès : **réduction d'au moins 50 % du temps de rédaction par devis** [10].

Résultats à 9 mois

Mesure janvier 2026, sur les 6 derniers mois d'exploitation en production [10] :

- **275 heures économisées par mois** [10] sur la rédaction des devis. Soit l'équivalent de 1,7 ETP libéré, redéployé sur le suivi commercial post-vente et les visites client en amont.
- **+18 % de chiffre d'affaires sur le périmètre commercial concerné** [10]. La capacité supplémentaire dégagée a permis de répondre à plus d'appels d'offres et de faire plus de visites client en amont. La conversion devis → commande est passée de 38 % à 44 % [10].
- **ROI de 182 % documenté à 9 mois de mission, consolidé sur 12 mois** [10]. Coût total de la mission (32 k€ d'accompagnement + 8 k€ de licences + 4 k€ de structuration interne) rentabilisé au 5^e mois.

Bascule confirmée vers Orchestre stabilisé

Au-delà des chiffres, l'enseignement est ailleurs : INDUSTEC est passé d'une situation où l'IA était un sujet de curiosité personnelle (Artisan) à une situation où l'IA est inscrite dans le processus commercial, mesurée mensuellement en COMEX, et étendable (Orchestre). Deux cas d'usage suivants ont été cadrés pour 2026 : la veille concurrentielle, et l'aide à la rédaction des comptes rendus de visite client.

La trajectoire vers Architecte est tracée pour 2027-2028, avec la construction d'une base de connaissance technique propriétaire (10 ans d'archives métier nettoyées) et l'automatisation supervisée de la réponse aux appels d'offres.

Étude de cas anonymisée. Toute ressemblance avec une entreprise existante serait fortuite. Les résultats présentés sont issus d'une mission spécifique et ne constituent pas une garantie de résultats équivalents pour d'autres organisations. Le retour sur investissement d'une transformation IA dépend du contexte sectoriel, de la maturité initiale de l'entreprise, de l'engagement des équipes et de la qualité d'exécution. Les performances passées ne préjugent pas des résultats futurs.

Chapitre 8 : Feuille de route 12 mois

La feuille de route ci-dessous est le squelette type d'une mission Artisan → Orchestre. Les libellés des cas d'usage sont à remplacer par les cas spécifiques retenus en phase 1.

Trimestre 1 : Diagnostic et fondations

- Diagnostic 360° et cadrage des 3 à 5 cas d'usage prioritaires.
- Charte d'usage IA validée en COMEX.
- Désignation des 2 à 4 référents internes.
- Plan de formation à trois niveaux validé.
- Premiers éléments ISO 42001 (politique, registre minimal).

Indicateur de bonne fin : rapport diagnostic restitué, charte signée, référents nommés et briefés, registre IA initial constitué.

Trimestre 2 : Premier déploiement pilote

- Sprint 1 sur le cas d'usage prioritaire (4 à 6 semaines).
- Sprint 2 sur le second cas d'usage.
- Formation des opérationnels sur les outils retenus.
- Mise en place du LLM gateway et du dashboard FinOps initial.
- Première revue mensuelle COMEX avec mesure d'adoption.

Indicateur de bonne fin : au moins un cas d'usage en production avec ses indicateurs de succès atteints ; LLM gateway opérationnel ; dashboard FinOps en place.

Trimestre 3 : Industrialisation et extension

- Sprint 3 et sprint 4 sur les cas d'usage suivants.
- Industrialisation du cas d'usage trimestre 2 (intégration outils, support).
- Mise en place du registre AI Act au format complet.
- Premiers gains documentés en COMEX, premières optimisations FinOps.

Indicateur de bonne fin : 3 cas d'usage en production, gains mesurés sur 2 mois consécutifs, optimisations FinOps mesurables (-15 à -25 % vs. trajectoire initiale).

Trimestre 4 : Consolidation et roadmap année 2

- Optimisation des cas en production.
- Mesure consolidée des bénéfices sur 12 mois.
- Comité IA trimestriel constitué et premier comité tenu.
- Roadmap année 2 validée en COMEX, incluant trajectoire ISO 42001 et plan Architecte.

Indicateur de bonne fin : bilan de mission documenté, ROI mesuré, gouvernance pérenne installée, trajectoire Architecte cadrée pour année 2.

Chapitre 9 : Conclusion & passer à l'action

La fenêtre est encore ouverte

La transformation par l'IA n'est pas un sujet technologique. C'est un sujet de discipline d'exécution. Les outils sont disponibles, les méthodes sont documentées. Ce qui manque dans la plupart des PME, ce ne sont pas les moyens : c'est le cadre qui permet de transformer ces moyens en résultats.

L'Échelle Méthode Junyr™ en 5 niveaux et la démarche en 5 phases présentées dans ce livre blanc sont une réponse à ce manque. Elles ne sont pas magiques : elles ne dispensent ni du travail de cadrage, ni de l'engagement du COMEX, ni de l'investissement en formation. Mais elles tracent un chemin praticable, mesuré, et reproductible.

Le calendrier post-Digital Omnibus (chap. 1.4) offre une fenêtre (18 à 24 mois) pour installer une gouvernance proprement, avant que les obligations haut risque ne deviennent contraignantes en fin 2027 / mi-2028. Les PME qui utiliseront cette fenêtre pour passer Artisan → Orchestre, puis Orchestre → Architecte, prendront un avantage de gouvernance et de discipline que les retardataires ne pourront rattraper qu'en urgence, toujours plus cher, toujours plus risqué.

Où va la frontière ? Ce que disent ceux qui la construisent

Un dirigeant a le droit de se demander si tout cela ne sera pas balayé par la prochaine vague. Voici ce que **déclarent**, sans qu'on puisse en garantir le calendrier, ceux qui construisent la frontière, en juin 2026 :

- **Demis Hassabis** (Google DeepMind, mai 2026) a resserré sa fenêtre pour une IA de niveau humain (« AGI ») à **2029-2030**, contre 2030-2035 un an plus tôt [45].
- **Dario Amodei** (Anthropic, janvier 2026) estime qu'une « IA puissante » pourrait n'être qu'à **un ou deux ans**, « même si cela pourrait aussi être nettement plus loin », précise-t-il lui-même [46].
- Côté mesures, le **Stanford AI Index 2026** documente une accélération réelle : la performance sur le benchmark de codage SWE-bench est passée de ~60 % à près de 100 % du niveau humain en un an [29].

Mais la même édition de l'AI Index pose le **contrepois honnête** : ces modèles qui décrochent l'or à l'Olympiade internationale de mathématiques ne lisent correctement une **horloge analogique que 50,1 % du temps** [29]. La capacité est « en dents de scie » : surhumaine sur une tâche, défailante sur la voisine.

Quelle conclusion pour un dirigeant ? La date de l'AGI est contestée ; sa trajectoire ne l'est pas : la capacité brute s'accélère et se banalise. Or, plus la technologie devient une commodité que tout le monde peut acheter, **plus le seul avantage durable est organisationnel**. BCG le chiffre : une entreprise dotée d'une **stratégie IA claire** capte **+25 points d'impact**, contre **+5 points** seulement pour celle qui se contente d'acheter de meilleurs outils [47]. La bonne réponse à l'incertitude n'est pas d'attendre. C'est d'être prêt : monter en maturité maintenant, méthodiquement.

Et après ? Du diagnostic à l'industrialisation

Ce livre blanc répond à la première question : *où en êtes-vous ?* La suivante arrive vite : *comment passer des premiers pilotes à l'usage quotidien, mesuré, gouverné ?* C'est l'objet du second livre blanc de la collection, « **Du POC à l'industrialisation : conduire le changement IA dans les PME européennes** » (édition juin 2026, FR + EN). Il documente les sept frictions du « dernier kilomètre », les cinq leviers de conduite du changement et

l'AgentOps, car seuls 46 % des POC IA atteignent la production [27], et ce mur est à 70 % humain, pas technologique [28].

→ Lire le livre blanc n°2 : paulantoinetual.fr/livre-blanc-industrialisation

Diagnostic IA Express : le premier pas

Pour positionner votre entreprise sur l'Échelle Méthode Junyr™ et tracer une feuille de route à 90 jours, le **Diagnostic IA Express** est conçu comme un premier pas calibré :

- 60 minutes en visioconférence, sans engagement.
- Positionnement de votre PME sur les 5 niveaux.
- Identification de 3 cas d'usage prioritaires pour votre contexte.
- Première feuille de route à 90 jours, livrée par écrit dans la semaine.

C'est ce premier pas, court mais engageant, qui fait souvent la différence entre les projets qui démarrent et ceux qui restent au stade de l'intention. La démarche est conforme au cadre **Osez l'IA** (Bpifrance) [14] et peut s'inscrire dans une trajectoire de financement public.

→ Réserver un Diagnostic IA Express : junyr.fr/diagnostic

SOURCES : vérifiables, juin 2026

[1] **Bpifrance Le Lab et Rexecode**, baromètres de conjoncture des TPE-PME. 82^e baromètre semestriel (janvier 2026) : 55 % d'usage IA déclaré fin 2025 ; baromètre du 2^e trimestre 2026 (terrain 13-26 avril 2026, 1 135 répondants, publié le 19 mai 2026) : **58 % d'usage déclaré, dont près de la moitié quotidiennement** ; 43 % des dirigeants constatent un gain de productivité ; 1^{re} barrière = difficulté à identifier les cas d'usage pertinents (~54 %). <https://presse.bpifrance.fr> : usages dominants, croissance sur 18 mois.

[2] **KPMG**, *Global AI Pulse, 1^{er} trimestre 2026* (enquête du 17 février au 17 mars 2026, 2 110 dirigeants, 20 pays), 31 mars 2026. <https://kpmg.com/xx/en/media/press-releases/2026/03/kpmg-global-ai-pulse-survey.html> : **11 %** d'organisations « AI leaders » ; **82 %** des leaders rapportent une valeur métier significative contre **62 %** des autres ; **32 %** déploient ou passent à l'échelle des agents IA. Source de la part « ≈11 % » d'organisations tirant une valeur mesurée de l'IA, utilisée dans ce livre blanc.

[3] **Bpifrance Le Lab**, *L'IA dans les PME et ETI françaises* (enquête terrain octobre-décembre 2024, 1 209 dirigeants) : perception « enjeu de survie », absence de stratégie IA formalisée, non-analyse des données. Complété par **INSEE Première n°2061**, *Les TIC dans les entreprises en 2024* (juillet 2025) : 10 % des entreprises de 10 salariés ou plus utilisent l'IA en 2024. <https://www.insee.fr/fr/statistiques/8604126>

[4] **Gartner**, *Hype Cycle for Agentic AI* (avril 2026) : au moins 50 % des projets d'IA générative abandonnés après la phase pilote ; **plus de 40 % des projets d'IA agentic abandonnés d'ici 2027** (coûts, valeur floue, contrôle des risques). *AI Maturity Model Toolkit* (modèle 5 niveaux), 2025-2026. <https://www.gartner.com/en/articles/hype-cycle-for-agentic-ai>

[5] **HCLTech**, *AI Impact Imperatives 2026* (20 mai 2026, enquête auprès de 467 dirigeants d'entreprises de plus d'un milliard de dollars de chiffre d'affaires) : **43 % des grandes initiatives IA sont jugées vouées à l'échec**, principalement à cause d'un « execution gap » organisationnel plutôt que technologique. <https://www.hcltech.com/press-releases/hcltech-report-warns-43-enterprise-ai-initiatives-may-fail-leaders-face-shrinking>

[6] **McKinsey**, *The State of AI Trust in 2026* (25 mars 2026, terrain déc. 2025 à janv. 2026, ~500 organisations) : modèle de maturité de la confiance IA à 5 dimensions ; score moyen **2,3 sur 4, une organisation sur trois seulement au niveau 3 ou plus** ; la supervision de l'IA par la direction figure parmi les facteurs les plus corrélés à l'impact au compte de résultat. <https://www.mckinsey.com/capabilities/quantumblack/our-insights>

[7] **Denis Atlan**, *Baromètre IA & ROI PME France 2022-2025*, 2026. Analyse de plus de 200 déploiements IA B2B en PME françaises : ROI médian de 159,8 % mesuré sur 24 mois pour les projets bien cadrés, taux de succès de 73 %. <https://www.denisatlan.fr/barometre-ia-pme> (DOI : 10.5281/zenodo.17795133).

[8] **Croissance et Transitions**, retours terrain consolidés : plus de 30 missions IA documentées en PME-ETI françaises depuis 2024 ; budgets observés, ratios cadrage/déploiement, causes d'échec récurrentes, durées de répliation concurrentielle.

[9] **Règlement (UE) 2024/1689** du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle (*Artificial Intelligence Act*). <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> : entrée en vigueur 1^{er} août 2024, application progressive 2025-2027, calendrier actualisé par le Digital Omnibus [15].

[10] **Étude de cas INDUSTEC** (nom anonymisé) : mission Méthode Junyr™ avril 2025 à janvier 2026, mesures consolidées sur les 6 derniers mois d'exploitation en production (juillet 2025 à décembre 2025). Données issues du reporting interne client, validées en COMEX mensuel.

[11] **ADP Research**, *People at Work 2026* (près de 40 000 actifs, 36 pays ; volet France n = 1 051, diffusion juin 2026) : en France, **11 %** des salariés utilisent l'IA presque quotidiennement, 24 % plusieurs fois par semaine, **29 % jamais**, contre 20 % d'usage quasi quotidien au niveau mondial. <https://www.adpresearch.com/research/ai-powers-into-the-workplace>

[12] **DGE / France Num**, *Baromètre France Num 2025*, 6e édition, septembre 2025, 11 021 entreprises. <https://www.francenum.gouv.fr/guides-et-conseils/strategie-numerique/comprendre-le-numerique/barometre-france-num-2025-le> : 26 % des TPE-PME utilisent au moins un outil d'IA en 2025 (×2 vs 2024) ; par taille : 23 % (1-4), 31 % (10-19), 35 % (20-49), 42 % (50-249).

[13] **Microsoft**, *Work Trend Index 2026: Annual Report*, 5 mai 2026. <https://www.microsoft.com/en-us/worklab/work-trend-index> : « Frontier Firms » (19 % des utilisateurs en organisation Frontier, 16 % « Frontier Professionals » orchestrant des agents), « agent boss » / human agency, écart dirigeants/employés sur les agents (67 % vs 40 %), Owned Intelligence.

[14] **Bpifrance**, *Plan Osez l'IA*, juillet 2025 ; *Diag Data IA*. <https://www.bpifrance.fr/catalogue-offres/diag-data-ia> ; <https://bpifrance-creation.fr/entrepreneur/actualites/lancement-du-plan-national-osez-lia> : enveloppe initiale 200 M€, dynamique globale 10 Md€, cible 80 % PME-ETI au niveau Orchestre+ d'ici 2030. Depuis le 1^{er} janvier 2026, le Diag Data IA est un diagnostic forfaitaire de 10 000 € HT pris en charge à 25 % (ETI non éligibles) ; l'IA Booster finance jusqu'à 80 %.

- [15] **Conseil de l'Union européenne**, *Digital Omnibus : accord politique sur la simplification des règles numériques*, communiqué du 7 mai 2026. <https://www.consilium.europa.eu/en/press/press-releases/2026/05/07/> : décalage AI Act haut risque à fin 2027 / 2 août 2028, réduction -35 % de la charge administrative PME d'ici 2029.
- [16] **Gartner**, *AI Maturity Model: 5 levels* (Awareness, Active, Operational, Systemic, Transformational), 2025-2026. <https://www.gartner.com/en/chief-information-officer/research/ai-maturity-model-toolkit>
- [17] **PwC**, *The AI-Native Enterprise: 5-level maturity model*, 2026. <https://www.pwc.com.au/services/artificial-intelligence/the-ai-native-enterprise.html>
- [18] **Microsoft**, *Agentic AI Adoption Maturity Model (Copilot Studio): L100 to L500*, 2025-2026. <https://learn.microsoft.com/en-us/microsoft-copilot-studio/guidance/maturity-model-overview>
- [19] **BCG**, *AI Radar 2026: As AI Investments Surge, CEOs Take the Lead* (15 janvier 2026, 2 360 dirigeants dont 640 CEO, 16 marchés). <https://www.bcg.com/publications/2026/as-ai-investments-surge-ceos-take-the-lead> : archétypes de dirigeants **Trailblazers ~15 % / Pragmatists ~70 % / Followers ~15 %** ; 72 % des CEO se déclarent principal décideur IA (×2 en un an) ; > 30 % du budget IA alloué à l'agentique.
- [20] **Deloitte**, *State of AI in the Enterprise 2026: The Untapped Edge* (21 janvier 2026, 3 235 dirigeants, 24 pays). <https://www.deloitte.com/us/en/about/press-room/state-of-ai-report-2026.html> : **30 % seulement** des organisations reconçoivent leurs processus clés autour de l'IA, **37 %** l'emploient en surface, ~25 % constatent un effet transformateur ; 23 % utilisent l'IA agentique au moins modérément, ~74 % prévoient de le faire sous deux ans, **21 % seulement** ont une gouvernance mature des agents.
- [21] **ISO/IEC 42001:2023**, *Artificial Intelligence Management Systems: Requirements*, ISO, décembre 2023. <https://www.iso.org/standard/81230.html> : premier référentiel international AIMS certifiable.
- [22] **CLOUD Act** (États-Unis, 2018), §2713 : *Clarifying Lawful Overseas Use of Data Act*. <https://www.congress.gov/bill/115th-congress/house-bill/4943>
- [23] **ANSSI**, *Avis sur la migration vers la cryptographie post-quantique*, première publication 2022, mises à jour 2024 puis 2026 (fin des certifications sans PQC à 2027). <https://cyber.gouv.fr/> : calendrier national 2027 / 2030. L'échéance 2035 relève de la **feuille de route PQC coordonnée de l'UE** (NIS Cooperation Group, juin 2025).
- [24] **Directive (UE) 2022/2555 (NIS2)** du Parlement européen et du Conseil du 14 décembre 2022 concernant des mesures pour un niveau commun élevé de cybersécurité dans l'Union. <https://eur-lex.europa.eu/eli/dir/2022/2555/oj>
- [25] **Règlement (UE) 2023/2854 (Data Act)** du 13 décembre 2023 concernant des règles harmonisées portant sur l'équité de l'accès aux données et de leur utilisation. <https://eur-lex.europa.eu/eli/reg/2023/2854/oj> : dispositions principales applicables depuis le **12 septembre 2025**.
- [26] **Croissance et Transitions / Méthode Junyr™**, article *La fin du prompt engineering : du prompt au context engineering*, mai 2026 (28 sources). <https://paulantoinetual.fr/blog/la-fin-du-prompt-engineering-pourquoi>
- [27] **IDC / Lenovo**, *CIO Playbook 2026: The Race for Enterprise AI* (janvier 2026, n = 800, Europe & Moyen-Orient) : **seuls 46 % des POC atteignent la production** ; 94 % des organisations anticipent un ROI positif, avec **2,78 \$ de valeur attendue par dollar investi**. <https://news.lenovo.com/pressroom/press-releases/research-reveals-ai-is-paying-off-cios-arent-ready/>

- [28] **BCG**, *AI Transformation Is a Workforce Transformation* (2026). Réaffirmation de la règle **10-20-70** : 10 % algorithmes, 20 % technologie et données, **70 % refonte du facteur humain et des processus**. <https://www.bcg.com/publications/2026/ai-transformation-is-a-workforce-transformation>
- [29] **Stanford HAI**, *The 2026 AI Index Report*, 20 avril 2026 : 88 % des organisations utilisent l'IA dans ≥ 1 fonction (vs 78 % l'an passé), < 10 % à l'échelle ; SWE-bench ~ 60 % $\rightarrow \sim 100$ % du niveau humain en un an ; capacité « en dents de scie » (or à l'OIM mais horloges analogiques lues à 50,1 %). <https://hai.stanford.edu/ai-index/2026-ai-index-report>
- [30] **OCDE**, *AI adoption by small and medium-sized enterprises* (décembre 2025) : adoption IA dans l'OCDE de 5,6 % (2020) à 14 % (2024) ; 29 % seulement des PME utilisatrices d'IA générative l'emploient dans leur cœur de métier. Voir aussi *Compendium of Productivity Indicators 2025* (écart de productivité PME / grandes entreprises). <https://www.oecd.org/>
- [31] **Mistral AI**, *Mistral Large 3* (2 décembre 2025, MoE 675B/41B, licence Apache 2.0) : modèle open-weight de frontière le plus récent du laboratoire français. <https://mistral.ai/news/mistral-3/> ; depuis : *Mistral Medium 3.5* (30 avril 2026, 128B, open-weight, licence MIT modifiée, gamme intermédiaire).
- [32] **Google DeepMind**, *Gemma 4* (2 avril 2026, licence Apache 2.0) : poids non quantifiés tenant sur un seul GPU de 80 Go ; versions quantifiées exécutables hors ligne sur téléphone, Raspberry Pi ou NVIDIA Jetson. <https://blog.google/innovation-and-ai/technology/developers-tools/gemma-4/>
- [33] **Artificial Analysis**, *Intelligence Index* : au 30 avril 2026, meilleurs modèles ouverts (Kimi K2.6, MiMo V2.5 Pro) à **54** contre **60** pour GPT-5.5 (écart 6 points), contre ~ 22 un an plus tôt ; le 16 juin 2026, **GLM-5.2** (Z.ai, MIT) devient le 1^{er} modèle open-weight de l'index et fait jeu égal avec GPT-5.5 sur le sous-benchmark agentique GDPval-AA. Rappel factuel (SimpleQA) encore en retrait pour les modèles ouverts. <https://artificialanalysis.ai/articles/glm-5-2-is-the-new-leading-open-weights-model-on-the-artificial-analysis-intelligence-index>
- [34] **Sesame Disk**, *Quantization Techniques for AI Inference in 2026* (14 mai 2026). Qualité retenue vs BF16 : ~ 99 % (FP8), $\sim 97-98$ % (INT8), ~ 95 % (INT4 AWQ/GPTQ) ; dégradation 4 bits mesurée à $-1,2$ à $-2,5$ % (MMLU/HumanEval/MATH). <https://sesamedisk.com/quantization-techniques-ai-inference-2026/>
- [35] **VRLA Tech**, *LLM Quantization Explained: INT4, INT8, FP8, AWQ and GPTQ in 2026* (avril 2026) : échelle qualité/mémoire par format (ordres de grandeur). <https://vrlatech.com/llm-quantization-explained-int4-int8-fp8-awq-and-gptq-in-2026/>
- [36] **Epoch AI**, *AI supercomputers performance share by country* (mai 2025, dernière mesure disponible) : US $\sim 74,5$ % / Chine $\sim 14,1$ % / UE $\sim 4,8$ %. <https://epoch.ai/data-insights/ai-supercomputers-performance-share-by-country> ; concentration confirmée par le **CNAS**, *Sovereign AI Index* (avril 2026) : États-Unis et Chine ~ 90 % de la puissance de calcul de frontière. <https://interactives.cnas.org/reports/sovereign-ai-index/>
- [37] **EU Perspectives**, *Europe bets on EuroStack* (février 2026) : hyperscalers US > 85 % du marché cloud UE ; EuroStack ~ 300 Md€ ; cloud souverain UE $\times 3 \rightarrow 23$ Md\$ d'ici 2027 (Gartner). <https://euperspectives.eu/2026/02/europe-bets-on-eurostack/>
- [38] **OVHcloud**, *AI Endpoints* (40+ modèles ouverts, rétention nulle, hébergement UE) <https://www.ovhcloud.com/fr/public-cloud/ai-endpoints/> ; **Scaleway**, *Generative APIs* (clusters H100, compatibles OpenAI, données UE) <https://www.scaleway.com/fr/generative-apis/>

- [39] **Maddyness**, *Mistral AI vise une capacité de calcul d'un gigawatt à l'horizon 2030* (28 mai 2026) : 200 MW en 2027, 1 GW en 2030, 4 Md€ ; data center souverain Bruyères-le-Châtel (levée de dette 830 M\$ menée par Bpifrance, mars 2026). <https://www.maddyness.com/2026/05/28/mistral-ai-vise-une-capacite-de-calcul-dun-gigawatt-a-lhorizon-2030/>
- [40] **Commission européenne**, *InvestAI / AI Gigafactories* : ~200 Md€ vers 2030 dont 20 Md€ pour 4-5 gigafactories. <https://digital-strategy.ec.europa.eu/>
- [41] **Gundlach, Lynch, Mertens & Thompson**, *The Price of Progress: Price Performance and the Future of AI* (arXiv:2511.23455, version révisée du 23 mars 2026) : coût d'inférence à performance constante divisé par **~5 à 10× par an** (jusqu'à 31×/an sur le segment le plus performant) ; efficacité algorithmique ~3×/an. <https://arxiv.org/abs/2511.23455>
- [42] **Ramp**, *The cost of AI tokens for businesses* (8 juin 2026) : entre janvier 2025 et avril 2026, consommation de tokens **+1 001 %** contre dépense **+497 %** ; tarif observé moyen **0,72 \$/M tokens** (jusqu'à 0,07 \$ pour les modèles légers). <https://ramp.com/blog/ai-token-cost-for-businesses> ; voir aussi **TechCrunch**, *The token bill comes due* (5 juin 2026) : Uber a épuisé son budget IA 2026 dès avril ; une tâche agentique consomme 5-30× plus de tokens (Gartner). <https://techcrunch.com/2026/06/05/the-token-bill-comes-due-inside-the-industry-scramble-to-manage-ai-runaway-costs/>
- [43] **DeepSeek**, grille tarifaire API (2026) : DeepSeek V4-Flash à **~0,14 \$ / 0,28 \$ par million de tokens** (entrée/sortie), licence MIT. https://api-docs.deepseek.com/quick_start/pricing
- [44] **InfoWorld**, *Small language models: rethinking enterprise AI architecture* (4 mai 2026, citant NVIDIA, Gartner et Info-Tech) : le routage vers des SLM réduit le coût d'inférence **jusqu'à 90 %** sur les tâches répétitives à fort volume ; NVIDIA qualifie désormais les SLM d'« avenir de l'IA agentique ». <https://www.infoworld.com/article/4160404/small-language-models-rethinking-enterprise-ai-architecture.html>
- [45] **Demis Hassabis** (Google DeepMind), fenêtre AGI resserrée à 2029-2030 (déclarations mai 2026, Google I/O / interview Axios).
- [46] **Dario Amodei** (Anthropic), *The Adolescence of Technology* (janvier 2026) : « powerful AI » possible à 1-2 ans, avec réserve explicite. <https://darioamodei.com/essay/the-adolescence-of-technology>
- [47] **BCG**, *AI at Work 2026: Why Strategy Matters More Than Tools* (3 juin 2026, 11 749 salariés, 14 marchés) : une stratégie IA claire = +25 points d'impact, contre ~+5 points pour de meilleurs outils seuls ; 74 % des cols blancs de première ligne sont des utilisateurs réguliers. <https://www.bcg.com/publications/2026/ai-at-work-why-strategy-matters-more-than-tools>
- [48] **Gartner**, *Worldwide Sovereign Cloud IaaS Spending* (communiqué du 9 février 2026) : 80 Md\$ en 2026 (+35,6 %) ; Europe de 6,9 à 12,6 Md\$ (+83 %), dépassant l'Amérique du Nord en 2027. <https://www.gartner.com/en/newsroom/press-releases/2026-02-09-gartner-says-worldwide-sovereign-cloud-iaas-spending-will-total-us-dollars-80-billion-in-2026>
- [49] **AION** (Ardian, Artefact, Bull, EDF, Capgemini, Iliad, Orange, Scaleway), candidature à une *AI Gigafactory* française dans le cadre d'InvestAI (5 juin 2026). <https://www.storagenewsletter.com/2026/06/05/>

Livre blanc « Maturité IA des PME françaises 2025-2026 : Méthode Junyr™ », par Paul-Antoine TUAL, AI Transformation Leader · Croissance et Transitions. Édition initiale Mai 2026 (golden standard, v2.0). Publication : 23 mai 2026. Révision v2.1 : 12 juin 2026. Révision v2.2 : 22 juin 2026 (deep research : correction d'attributions, enrichissement souveraineté & FinOps, section horizon AGI). Dernière révision : 23 juin 2026. **Édition Juin 2026, v2.3** : passe de fraîcheur stricte (toutes les sources de marché IA de plus de six mois remplacées par leurs équivalents 2026 : KPMG Global AI Pulse T1 2026, HCLTech, Deloitte / Gartner / BCG 2026, IDC-Lenovo CIO Playbook 2026, ADP People at Work 2026, paysage open-weight et FinOps de printemps 2026 ; régulations et preuves fondamentales conservées, datées). Révision suivante prévue : novembre 2026.

Document maître : maintenu en Markdown (`content/livre-blanc-maturite-ia-pme.md`), généré en HTML canonique sur junyr.fr/livre-blanc et en PDF distribué.