



Do Tank « Impacts de l'IA sur les activités d'ingénierie logicielle »

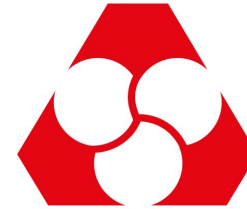
GT 5/6

« Optimisation du support »

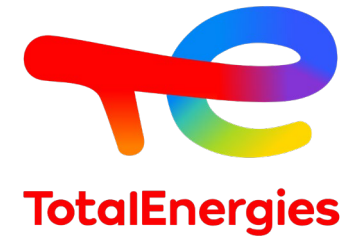
Les sociétés ayant participé au GT



Pilote du GT :
Viet NGUYEN



Crédit  Mutuel



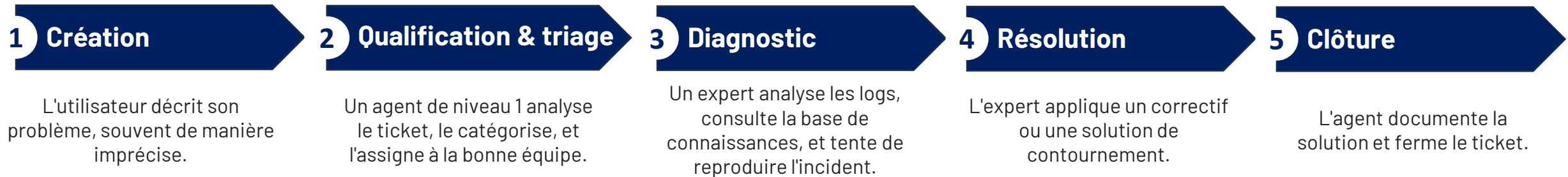
- 1. Introduction**
- 2. Analyse des opportunités de l'IA Générative pour le support IT**
- 3. Focus sur les serveurs MCP**
- 4. Conclusion**

01

Introduction

Introduction : les opportunités de l'IA Générative pour le support IT

Le modèle actuel de support IT repose majoritairement sur l'intervention humaine pour chaque étape du cycle de vie d'un ticket :



**Ce processus est coûteux, lent, sujet aux erreurs humaines et peu scalable.
L'IA Générative promet de réinventer chacune de ces étapes :**



Réduire le volume de tickets humains



Accélérer le MTTR (*Mean Time To Repair - Temps Moyen de Résolution*)



Renforcer l'expérience développeur (DevEx)



Améliorer la qualité, la perception et la cohérence du support



Réduire les coûts de Run

autre

02

Analyse des opportunités de l'IA Générative pour le support IT

Liste des personae au sein du support IT, potentiels bénéficiaires de l'intégration d'IA générative



Utilisateur technique

(Développeurs, Data Scientists)

- Demande potentiellement complexe (blocage, incident, etc.)



Utilisateur final non technique

- Recherche une assistance rapide et compréhensible pour un problème simple (ex. : accès bloqué, mot de passe oublié).



Technicien de support

(Support N1)

- Traite les demandes les plus simples
- Utilise l'IA pour obtenir des diagnostics, des recommandations de résolution ou automatiser certaines tâches répétitives.



Expert technique (Support N2/N3)

- Traite les demandes les plus complexes
 - Expert sur son domaine
- Utilise l'IA pour obtenir des informations techniques plus poussées afin de poser un diagnostic sur l'incident
- Utilise l'IA comme un outil de brainstorming, pour poser des questions et obtenir un résumé des interactions précédentes avec le demandeur.



Responsable d'équipe IT

- Assure le suivi des équipes (nombre de tickets, statistiques, etc.)
- Utilise l'IA pour obtenir des statistiques sur les problématiques récurrentes, les succès/échecs de résolution, et identifier les besoins en compétences (ex: pour des appels à prestation).

Des cas d'usage qui couvrent l'ensemble du cycle de vie du ticket

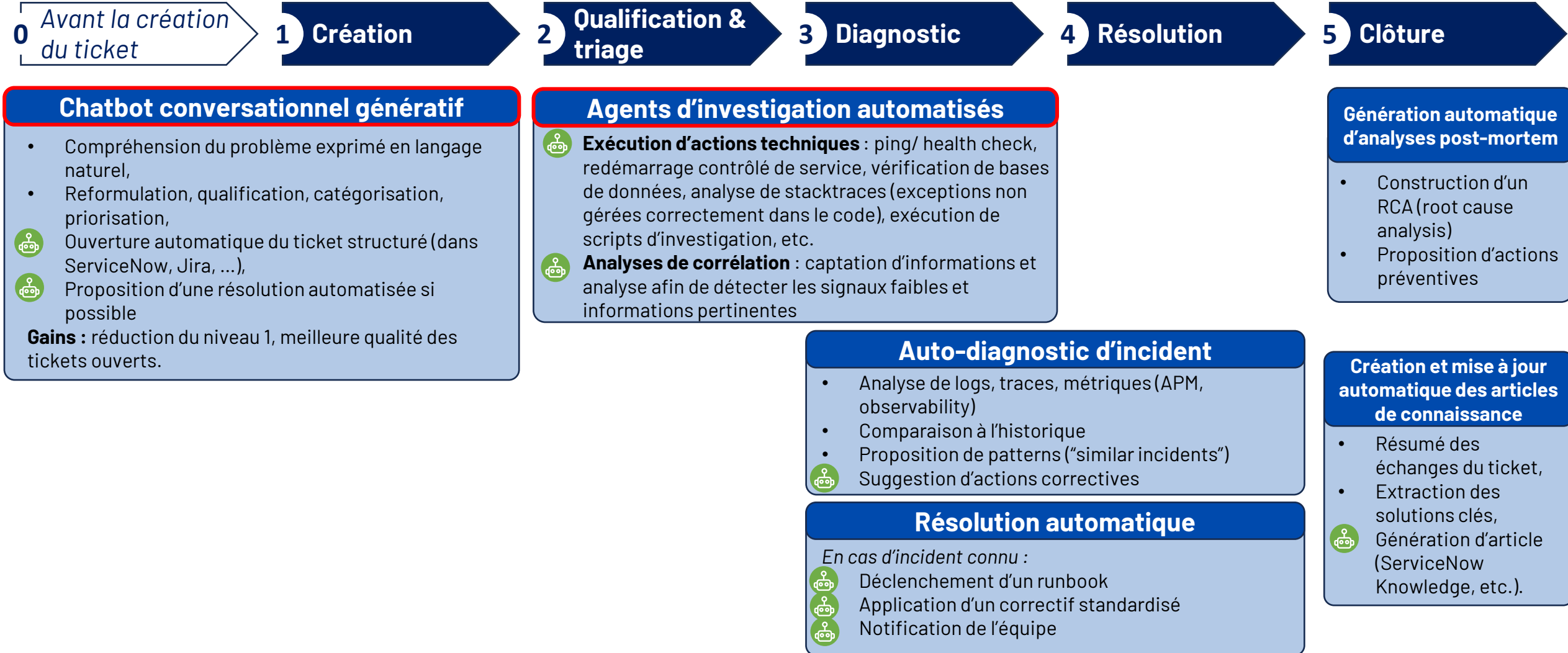




Gestion des connaissances : Capitalisation, et exploitation sur l'ensemble du cycle de vie du ticket

Cas d'usage en train d'être déployés par les membres du GT

- Utilisateur technique
- Utilisateur final
- Technicien support (N1)
- Expert technique (N2/N3)
- Responsable d'équipe

Capacités à acquérir en IA générative, et parfois en IA agentique



 Cas d'usage en train d'être déployés par au moins un des membres  Capacités agentiques (appel à des outils / tools)

REX sur les premiers cas d'usage déployés par les membres du GT



Prérequis et facteurs clés de succès

Mener au préalable un **chantier documentaire** :

- Centralisation de gisements documentaires souvent éparpillés
- Fusion des documents et/ou paragraphes en doublon
- Suppression des documents obsolètes
- Identification des paragraphes contradictoires



Intégrer et/ou interfacier des capacités d'IA avec les applications de ticketing utilisées

- L'export/import d'un outil à l'autre est trop fastidieux



Définir les **éléments de contexte métier** nécessaires à la réalisation des cas d'usage envisagés :

- Qu'est ce qu'un ticket long ?
- Qu'est ce qui fait un bon résumé ?
- Etc.



Défis à relever

Des **difficultés qui persistent** dans la gestion documentaire :

- Présence d'images dans la documentation
- Boucle entre références de documentation



Une **évaluation complexe des résultats** obtenus :

- Besoin de jeux de **données de test souvent spécifiques** et donc difficiles à constituer
- Activités nécessaires de **labellisation** des données, très **chronophages**
- **Variabilité dans le temps des résultats**, notamment, en cas de changement de LLM
- Équilibre à trouver dans les évaluations, entre élément **objectif** (test formel) et **subjectif** (dire d'expert)



Des **spécificités de chaque contexte** (processus, personae, critères de succès, etc.) qui complexifient l'intégration de ces capacités, et doivent faire l'objet d'une attention particulière pour en garantir la réussite.



Mesure actuelle et future des gains associés à l'IA dans le support

De premiers indicateurs mis en place



FCR – Résolution au Premier Contact

Nombre de tickets qui sont directement résolus par le 1er groupe affecté



MTTR – Temps Moyen de Résolution

Durée moyenne nécessaire à la résolution complète d'un incident



Temps gagné à dire d'expert

Retours quantitatifs et qualitatifs sur les gains réalisés grâce aux IA déployées



Satisfaction utilisateurs

NPS (*Net Promoter Score*), réalisation d'enquêtes post-interactions



Une mesure qui devrait s'enrichir, notamment à partir du « marquage » des tickets traités par IA



Taux de résolution automatisée des tickets



Taux d'erreurs dues à l'IA



Taux d'escalade vers un humain



Évolution du volume global d'incidents

03

Paradigmes d'architecture associés au support augmenté

Différentes alternatives, avec un niveau de complexité croissant

Complexité

1

IA intégrée à l'ITSM (*IT Service Management*): ServiceNow, Jira Service Management, etc.

2

LLM Centralisé avec architecture RAG (*Retrieval-Augmented Generation*)

3

Système d'agents décentralisés (*Vision agentique*)

Déploiement simple et intégration native

Maîtrise, confidentialité et scalabilité

Spécialisation, scalabilité et périmètre d'actions

Peu de maîtrise, coûts importants et risques de dépendance

Complexe, et induit des coûts conséquents de développement et de maintenance

Très complexe et encore peu mature à ce stade

Modèle 1 : IA Intégrée à l'ITSM (IT Service Management : ServiceNow, Jira Service Management...)

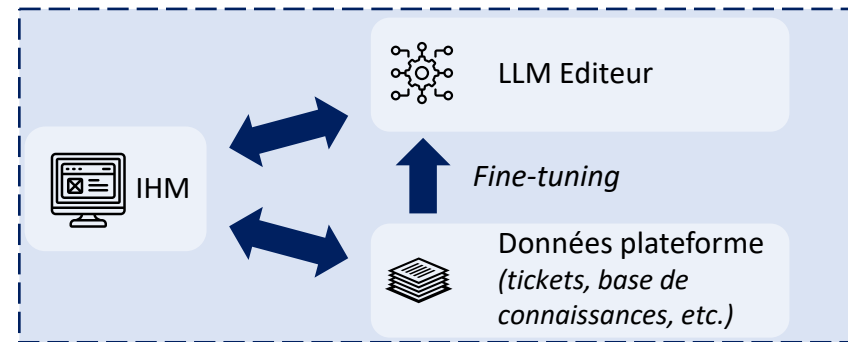


Description et architecture

Utilisation des fonctionnalités d'IA générative natives proposées par les grands éditeurs d'ITSM. Ces solutions sont souvent "clés en main".



Utilisateur



Plateforme ITSM



Avantages

- **Simplicité** : déploiement rapide, pas d'infrastructure à gérer.
- **Intégration parfaite** : accès natif à toutes les données de l'outil.
- **Support et maintenance** assurés par l'éditeur.



Inconvénients

- **Boîte noire** : peu de contrôle sur le modèle de langage, le prompt, ou la manière dont les données sont utilisées.
- **Coût** : souvent un add-on coûteux à la licence existante.
- **Périmètre limité** : l'IA n'a pas accès aux données externes (logs, monitoring, code source) sauf via des connecteurs spécifiques et souvent limités.
- **Dépendance** (*Vendor Lock-in*).

Modèle 2 : LLM Centralisé avec Architecture RAG (Retrieval-Augmented Generation)

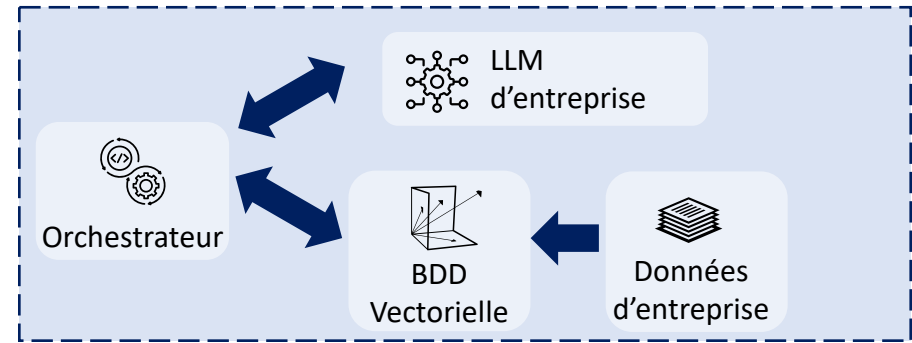


Description et architecture

C'est le modèle le plus populaire pour une solution sur mesure. On combine un LLM généraliste (ex Gemini 3.0, Mistral Large 3) avec une base de connaissances privée et contextuelle.



Utilisateur



Plateforme IA d'Entreprise



Avantages

- **Précision et Contrôle** : réduit drastiquement les "hallucinations" en forçant le LLM à se baser sur les données de l'entreprise.
- **Confidentialité** : les données restent dans le périmètre de l'entreprise (surtout en cas de LLM auto-hébergé).
- **Extensibilité** : il est facile d'ajouter de nouvelles sources de données (il suffit de les indexer).



Inconvénients

- **Complexité** : nécessite des compétences en ingénierie de données, ML et prompt engineering.
- **Coût de développement et de maintenance** de l'infrastructure (pipeline de données, base vectorielle, API du LLM).
- **Évaluation des résultats complexes** : besoin de données de tests spécifiques et souvent labellisées à la main, et variabilité des résultats.

Modèle 3 : Système d'Agents Décentralisés

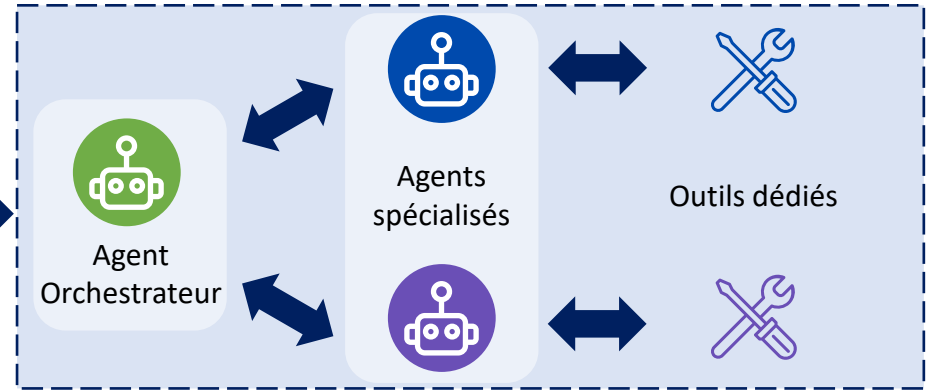


Description et architecture

C'est le futur du RUN autonome. Au lieu d'un seul système monolithique, on déploie une flotte d'agents spécialisés qui collaborent (tri de tickets, récupération d'informations, etc.)



Vision prospective, pas encore mise en place par les membres du GT mais qui devrait l'être dès 2026



Plateforme agentique d'entreprise



Avantages

- **Spécialisation extrême** : chaque agent est très performant dans sa tâche.
- **Scalabilité et Résilience** : on peut ajouter de nouveaux agents ou mettre à jour un agent sans impacter les autres.
- **Capacités étendues** : permettent de passer de la simple suggestion à l'action et à l'auto-réparation.



Inconvénients

- **Sécurité** : donner à des agents la capacité d'agir sur des systèmes de production est un risque majeur qui nécessite des garde-fous extrêmement robustes.
- **Complexité** : l'orchestration des agents et la gestion des erreurs sont des défis majeurs. Des plateformes clés en main (Azure Foundry, etc.) sont facilitantes par rapport aux frameworks comme LangChain Agents ou AutoGen.

Contraintes et bonnes pratiques



Sécurité et Conformité

Les réglementations en vigueur (AI Act, ISO 27001, RGPD, etc.) doivent être prises en compte quelle que soit l'architecture mise en œuvre. L'AI Act pose des exigences de [gouvernance](#) de l'ensemble des outils d'IA d'une organisation, et donc de ceux relatifs au support IT.



Données sensibles

L'intégration de données sensibles augmente la complexité de ces architectures, et exige la définition claire de politiques d'anonymisation des données, ainsi que de conservation et d'effacement des logs.



Adoption et accompagnement

L'embarquement des équipes (selon les personnes concernées) est critique pour la réussite de ces nouveaux outils. Leur mise en œuvre doit par conséquent s'accompagner de formations et de communications ad-hoc et dans la durée, ainsi que d'actions d'accompagnement au changement pour garantir l'adhésion.

Tableau Comparatif des architectures étudiées

Critère	Modèle 1: Intégré ITSM	Modèle 2: RAG Centralisé	Modèle 3: Agents Décentralisés
Complexité de mise en œuvre	★ Faible	~ Moyenne à Élevée	– Très Élevée
Coût (Build + Run)	~ Coûts de licence élevés	~ Coût de dev initial + coût d'infra/API	– Coût de R&D et d'infra très élevé
Contrôle & Personnalisation	– Faible	+ Élevé	★ Très Élevé
Précision & Pertinence	~ Moyenne	+ Élevée	★ Très Élevée (potentiellement)
Confidentialité des données	~ Dépend de l'éditeur	+ Élevée (si bien conçu)	+ Élevée (si bien conçu)
Capacité d'action autonome	– Nulle à Faible	– Faible (via connecteurs)	+ Élevée
Maturité de la solution	+ Mature	~ En cours de maturation	– Émergente / Expérimentale

04

Conclusion

Conclusion

- 1** L'IA est un **accélérateur considérable de l'ensemble des activités du support IT**. Leur caractère autonome est cependant plus limité que sur d'autres thématiques (développement, design, etc.) : la **présence d'humains (*Human In The Loop*)** reste absolument **nécessaire**.
- 2** Les **modèles d'architecture sur mesure restent complexes à déployer**, mais offrent une promesse de maîtrise des données, des modèles et des coûts afférents. L'**agentique** amplifie encore les opportunités de ces solutions internes : de premiers déploiements au sein des membres du GT sont attendus dès le 1^{er} trimestre 2026.
- 3** Les conséquences en termes d'**évolution des pratiques** restent à approfondir. Ceci vaut également pour la **mesure de la valeur** apportée par l'IA sur le support, qui diffère d'une entreprise, et même d'une équipe à une autre et bénéficierait du « marquage » des tickets traités par l'IA pour démontrer son impact.