HUB FRANCE IA

AGENTS EXPERTS IA

Janvier 2026

Table des matières

HUB FRANCE IA

Résumé exécutif

AGENTS EXPERTS IA
Janvier 2026

Résumé exécutif

Ce livre blanc traite des **agents experts IA**, technologie en plein essor qui transforme profondément la manière dont les entreprises et institutions gèrent les tâches complexes, automatisées et interactives.

Nous analysons successivement :

- les **cas d'usage sectoriels et métiers** : santé, finance, industrie, éducation, etc. ;
- la **typologie des agents experts IA** : des agents conversationnels aux systèmes multi-agents ;
- l'**architecture des agents experts IA** : modules techniques fondamentaux (perception, raisonnement ou décision) ;
- les **techniques avancées** : approches symboliques, neuro-symboliques, numériques et algorithmiques, pour garantir fiabilité et efficacité ;
- le **fonctionnement des agents experts IA** : architectures hiérarchiques ou décentralisées pour une optimisation globale des performances ;
- la **gouvernance des agents experts IA** : maîtrise des enjeux éthiques, sécurité des données, conformité réglementaire (règlement européen sur l'intelligence artificielle [RIA], RGPD[1]), gestion proactive des risques émergents.

Ce livre blanc élaboré par une équipe d'expertes et experts dans le domaine de l'IA, l'IA distribuée et leurs applications, se propose comme une contribution essentielle à la compréhension des enjeux complexes liés aux agents IA et à la maîtrise opérationnelle des solutions innovantes, au service d'une IA responsable et durablement performante. Il offre une analyse approfondie, intégrant les dimensions pratiques, métiers, technologiques et stratégiques, destinée aux décideurs et aux techniciens afin d'optimiser le déploiement sécurisé, efficace et responsable des systèmes à base d'agents IA.

Les **agents experts IA**, renforcés par les récentes avancées de l'IA générative, constituent un levier technologique et organisationnel d'une importance capitale pour l'avenir. Toutefois, la pleine réalisation de leur potentiel requiert un cadre rigoureux, garantissant la maîtrise technique, éthique et réglementaire de ces systèmes.

[1] Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=celex:32016R0679

HUB FRANCE IA

Guide de lecture rapide

AGENTS EXPERTS IA
Janvier 2026

Guide de lecture rapide

Nous recommandons la lecture des chapitres d'introduction (page 13) et de conclusion (page 99) pour tous les lecteurs.

Le document s'adresse à différents profils de lecteurs :

- **décideurs** (chapitres 2 et 0) : pour comprendre rapidement les bénéfices opérationnels et les impératifs de gouvernance ;
- **architectes et développeurs IA** (chapitres 0, 0 et **Error! Reference source not found.**) : pour maîtriser les dimensions techniques, méthodologiques et opérationnelles des agents experts IA ;
- **experts en conformité et gouvernance** (chapitre 0) : pour intégrer dès le départ les contraintes réglementaires et éthiques dans les projets.

La structure modulaire du document permet une lecture sélective en fonction des priorités et des intérêts du lecteur. Nous recommandons toutefois une lecture préalable du résumé exécutif afin de bénéficier d'une vue d'ensemble avant d'aborder les chapitres techniques.

HUB FRANCE IA

À propos de ce Livre Blanc

À propos de ce livre blanc

Le présent livre blanc, issu du travail collaboratif des expertes et experts du Hub France IA, propose une synthèse approfondie sur les agents experts dans le contexte de l'IA générative. Notre ambition est d'offrir aux décideurs, architectes techniques et responsables opérationnels une compréhension complète des potentiels, limites, et impératifs associés aux agents intelligents.

Ce document se structure autour de six chapitres clairement identifiés, allant des cas d'usages concrets à la gouvernance des agents, en passant par la typologie des agents, leurs architectures techniques et les techniques avancées de l'IA. Ce travail constitue ainsi une ressource précieuse à la fois pour l'action immédiate et pour la réflexion stratégique de moyen et long terme.

HUB FRANCE IA

1. Introduction

1. Introduction

L'essor rapide des modèles d'IA Générative a profondément modifié la manière dont les organisations conçoivent l'automatisation, l'analyse de données et l'interaction homme-machine. De nos jours, l'IA transcende les capacités limitées des fonctions isolées. Elle s'intègre désormais à des systèmes complexes et interconnectés, au sein desquels plusieurs agents interagissent de manière continue, tant entre eux qu'avec les Systèmes d'Information (SI). C'est ce que l'on appelle les **agents experts IA**, également connus sous différentes appellations : IA agentique, agents experts ou spécialisés, agents IA experts, agents génératifs ou encore agents experts *GenAI.* Chaque entreprise choisit sa propre terminologie en fonction de sa stratégie de déploiement de l'IA.

Dans un contexte technologique en pleine mutation, l'intégration des grands modèles de langage (*LLM*[2]) et des techniques avancées (symboliques, neuro-symboliques, etc.) transforme les capacités opérationnelles des entreprises et des institutions. Cette évolution soulève toutefois des enjeux majeurs en matière de gouvernance, de sécurité et d'éthique.

Cette transformation s'inscrit dans un contexte économique marqué par un contexte macroéconomique dégradé. En effet, selon le Capgemini *Research Institute*, les agents experts IA pourraient générer jusqu'à 450 milliards de dollars de valeur économique d'ici 2028[3]. Parallèlement, Gartner prévoit que 33 % des applications d'entreprise incluront l'IA agentique (agents experts IA) d'ici 2028, contre moins de 1 % en 2024[4]. Cette accélération conduit 93 % des dirigeants à penser que le déploiement rapide des Agents experts IA constituera un avantage concurrentiel majeur.

1.1. C'est quoi les agents experts IA ?

Indépendamment de la nomenclature employée, un agent IA se caractérise par une capacité d'action sur son environnement (logiciel ou même physique) et par un niveau

[2] *LLM* : *Large Language Model.*
[3] Capgemini Research Institute. Trust and human-AI collaboration set to define the next era of agentic AI, unlocking $450 billion opportunity by 2028. Capgemini, p.1–10. 16 July 2025. https://www.capgemini.com/news/press-releases/trust-and-human-ai-collaboration-set-to-define-the-next-era-of-agentic-ai-unlocking-450-billion-opportunity-by-2028/
[4] Gartner**.** Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027. Gartner, p.1–10. 25 June 2025. https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027

d'autonomie gradué. Dans ce livre blanc, l'autonomie ne signifie pas absence de supervision : elle désigne la capacité à enchaîner des étapes, à planifier et exécuter des actions dans un cadre de règles et de garde-fous. À l'inverse, de nombreux produits sont couramment qualifiés d'« agents » dans les communications, mais relèvent plutôt d'assistants LLM qui ne disposent pas d'outils, de droits d'action et d'une orchestration (mémoire, planification, déclencheurs). Un agent repose au minimum sur des modèles génératifs, mais pas uniquement. On retrouve également des agents multimodaux, intégrant aussi bien l'apprentissage automatique traditionnel que des modèles conversationnels.

On distingue généralement trois grandes catégories souvent qualifiées d'agents IA :

- **les assistants conversationnels** (ChatGPT, Gemini, Claude, Microsoft Copilot, Mistral Le Chat, etc.), centrés sur l'interaction en langage naturel et l'aide à la décision ;
- **les copilotes bureautiques** (Microsoft 365 Copilot, Google Duet AI, Notion AI, Zoho Zia, GrammarlyGO, etc.), centrés sur la productivité dans des suites applicatives ;
- les **agents experts**, spécialisés sur une tâche ou un processus métier précis, capables d'exécuter des actions via des outils/API dans un cadre de règles et de supervision.

Les assistants conversationnels et les copilotes sont généralement réactifs (déclenchés par l'utilisateur) et ne deviennent « agentiques » que lorsqu'ils disposent d'outils d'action, d'une mémoire/état et de mécanismes de planification/orchestration. L'autonomie doit donc être comprise selon une échelle (passif – réactif – proactif – dubitatif – intentionnel), détaillée dans la section 3.3.1.

Les agents experts IA s'intègrent aux systèmes existants pour renforcer la valeur métier par l'automatisation. Ils interagissent avec une grande variété de sources de données — e-mails, partages réseau, *SharePoint, ERP, CRM*, ou tout autre système d'information — et sont capables de traiter des contenus hétérogènes, que ceux-ci soient structurés ou pas.

En fonction du contexte métier, ces agents peuvent :

- lire les données des systèmes ;
- analyser les données structurées et non structurées pour détecter des tendances ou anomalies ;
- mettre à jour les données des systèmes ;
- rédiger des rapports ou des courriels adaptés au type d'interlocuteur ;
- synchroniser les informations entre plusieurs systèmes pour garantir la cohérence des données.

1.2. Pourquoi les agents experts IA ?

Les agents experts IA vont au-delà de la simple interaction textuelle. Ils sont conçus pour agir, s'intégrer, et automatiser des tâches dans des environnements dynamiques et complexes. Leur objectif est de transformer l'IA en outil opérationnel, capable de prendre des décisions, d'exécuter des actions, et de s'adapter à des contextes métier spécifiques. **Contrairement à un simple *prompt*/conversation, dont les résultats sont souvent imprévisibles et nécessitent plusieurs essais manuels pour obtenir une réponse satisfaisante,** les agents experts IA permettent de mieux contrôler la qualité des processus, de limiter les hallucinations des modèles, et d'assurer une explicabilité et une observabilité à chaque niveau du système. Toutefois, l'intégration d'un agent expert d'IA requiert la présence de règles prédéfinies pour atteindre un seuil de performance désiré, car le processus ne peut s'appuyer sur un utilisateur susceptible d'effectuer de multiples tentatives et de servir de mécanisme de contrôle au modèle. Chaque agent agit de manière autonome ou semi-automatique sur une tâche bien définie, selon des règles métier qui peuvent être établies par un humain ou apprises à partir de données, ce qui permet à l'utilisateur d'ajuster les résultats. Cette approche réduit significativement les erreurs humaines, renforce la fiabilité des décisions, et garantit la traçabilité et la conformité dans l'exécution des tâches.

Les résultats produits par l'IA (non symbolique) ne **garantissent** jamais un taux de détection ou de précision de 100 % (voir le tableau 1 en guise d'exemple). En pratique, **un agent IA est considéré comme fiable à partir d'un certain niveau de performance**, mais ce niveau dépend fortement de la **qualité des données** et de la **complexité du processus métier**. Il est donc essentiel de **maintenir l'humain dans la boucle** pour intervenir sur les cas complexes. Certains processus nécessitent toujours une part de travail manuel, et dans certains contextes spécifiques, même un humain peut rencontrer des difficultés d'interprétation. L'IA ne peut pas tout automatiser, mais elle peut considérablement améliorer la fiabilité et l'efficacité lorsqu'elle est utilisée de manière encadrée et complémentaire.

Caractéristiques	***Prompt*/conversation**	**Agent expert d'IA**
Interaction	Réactive, limitée à une session	Continue, proactive, contextuelle
Mémoire	Pas ou peu de capacité à mémoriser	Capacité à mémoriser et suivre des objectifs
Connexion au système	Non connecté	Connecté à des bases de données, outils métier (*CRM, ERP*, etc.)

Exécution des tâches	Génère du texte ou des réponses	Exécute des actions : mise à jour de systèmes, envoi d'e-mails, génération de rapports
Spécialisation métier	Généraliste	Spécialisé dans un domaine ou processus métier
Autonomie	Réactif : dépend des sollicitations de l'utilisateur ; pas d'initiative ni d'actions sur SI.	Autonomie encadrée : peut planifier et exécuter des actions via règles/outils, avec supervision, validations ou garde-fous selon le niveau de risque.

Tableau 1 : Tableau comparatif *prompt/conversation – Agents experts IA*

À partir des agents experts IA, il devient possible de concevoir des systèmes multi-agents (SMA), dans lesquels plusieurs agents, chacun doté d'une expertise métier ou fonctionnelle spécifique, communiquent, se coordonnent (notamment par la collaboration) et s'organisent pour contribuer à exécuter des tâches complexes. Un agent orchestrateur coordonne l'affectation d'objectifs aux agents experts en fonction de la nature des requêtes et de la compétence spécifique de chaque agent. Ces systèmes se distinguent par leur capacité à répartir les rôles, à s'adapter dynamiquement aux contextes, à interagir avec des environnements multiples (*ERP*, *CRM*, bases de données, *API*, etc.), et à fonctionner de manière distribuée, évolutive et résiliente. Chaque agent reste strictement limité à son périmètre fonctionnel et agit selon des règles métier préétablies par l'humain, garantissant ainsi la fiabilité, la traçabilité et la conformité des actions menées dans un environnement distribué et évolutif.

1.3. Les enjeux majeurs des agents experts IA

Les agents experts IA répondent à un besoin croissant de performance et d'adaptabilité dans des environnements dynamiques et complexes. Ces systèmes permettent :

- une **spécialisation** accrue des agents pour des performances optimisées ;
- une **résilience** renforcée, par la distribution des tâches entre agents multiples ;
- une **scalabilité** nécessaire, face à l'explosion des volumes de données et la complexité croissante des tâches cognitives automatisées.

Cependant, les agents experts IA ne sont pas sans défis. La collaboration des agents, la gestion des risques liés aux biais ou aux hallucinations des modèles, la sécurité des données manipulées par les agents et la conformité avec les réglementations émergentes (RIA, règlement général sur la protection des données [RGPD], règlement

européen sur les services numériques [DSA], etc.) constituent autant d'enjeux critiques que les organisations doivent impérativement maîtriser.

Par exemple, dans les véhicules autonomes, les agents experts IA opèrent selon une échelle d'autonomie de 0 (aucune autonomie) à 5 (autonomie complète).

1.4. Défis techniques, éthiques et organisationnels

Les agents experts IA ne se limitent pas à un simple assemblage de programmes informatiques. Ils nécessitent :

- **une conception technique rigoureuse** : architectures robustes (perception, raisonnement, mémoire), techniques hybrides intégrant l'IA générative, un système de connaissances, le *Machine Learning* et/ou l'IA symbolique ou neuro-symbolique ;
- une **maîtrise des biais et hallucinations** propres aux modèles génératifs ;
- une **gouvernance proactive**, à même de répondre rapidement aux défis éthiques, réglementaires et organisationnels : accès aux données, transparence algorithmique, conformité réglementaire et sécurité opérationnelle.

L'intégration réussie des agents experts IA dans les processus métier suppose donc une vision stratégique claire, appuyée sur une gouvernance adaptée et une architecture technique solide.

Une crise de confiance dans l'IA générative se profile à l'horizon. La confiance accordée aux agents totalement autonomes a connu une baisse significative de 43 % à 27 % au cours de la dernière année[5]. Cette diminution témoigne d'une prise de conscience accrue des défis concrets liés à l'implémentation de ces technologies, après une phase initiale d'enthousiasme. Par ailleurs, il convient de souligner que plus de 80 % des organisations ne disposent pas d'une infrastructure d'IA suffisamment mature pour permettre un déploiement à grande échelle[6].

[5] Capgemini *Research Institute*. L'essor de l'IA agentique : la confiance, clé de la collaboration humain–IA. Capgemini, p.1–10. Juillet 2025. https://www.capgemini.com/fr-fr/perspectives/publications/data-ia-essor-ia-agentique/

[6] Noam Kolt. Governing AI Agents. *arXiv preprint arXiv:2501.07913*. January 2025. https://arxiv.org/pdf/2501.07913

HUB FRANCE IA

2. Cas d'usage

2. Cas d'usage

En janvier 2024, nous avons publié un livre blanc consacré aux usages de l'IA générative[7], centré sur les modèles de langage de grande taille (*LLM*). À travers une analyse des applications dans six secteurs clés[8], ainsi que sur l'apport des *LLM* aux *chatbots*, nous concluions que « l'année 2023 a été l'année des *LLM*. Mais c'est l'année 2024 qui permettra d'établir les bonnes pratiques pour le déploiement des *LLM* ». Cette prévision s'est effectivement vérifiée.

L'année 2024 a constitué une étape significative avec l'introduction des agents experts IA. Si les grands *LLM* se concentraient principalement sur la génération de texte et l'assistance conversationnelle (notamment avec l'émergence de nouveaux modes d'interaction tels que la voix, le son, l'image, la vidéo et les données structurées), l'essor des IA multimodales s'est accentué durant cette période. Comme nous allons le voir dans cette section, les agents ont ainsi permis d'étendre les capacités de l'IA au-delà de ces fonctions initiales :

- **dans le secteur financier** : des copilotes dotés de capacités d'automatisation de la génération de rapports et de conformité réglementaire sont intégrés aux systèmes existants ;
- **dans l'industrie manufacturière** : des agents de maintenance prédictive sont incorporés aux systèmes de production afin d'optimiser les opérations et de minimiser les temps d'arrêt non programmés ;
- **dans les ressources humaines** : des assistants de recrutement sophistiqués sont déployés pour analyser et présélectionner efficacement un volume important de candidatures ;
- **dans le domaine administratif** : des agents de traitement automatisé de documents sont mis en œuvre pour réduire significativement les délais de traitement et améliorer l'efficacité globale (par exemple, le compte-rendu automatique de conversations, de réunions, la synthèse de documents).

Ces cas d'usage démontrent la manière dont les agents experts IA optimisent et amplifient les capacités des *LLM*. Au-delà de la simple génération de contenu, ces

[7] Hub France IA. Les usages de l'IA Générative. *Vol. 1 Les LLMs. Livre blanc. Janvier 2024*. https://www.hub-franceia.fr/wp-content/uploads/2024/02/Livre-blanc_Les-usages-de-lia-generative-01.2024.pdf

[8] Cybersécurité, industries culturelles et créatives, ressources humaines, développement informatique, éducation et marketing.

agents interagissent directement avec les systèmes d'information et exécutent des actions concrètes, contribuant ainsi à une intégration plus poussée et efficace au sein des processus métier.

Pour bien préciser les différences avec les LLM et les apports des agents, nous proposons quelques exemples illustratifs.

Contexte	**LLM seul**	**Ce que fait l'agent**
Support client	Rédiger un email de réponse client	Lire le ticket, interroger le CRM, rédiger la réponse, l'envoyer, mettre à jour le ticket
Juridique	Résumer un contrat	Extraire les clauses à risque, les comparer à la base réglementaire, alerter le juriste, générer un rapport
Logiciel	Suggérer du code (quelques lignes sur un sujet précis)	Concevoir et rédiger le code (plusieurs fichiers), le tester, corriger les bugs, l'insérer dans le référentiel

Tableau 2 : Tableau comparatif *LLM seul – Agent expert IA*

Cette évolution prépare le terrain à ce que nous désignons par le terme « **agentique** » : une nouvelle phase caractérisée par la collaboration, la coordination et l'orchestration d'agents experts IA au sein de systèmes multi-agents (SMA). Si les *LLM* ont ouvert la voie à l'automatisation du langage, l'agentique, quant à elle, ouvre la voie à l'automatisation de processus complexes, grâce à sa capacité d'adaptation, de supervision et de gouvernance à grande échelle. Beaucoup des exemples ci-dessous contiennent un *LLM*, doté de fonctions agentiques qui s'enrichissent progressivement.

À ce jour, une large gamme d'agents experts IA est accessible[9] pour des tâches spécifiques et relativement simples, encore très proches des LLM, telles que la rédaction de résumés structurés, la transcription de conversations avec un style prédéfini, ou encore l'extraction automatique des actions à partir de notes de réunion. Ils sont souvent appelés **assistants IA**.

Au-delà de ces applications initiales, on constate l'émergence de solutions d'une portée considérablement plus importante au sein des entreprises. Ces nouvelles générations d'agents ne se contentent plus d'une fonction d'assistance ; elles automatisent,

[9] Ranjan Sapkota, Konstantinos I. Roumeliotis, Manoj Karkee. AI agents vs. agentic AI: A conceptual taxonomy, applications and challenge. *arXiv preprint arXiv:2505.10468*. May 2025. https://arxiv.org/abs/2505.10468.
And published in *Information Fusion*, Volume 126, Part B 103599, Elsevier, p. 1-30. February 2026. https://www.sciencedirect.com/science/article/pii/S1566253525006712?via%3Dihub

exécutent et sécurisent l'intégralité des processus métier, générant ainsi une valeur opérationnelle et stratégique substantielle.

Afin d'illustrer cette dynamique, nous présentons une série de cas d'usage selon deux perspectives complémentaires :

- **par secteur d'activité** : santé, finance, industrie, éducation, administration, etc. ;
- **par fonction métier** : *marketing, supply chain*, ressources humaines, juridique, développement logiciel, etc.

Cette double lecture a pour objectif d'inspirer un large éventail de décideurs et de praticiens. Elle ne prétend pas couvrir l'intégralité des déploiements en cours, mais plutôt fournir des repères concrets et illustratifs, afin d'éclairer les choix stratégiques et d'anticiper les prochaines étapes du déploiement des agents experts IA dans les organisations. Les cas d'usage décrits ci-après sont soit déjà mis en œuvre, soit en cours de développement, soit envisagés pour l'avenir. Dans les deux premiers cas, des références sont fournies. Le rythme soutenu des annonces et des usages est tel que ce qui n'est actuellement qu'une perspective pourrait se concrétiser au moment de la lecture de ce document.

2.1. Cas d'usage par secteurs d'activités

2.1.1. Santé et médico-social

Agents d'accompagnement au parcours de soins personnalisés

On savait que les LLM seuls pouvaient assister les soignants de plusieurs façons : transcrire le résultat d'une consultation, rédiger un compte-rendu d'analyse, répondre aux questions d'un patient via un chatbot. D'ailleurs, de tels assistants conversationnels thérapeutiques sont proposés par *Character.ai*, offrant ainsi une assistance 24 heures sur 24[10]. Bien qu'ils ne remplacent pas un thérapeute qualifié, de nombreuses personnes les utilisent à cette fin.

Mais les agents, avec leur autonomie, permettent d'aller plus en profondeur en orchestrant de bout en bout une étape du parcours patient (de la prise de rendez-vous à la mise à jour du dossier médical, en passant par l'envoi des documents au bon

[10] Hélène Pagesy. Comment l'IA bouscule le milieu de la santé mentale. *Le Monde*. 06 août 2024. https://www.lemonde.fr/pixels/article/2024/08/06/comment-l-ia-bouscule-le-milieu-de-la-sante-mentale-plutot-que-de-payer-une-nouvelle-seance-chez-le-psy-j-allais-sur-chatgpt_6270640_4408996.html

interlocuteur). Les personnels de santé restant les « *touch points* » de l'expérience patient.

Par exemple, des agents sont déployés pour gérer le parcours post-consultation : l'agent récupère le compte-rendu dicté par le médecin, le structure selon le format requis, met à jour le dossier patient dans le système d'information hospitalier (SIH), identifie les examens complémentaires prescrits, déclenche les demandes de rendez-vous auprès des services concernés, et envoie au patient un récapitulatif personnalisé avec les prochaines étapes. L'ensemble s'exécute de manière autonome, avec une validation humaine uniquement sur les points critiques.

En France, Doctolib a lancé fin 2025 un assistant téléphonique fondé sur l'IA, capable de dialoguer naturellement avec les patients, de prendre des rendez-vous intégrés directement à l'agenda du praticien, et d'enregistrer les demandes administratives (renouvellements d'ordonnance, certificats) dans la messagerie du cabinet[11].

Agents experts en recherche biomédicale

Des agents experts sont déployés afin d'accélérer la recherche biomédicale et la découverte de médicaments en allant bien au-delà de ce que les seuls LLM pouvaient déjà apporter (résumer des articles scientifiques, suggérer des hypothèses de recherche, générer des descriptions de molécules). Ils vont piloter de manière autonome un cycle de recherche : formuler une hypothèse, interroger les bases de données, planifier des expériences *in silico*, analyser les résultats, ajuster l'hypothèse, et itérer jusqu'à convergence.

Ces agents planifient des flux de travail, procèdent à une auto-évaluation afin d'identifier et de combler leurs lacunes, et utilisent des modèles génératifs et de langage pour un apprentissage continu. Ils intègrent des connaissances scientifiques et biologiques afin d'influencer des domaines tels que la simulation cellulaire, le contrôle des phénotypes, la conception de circuits cellulaires et le développement de thérapies ciblées[12].

[11] Intelligence artificielle : Doctolib lance un Assistant téléphonique et poursuit le déploiement de l'Assistant de consultation. *About Doctolib – France*. 02 Décembre 2025. https://about.doctolib.fr/news/doctolib-lance-un-assistant-telephonique/

[12] Shanghua Gao et al. Empowering biomedical discovery with AI agents. *Cell*, vol. 187, no 22, 6125-51. October 2024. ScienceDirect. https://www.cell.com/cell/fulltext/S0092-8674(24)01070-5?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867424010705%3Fshowall%3Dtrue

Par exemple, le système *BioDiscoveryAgent* développé par des chercheurs de Stanford et CMU démontre qu'un agent autonome peut concevoir des expériences de génétique, les exécuter via des outils de biologie computationnelle, et améliorer ses performances au fil des itérations, surpassant les approches non-agentiques sur des tâches de découverte génétique[13].

2.1.2. Finance et assurance

Agents d'automatisation des processus financiers

Des agents experts IA spécialisés dans l'analyse de bilans d'entreprise automatisent l'extraction des informations clés depuis les états financiers, mettent à jour les systèmes ERP et CRM, etc.

Dans le domaine de la conformité, des agents dédiés aux processus *Know Your Customer* (KYC) assurent une gestion intégrée englobant la reprise de stock documentaire, la remédiation continue et l'automatisation des contrôles. Ces agents interrogent les bases de données réglementaires, croisent les informations clients, détectent les incohérences et génèrent les rapports de conformité — réduisant significativement les délais de traitement tout en renforçant la traçabilité.

En assurance, des agents de détection de fraude examinent en temps réel les flux de déclarations de sinistres. Ils analysent les patterns suspects, croisent les données avec l'historique client et les bases sectorielles, et déclenchent des alertes graduées selon le niveau de risque identifié[14].

BNP Paribas a déployé sur son portail corporate un agent virtuel, Noa, accessible à 100000 utilisateurs. Cet agent analyse les demandes des clients, génère des réponses lorsqu'il dispose des informations nécessaires via API, ou crée un ticket orienté vers l'équipe appropriée dans le cas contraire. Il automatise également la production de documents officiels, tels que les confirmations de paiement. La banque prévoit d'intégrer l'IA générative pour améliorer la compréhension des requêtes, tout en conservant des

[13] Yusuf Roohan, et al. BioDiscoveryAgent: An AI Agent for Designing Genetic Perturbation Experiments. *arXiv preprint arXiv:2405.17631.* March 9, 2025. https://arxiv.org/pdf/2405.17631

[14] Pipefy. AI Agents for Insurance Companies: How They Help in Fraud Detection. *Pipefy*, March 2025. https://www.pipefy.com/blog/ai-agents-for-insurance-companies/

traitements déterministes pour la génération des réponses — une approche hybride visant à maîtriser les risques d'hallucination[15].

Des institutions financières développent des agents conversationnels spécialisés pour les fonctions de trésorerie. J.P. Morgan Payments travaille sur un assistant capable de comprendre le métier de trésorier et de répondre à des requêtes analytiques complexes en langage naturel comme par exemple, générer instantanément une analyse de liquidité avec graphiques et contexte, sans écrire de code. L'évolution vers des "agents IA" capables d'exécuter des actions multi-étapes de manière autonome est envisagée comme prochaine étape[16].

Par ailleurs, le Lab IA de BNP Paribas Global Markets a développé un système automatisé qui analyse quotidiennement les publications des analystes et économistes de la banque, génère des scores de sentiment sur l'inflation, la croissance et l'attitude des banques centrales, et publie ces indicateurs à destination des clients. Ce workflow illustre l'intégration d'agents dans l'offre produit elle-même, au-delà des seuls gains de productivité internes.

Citons aussi le cas d'étude académique sur l'"*Agentic AI for Financial Crime Compliance".* Elle montre un système d'agents orchestrant l'onboarding, la surveillance des transactions, l'investigation des alertes et la génération de rapports, avec des rôles clairement bornés et une traçabilité des décisions[17].

Enfin, dans la banque de détail, un système multi-agents pour l'évaluation de crédit : des agents spécialisés collectent les données, analysent le risque, vérifient la conformité des dossiers, proposent une décision et, si besoin, escaladent vers un humain. Là où la GenAI seule se limiterait à assister l'analyste dans la lecture et la synthèse de documents, les agents structurent le processus de bout en bout, automatisent les tâches répétitives et n'envoient aux humains que les cas ambigus ou à fort enjeu[18].

---

[15] Reynald Fléchaux. La banque d'affaires de BNP Paribas saupoudre ses offres produits avec de l'IA. *CIO*. 12 décembre 2025. https://www.cio-online.com/actualites/lire-la-banque-d-affaires-de-bnp-paribas-saupoudre-ses-offres-produits-avec-de-l-ia-16725.html.

[16] J.P. Morgan Payments. AI Agents: Your Next Co-Workers? Payments Unbound Vol. 4. *J.P. Morgan & Wired*. May 2024. https://www.jpmorgan.com/content/dam/jpmorgan/documents/payments/Payments-Unbound-Volume4.pdf

[17] Henrik Axelsen, Valdemar Licht, Jan Damsgaard. Agentic AI for Financial Crime Compliance. *arXiv preprint arXiv:2509.13137*. September 16, 2025. https://arxiv.org/pdf/2509.13137

[18] Subhadip Mitra. Engineering Multi-Agent Systems - A Retail Banking Case Study. *Blog Subhadip Mitra*. December 28, 2024. https://subhadipmitra.com/blog/2024/retail-bank-multi-agent-system/.

2.1.3. Industrie et logistique

Agents d'orchestration de la chaîne d'approvisionnement

Des agents experts IA sont déployés pour piloter de manière autonome les réseaux logistiques de bout en bout. Ces agents surveillent en continu les conditions opérationnelles, détectent les perturbations potentielles et déclenchent les actions correctives appropriées, souvent avant que l'impact ne se matérialise.

L'éditeur Blue Yonder propose un agent "Network Ops" capable de gérer les perturbations à l'échelle d'un réseau multipartenaires. Lorsqu'un fournisseur annonce un retard de livraison, l'agent recherche automatiquement un fournisseur alternatif, évalue l'option de reporter les livraisons clients, reprogramme les ressources d'entrepôt (déchargement, préparation, expédition) et arbitre entre coût, niveau de service et durabilité, le tout avec une intervention humaine minimale. Les équipes logistiques gardent le contrôle en définissant à l'avance les règles, objectifs et contraintes ; l'agent exécute selon ces paramètres[19].

Ces agents fonctionnent en continu, 24 heures sur 24, et s'améliorent grâce à l'apprentissage automatique en tirant les leçons des résolutions antérieures. Ils offrent des fonctionnalités telles que la prévision des heures d'arrivée sur des itinéraires multi-étapes, la consolidation intelligente des envois, le calcul des délais de livraison promis au client, ou encore l'anticipation des ruptures de stock avec proposition de solutions proactives[19].

Une étude menée par des chercheurs de Harvard, MIT et Georgia Tech a par ailleurs démontré qu'un système multi-agents basé sur des LLM pouvait gérer efficacement une simulation de chaîne d'approvisionnement (le "Beer Game"), en coordonnant les décisions d'approvisionnement entre plusieurs acteurs autonomes[20].

Agents de maintenance prédictive

Des agents experts IA sont déployés pour anticiper les défaillances des équipements industriels et déclencher de manière autonome les actions de maintenance appropriées. Ces agents ne se limitent pas à la prédiction : ils analysent en continu les données capteurs, détectent les dérives, évaluent l'urgence, planifient les interventions

---

[19] Agents d'IA pour la chaîne d'approvisionnement. BlueYonder. https://fr.blueyonder.com/why-blue-yonder/ai-and-machine-learning/ai-agents . Consulté le 19 décembre 2025.

[20] Carol Long, David Simchi-Levi, Andre P. Calmon, Flavio P. Calmon. When Supply Chains Become Autonomous. *Harvard Business Review*. December 11, 2025. https://hbr.org/2025/12/when-supply-chains-become-autonomous.

et peuvent automatiquement générer des ordres de travail ou commander des pièces de rechange[21].

2.1.4. Commerce et marketing

Agents de personnalisation dynamique de l'expérience client

Des agents experts IA analysent le comportement des clients en temps réel : navigation, historique d'achats, interactions cross-canal, l'objectif étant d'adapter dynamiquement les offres, les recommandations de produits et les parcours d'achat. Ces agents intègrent des données provenant de multiples sources (site web, application mobile, CRM, points de vente) et déclenchent des actions personnalisées : affichage d'une promotion ciblée, modification de l'ordre des produits affichés, envoi d'une notification push contextuelle.

L'objectif n'est plus seulement de générer du contenu marketing, mais d'orchestrer une expérience client individualisée où chaque interaction est optimisée en fonction du contexte et des objectifs commerciaux (conversion, fidélisation, panier moyen) [22 & 23].

Agents pour la vente et l'achat

Deux types d'agents émergent dans le commerce en ligne. D'un côté, des agents acheteurs permettent aux consommateurs d'automatiser leurs achats : surveillance de la disponibilité d'un produit, comparaison de prix entre plateformes, achat automatique dès qu'un critère est rempli (prix cible atteint, produit en stock). Ces agents agissent pour le compte de l'utilisateur sans intervention manuelle.

De l'autre côté, des agents vendeurs sont déployés par les entreprises pour gérer de manière autonome des interactions commerciales complexes. Au-delà du simple chatbot de support, ces agents peuvent découvrir les besoins du client par le dialogue, personnaliser les offres en temps réel, négocier les conditions (remises, options) et accompagner jusqu'à la conclusion de la vente. Ils opèrent sur divers canaux — site web, messageries, voix — et s'intègrent aux systèmes de gestion commerciale pour accéder aux stocks, tarifs et historiques clients.

---

[21] Navdeep Singh Gill. Agentic AI for Predictive Maintenance: Prevent Downtime & Cut Costs. *NexaStack*. August 14, 2025. https://www.nexastack.ai/blog/predictive-maintenance-agentic-ai. Consulté le 19 décembre 2025.

[22] Charlie Mitchell. 25 Use Cases for Generative AI In Customer Service. *CX Today*. August 28, 2024. https://www.cxtoday.com/contact-center/20-use-cases-for-generative-ai-in-customer-service/

[23] Monika Lončarić. The ultimate guide to generative AI chatbots for customer service. *Infobip*. https://www.infobip.com/blog/generative-ai-for-customer-service

À titre d'exemple, prenons Growify. Il se positionne explicitement sur "*Agentic Commerce*" avec un agent de vente qui engage proactivement les visiteurs, recommande des produits, répond aux objections, propose des remises et relance par SMS/email. Par rapport à un chatbot génératif qui répond aux questions, l'agent se déclenche sur des signaux (temps passé, abandon de panier), enchaîne plusieurs actions (conversation, recommandation, remise, relance) et gère un cycle de vente bout-en-bout, ce qui en fait un acteur autonome du tunnel de conversion plutôt qu'un simple outil de réponse[24].

Cette évolution s'inscrit dans la tendance plus large des agents capables de piloter des interfaces numériques de manière autonome, à l'instar des solutions *Computer Use* (Anthropic) ou *Operator* (OpenAI), qui préfigurent des agents capables d'effectuer des transactions complètes sans supervision humaine. Cette tendance est elle-même amplifiée par la vague des navigateurs dopés à l'IA (Google Chrome, Microsoft Edge, Atlas d'OpenAI, Comet de Perplexity, et même Firefox[25]).

2.1.5. Éducation et formation

Agents d'accompagnement pédagogique adaptatif

Les agents experts IA en éducation s'inscrivent dans la lignée des systèmes tuteurs intelligents développés depuis les années 1970. L'apport des LLM réside dans leur capacité à fournir un accompagnement personnalisé, disponible en permanence, et adapté au contexte spécifique de chaque cours.

À l'Université du Michigan, un assistant IA a été déployé pour un cours de programmation de systèmes distribués. Entraîné sur les supports de cours et les discussions des forums des années précédentes, l'agent répond aux questions conceptuelles et de spécification des étudiants à tout moment[26].

À l'Université George Washington, un prototype d'assistant pédagogique IA analyse les travaux des étudiants pour détecter des patterns récurrents (mots suremployés, répétitions) et croise automatiquement plusieurs documents du cours pour construire des réponses complètes. Des garde-fous garantissent que l'agent ne répond qu'à partir des contenus téléchargés par l'enseignante[26]. Ces déploiements confirment que les

[24] Agentic Commerce: How Growify's AI Sales Agent Is Revolutionizing Ecommerce. *Growify*. https://growify.ai/agentic-commerce/. Consulté le 19 décembre 2025
[25] Ajit Varma. Introducing AI, the Firefox Way: A Look at What We're Working on and How You Can Help Shape It. *The Mozilla Blog*. November 13, 2025. https://blog.mozilla.org/en/firefox/ai-window/.
[26] Erin Brereton. AI-Powered Teaching Assistants Can Drive Student Success. *EdTech*. March 24, 2025. https://edtechmagazine.com/higher/article/2025/03/ai-powered-teaching-assistants-perfcon.

agents pédagogiques ne remplacent pas les enseignants, mais permettent aux étudiants de faire un usage plus efficace du temps passé avec eux.

Ces agents analysent en continu les réponses, les hésitations et les progrès de l'étudiant pour ajuster dynamiquement le rythme, le niveau de difficulté et le style pédagogique. Contrairement à un LLM généraliste qui répond à des questions ponctuelles, l'agent pédagogique maintient un modèle de l'apprenant qu'il enrichit au fil des interactions, identifie les lacunes récurrentes et propose des exercices de remédiation ciblés[27].

L'évaluation de la qualité du tutorat constitue un enjeu majeur. Des systèmes multi-agents comme TRAVER et DICT abordent cette problématique : un agent tuteur interagit avec l'apprenant tandis qu'un agent évaluateur analyse la pertinence des interventions pédagogiques et suggère des améliorations. Cette architecture permet une amélioration continue du système sans supervision humaine constante[28].

2.1.6. Administration publique et collectivités

Agents d'automatisation des processus administratifs

Des agents experts IA sont déployés pour automatiser des workflows administratifs de bout en bout. Dans le secteur du logement social, des agents orchestrent le traitement des attestations d'assurance et des diagnostics techniques : vérification de la conformité des documents reçus, mise à jour de la Gestion Électronique de Documents (GED) et du système ERP, création des tickets de suivi et déclenchement de l'envoi d'emails aux parties prenantes. Cette automatisation assure une conformité fluide et centralisée, avec intervention humaine limitée aux cas d'exception[29].

À terme, ces agents pourraient également jouer un rôle de médiation entre citoyens et administrations, facilitant l'accès aux services et recueillant les retours d'expérience pour améliorer les politiques publiques[30].

---

[27] Subhankar Maity, Aniket Deroy. Generative AI and its impact on personalized intelligent tutoring systems. *arXiv preprint arXiv:2410.10650*. October 2024. https://arxiv.org/abs/2410.10650

[28] Jian Wang et al. Training Turn-by-Turn Verifiers for Dialogue Tutoring Agents: The Curious Case of LLMs as Your Coding Tutors. *arXiv:2502.13311, arXiv*, May 25, 2025. https://arxiv.org/pdf/2502.13311

[29] Georges Acar, Annabelle Luisetti. Réinventer l'efficacité opérationnelle grâce à l'intelligence artificielle. *La Jaune et la Rouge*. Magazine N°805. Mai 2025. https://www.lajauneetlarouge.com/reinventer-l-efficacite-operationnelle-grace-a-lintelligence-artificielle/

[30] Damien Bruce et al. Unlocking the potential of generative AI: Three key questions for government agencies. *McKinsey & Company*. December 7, 2023. https://www.mckinsey.com/industries/public-sector/our-insights/unlocking-the-potential-of-generative-ai-three-key-questions-for-government-agencies

D'autres cas montrent des agents IA qui automatisent la prise en compte des demandes entrantes (prestations sociales, licences, subventions) : analyse des pièces, contrôle de complétude, premiers contrôles de conformité et routage vers le bon service, avec des temps de traitement pouvant être divisés par deux. Par rapport à une GenAI qui aiderait à "remplir un formulaire", l'agent suit le dossier dans le temps, décide des étapes suivantes (demander un justificatif, transmettre à un instructeur, clôturer) et journalise ses actions, ce qui en fait un véritable gestionnaire de cas à faible complexité[31].

2.2. Cas d'usage par fonction métier

2.2.1. Création de contenu et communication

Une agence marketing ("Be Known") a mis en place un agent de création de contenu qui, à partir d'un simple formulaire, génère plus de 15 types de contenus : pages de destination, séquences d'emails, scripts de webinar, posts réseaux sociaux, puis classe automatiquement les fichiers et met à jour la tâche dans l'outil de gestion de projet. Par rapport à une GenAI qui rédige un texte à la demande, l'agent pilote l'ensemble du flux : validation des entrées, récupération du contexte client (en interne et/ou sur le Web), génération multiformats, organisation des livrables, ce qui a fait passer la production d'un asset de 4-6 heures à moins de 10 minutes.[32]

Pour les professionnels du secteur médiatique, des solutions innovantes offrent une interface intuitive pour la création vidéo. Le processus se déroule en plusieurs phases :

- **scénarisation** : un agent scénariste interroge le client sur l'histoire, les personnages et l'ambiance générale pour élaborer un script modifiable ;
- **direction artistique** : un agent génère des visuels initiaux en adéquation avec les préférences exprimées. Ces visuels sont ensuite animés par un autre agent ;
- **audio** : un agent assiste le client dans le choix d'une voix off et d'un texte, tandis qu'un compositeur virtuel propose une musique d'ambiance appropriée.
- **intégration finale** : un agent final assemble l'ensemble des éléments et permet au client d'ajuster les détails avant la finalisation du projet.

---

[31] Bernhard Huber. AI Agents in Public Administration: Automating Processes & Citizen Services. *Primotly*. June 13, 2025. https://primotly.com/article/ai-agents-in-public-administration-automating-processes-and-improving-citizen-services.

[32] AI Agent That Took Over Content Creation and Boosted Capacity. *Atomic Actions*. https://www.atomicactions.io/case-study/ai-agent-that-took-over-content-generation. Consulté le 19 décembre 2025.

De nombreuses entreprises offrent ce type de service. Parmi elles, on peut citer l'entreprise française *TheNextStories*[33], qui met à la disposition de ses utilisateurs une équipe de réalisation entièrement composée d'agents IA générative. Ces agents exploitent les modèles les plus performants dans leurs domaines respectifs : texte, son, images, narration, musique, vidéo et animation.

2.2.2. Achats et Approvisionnements

Des agents experts IA accompagnent les acheteurs tout au long du cycle d'approvisionnement. En amont, ils assistent à l'élaboration du cahier des charges en suggérant des critères pertinents basés sur l'historique des achats similaires et les standards du marché.

C'est dans l'analyse des réponses aux appels d'offres que ces agents déploient leur pleine valeur ajoutée. Ils examinent automatiquement l'ensemble des propositions reçues, identifient les écarts par rapport au cahier des charges, détectent les incohérences internes (prix unitaires vs prix globaux, délais contradictoires) et évaluent les risques potentiels (solidité financière du fournisseur, dépendances critiques). L'agent génère des grilles d'analyse comparatives et des rapports de synthèse, permettant à l'acheteur de se concentrer sur la décision finale plutôt que sur le dépouillement.

Ainsi, l'éditeur français Ivalua propose un assistant virtuel intelligent intégré à sa plateforme achats, capable d'accompagner les utilisateurs dans leurs tâches quotidiennes et d'automatiser les workflows de validation[34].

De même, des agents dédiés au sourcing surveillent la performance fournisseurs (délais, qualité, risques), détectent des dérives, suggèrent des fournisseurs alternatifs et préparent ou déclenchent des actions (renégociation, redistribution de volumes, ajout d'une seconde source). L'apport agentique, par rapport à un simple LLM qui rédigerait une analyse, est que l'agent parcourt plusieurs systèmes, corrèle les données, propose des plans d'action et peut exécuter les workflows de changement de fournisseur dans un cadre de règles, améliorant la résilience sans micro-pilotage humain constant.[35]

---

[33] https://thenextstories.com/

[34] L'IA générative pour les achats. *Ivalua*. https://fr.ivalua.com/technologie/la-plateforme-dachats-divalua/ivaintelligent-virtual-assistant/. Consulté le 19 décembre 2025.

[35] Supply chain sourcing. *AI Agents Agency*. August 31, 2025, https://www.aiagentsagency.ca/case-studies/supply-chain-sourcing/.

2.2.3. Service client et support

Des systèmes multi-agents sont déployés pour gérer les interactions client de bout en bout. Un agent principal analyse la requête entrante, sollicite des clarifications si nécessaire, puis oriente la demande vers des agents spécialisés. Ces derniers collectent les documents requis, analysent les données pertinentes et exécutent les actions appropriées : mise à jour des systèmes CRM et ERP, gestion des commandes, génération de courriels professionnels. L'ensemble fonctionne avec une intervention humaine minimale, réservée aux cas d'exception[36 & 37].

Cette architecture modulaire peut être adaptée aux exigences spécifiques de chaque métier et transposée à des agents internes au bénéfice des collaborateurs : support RH, helpdesk informatique, gestion logistique.

Le Bon Coin, plateforme française de petites annonces, a mis en place une équipe d'agents experts IA coordonnés par un agent maître nommé Markus. Celui-ci assure la cohérence des échanges et rédige les premières réponses aux tickets utilisateurs en s'appuyant sur une base documentaire interne (accessible via un agent dédié) et sur les données utilisateur (fournies par un agent spécialisé). Cette structure permet de pallier les limitations d'un chatbot unique, notamment le manque d'adaptabilité contextuelle et le risque d'hallucinations[38].

Ping An Insurance (Chine) exploite une plateforme de service client basée sur l'IA depuis 2019. Fin 2022, 49 % des ventes de produits étaient réalisées par des représentants assistés par IA et 82 % des interactions de service étaient gérées par des agents logiciels. Ces déploiements ont généré des économies de 600 millions de RMB en coûts de main-d'œuvre, tout en permettant à l'entreprise de fournir ses services IA à trente autres établissements bancaires chinois[39].

---

[36] Jeanne Bigot. L'IA générative au service du client, du conseiller et des ressources humaines. *Les Echos*. 30 septembre 2024. https://www.lesechos.fr/thema/articles/lia-generative-au-service-du-client-du-conseiller-et-des-ressources-humaines-2122022

[37] Rodrigo Rodrigues Pires de Mello, Thiago Ângelo Gelaim, Ricardo Azambuja Silveira. Negotiation strategies in multi-agent systems for meeting scheduling. *2018 XLIV Latin American Computer Conference (CLEI)*. IEEE, p. 242-250. 2018. https://clei.org/clei2018/docs/SLIOIA/182509.pdf

[38] Charles Fandre. leboncoin : comment la plus grande marketplace française s'est transformée grâce à l'IA. *Media Thiga*. 22 septembre 2025. https://www.media.thiga.co/leboncoin-comment-la-plus-grande-marketplace-francaise-sest-transformee-grace-ia

[39] AI Banker: Reshaping Retail Banking for the Digital Age. *Ping An Group*, August 26, 2019, https://group.pingan.com/media/perspectives/AI-Banker-Reshaping-Retail-Banking-for-the-Digital-Age.html.

2.2.4. Développement logiciel

Équipes de développement multi-agents

Des systèmes multi-agents reproduisent la distribution des tâches d'une équipe de développeurs humains. Plusieurs agents spécialisés collaborent en synergie (et cohérence) :

- un **agent de génération** produit des blocs de code à partir de spécifications détaillées ;
- un **agent de tests** vérifie la fiabilité et la robustesse du code généré ;
- un **agent QA** contrôle la conformité aux standards et identifie les anomalies ;
- un **agent de correction** optimise le code ou corrige les bogues détectés.

Cette boucle autonome (génération, test, détection, correction) permet d'automatiser et d'accélérer significativement le cycle de développement, notamment pour les projets complexes ou répétitifs[40].

Donc des frameworks dédiés (comme MetaGPT[41]) déploient une "équipe" d'agents spécialisés (chef de projet, architecte, développeur, reviewer) qui se répartissent la conception, l'écriture de code et la revue. Par rapport à un unique LLM de complétion voire de génération de code, l'approche agentique apporte la division du travail, la coordination (via un agent "manager") et la capacité à itérer sur plusieurs versions d'une même fonctionnalité, ce qui rapproche le système d'un vrai workflow d'équipe plutôt que d'un outil individuel de génération ou d'auto-complétion.

Selon une étude du MIT Sloan Management Review, 69 % des experts interrogés considèrent les agents IA autonomes comme un changement de paradigme dans la gestion des tâches complexes, en raison de leur capacité à exécuter des tâches à grande échelle avec une efficacité supérieure à celle des équipes traditionnelles[42].

Frameworks et outils émergents

---

[40] Jian Yang et al. From Code Foundation Models to Agents and Applications: A Comprehensive Survey and Practical Guide to Code Intelligence. *arXiv preprint arXiv:2511.18538.* 303 p. December 2025. https://arxiv.org/pdf/2511.18538.

[41] Foundation Agents/MetaGPT. *GitHub*. December 20, 2025. https://github.com/FoundationAgents/MetaGPT.

[42] Elizabeth M. Renieris, David Kiron, Steven Mills, Anne Kleppe. Agentic AI at Scale: Redefining Management for a Superhuman Workforce. *MIT Sloan Management Review*. September 16, 2025. https://sloanreview.mit.edu/article/agentic-ai-at-scale-redefining-management-for-a-superhuman-workforce/

Depuis 2023, les efforts de recherche se concentrent sur le développement d'agents autonomes capables d'exécuter des tâches complexes au sein de projets logiciels réels. Le framework CodeCoR, par exemple, met en œuvre une architecture multi-agent autoréflexive : les agents évaluent mutuellement leurs productions et itèrent jusqu'à convergence vers un code de qualité[43].

Des outils commerciaux comme GitHub Copilot Workspace, Cursor ou Claude Code intègrent progressivement ces approches agentiques, permettant aux développeurs de déléguer des tâches complètes (implémentation d'une fonctionnalité, correction d'un bug, refactoring) plutôt que de simples complétions de code.

Enfin, citons le cas des solutions à base d'agents qui assurent des tâches de fond comme piloter la maintenance continue d'un dépôt, appliquer des corrections chaque nuit, supprimer le code mort chaque mois, mettre à jour des dépendances et réparer des tests de manière récurrente, sur la base de règles de l'équipe. Voire même de prendre en charge des traductions de base de code d'un langage à l'autre[44]. Ici, l'apport agentique est la proactivité et la gestion du temps : l'agent ne se contente pas de proposer une refactorisation sur demande, il surveille le dépôt, planifie et exécute des tâches répétitives en continu, avec revue humaine ciblée, ce qui maintient la base de code saine sans mobilisation permanente des développeurs.

2.2.5. Agents experts IA pour les RH

Les départements des ressources humaines s'efforcent d'automatiser les processus de recrutement, compte tenu du volume important de candidatures, d'intégrer les nouvelles recrues de manière dynamique (*onboarding*) et d'optimiser la gestion des talents afin d'assurer un accompagnement plus efficace des collaborateurs dans leur parcours professionnel, tout en personnalisant l'expérience collaborateur.

Des agents de recrutement autonomes[45] (screening, présélection, planification des entretiens) scannent les CV, mènent des entretiens structurés, évaluent les compétences sur la base de grilles prédéfinies et proposent une liste réduite de candidats aux recruteurs. Par rapport à un LLM seul qui rédige une annonce, l'agent gère

[43] Ruwei Pan, Hongyu Zhang, Chao Liu. CodeCoR: An LLM-Based Self-Reflective Multi-Agent Framework for Code Generation. *arXiv preprint arXiv:2501.07811*. January 14, 2025. https://arxiv.org/pdf/2501.07811.
[44] Devonair. AI for Codebase Maintenance: Automate the Work Nobody Wants to Do. *Devonair*. October 27, 2025. https://devonair.ai/blog/use-cases/ai-codebase-maintenance.
[45] Bryan Peereboom. AI Friday powered by RecruitAgent.ai - How Autonomous Agents Are up and Coming for HR & Recruitment. *ToTalent*. September 27, 2024, https://totalent.eu/ai-friday-powered-by-recruitagent-ai-how-autonomous-agents-are-up-and-coming-for-hr-recruitment/.

toute la séquence : *sourcing, scoring*, coordination des entretiens, suivi candidat, avec des règles d'équité et de non-discrimination intégrées et auditables[46].

Il est à noter que la reconnaissance des émotions (proposées dans certaines plateformes agentiques – cf. celle mentionnée ci-dessus chez KPMG) n'est pas acceptable au titre de *l'AI Act* européen[47].

[46] Autonomous Recruitment. *KPMG*. November 27, 2025, https://kpmg.com/ng/en/home/ai-resource-hub/Autonomous%20Recruitment%20Agent.html.

[47] Alice Vitard. Reconnaissance des émotions : La Cnil menace de sanctions en vertu de l'AI Act. *L'Usine Digitale*. 11 avril 2025. https://www.usine-digitale.fr//article/reconnaissance-des-emotions-la-cnil-menace-de-sanctions-en-vertu-de-l-ai-act.N2230488

HUB FRANCE IA

3. Typologie des agents experts d'IA générative

3. Typologie des agents experts IA générative

3.1. Définition et cadre de référence

Un agent expert IA est un programme informatique capable de percevoir son environnement par le biais de capteurs, d'*APIs* et de flux de données[48], et d'agir sur cet environnement à l'aide d'actuateurs logiciels, de commandes, d'écritures en base de données et d'appels d'API. Cette boucle perception-action, formalisée par Russell et Norvig[49], constitue le fondement de l'évaluation de tout agent, notamment en termes de réactivité, d'adaptation, de sûreté et de traçabilité.

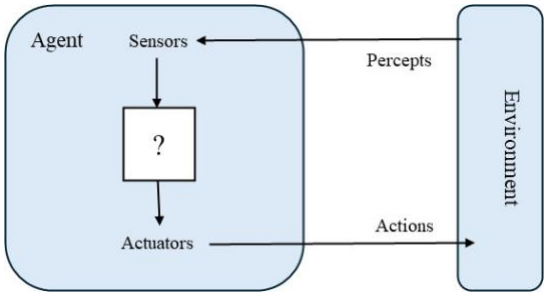

Figure 1 – Les agents interagissent avec l'environnement à travers capteurs et actionneurs. D'après[49].

Un agent possède une mémoire, une base de connaissance plus ou moins grande, ainsi qu'un nombre d'actions et de stratégies variables en fonction de sa nature pour atteindre un objectif personnel et parfois un objectif collectif.

La littérature[50] distingue différentes familles d'agents en fonction de leur niveau de sophistication. On peut regrouper ces types d'agents en trois grandes familles :

---

[48] Barbara Pazur, Dina Sostarec, Ishita Jaiswal. 13 Types of AI Agents (with Examples). *Multimodal*. November 25, 2024. https://www.multimodal.dev/post/13-types-of-ai-agents. Outre le rôle d'une vision externe, caméras, LiDAR, flux API ou données multimodales offrent à l'agent une compréhension affinée de son contexte, qu'il s'agisse de piloter un véhicule autonome ou d'interpréter des images médicales. Cette perception enrichie s'accompagne de questions éthiques sur la vie privée et les biais, qui seront abordées plus loin.

[49] Stuart Russell, Peter Norvig. Artificial Intelligence: A Modern Approach. 4th ed. *Pearson Education Limited*. 2016. https://aima.cs.berkeley.edu/4th-ed/pdfs/newchap02.pdf

[50] Emmanuel Adam. Multi-Agent Systems. *College SMAA*. November 2025. https://www.college-smaa.fr/collegesmaa/

Sarah Chudleigh. 7 Main Types of AI Agents [with Examples]. Botpress. June 2025. https://botpress.com/blog/types-of-ai-agents Consulté le 3 octobre 2025.

- **l'agent réflexe** (réactifs) : il possède un nombre limité d'actions, ainsi qu'un niveau limité de détail du contexte qui l'entoure. Ses décisions sur l'action à activer selon ses perceptions du contexte sont donc rapides et souvent prédéfinies. Un exemple concret serait celui d'un robot aspirateur ;
- **l'agent cognitif** (délibératif) : il possède une représentation du monde plus complexe, ainsi que des ensembles d'actions, regroupées en stratégies à activer selon ses perceptions et l'utilité attendue de celles-ci. Suite à l'application d'une stratégie, une réévaluation de l'utilité est effectuée, ce qui permet à l'agent d'être résilient, de s'adapter aux changements. Un système de navigation calculant l'itinéraire optimal et s'adaptant à la réalité du trafic constitue un exemple pertinent. De même un agent contrôlant une navette autonome dans un atelier de fabrication est un agent de ce type ;
- **l'agent avatar** : il est en interaction avec une personne et ajuste son comportement en fonction de l'expérience acquise de ses échanges. Un agent de recommandations qui intègre les actions de l'utilisateur en est une illustration.

Dans le cadre de ce livre blanc, nous adoptons une approche de classification orientée usages, plus pragmatique et applicable aux entreprises. Cette typologie permet d'identifier un agent (cf. les cas d'usage) afin de le "cartographier". En principe, un agent, en raison de sa nature spécialisée, appartient à un seul type.

En outre, chaque agent possède des attributs spécifiques, en sus de sa classification. Ces attributs sont détaillés dans la seconde partie de ce chapitre. Il est donc possible que deux agents partagent certains attributs tout en appartenant à des classifications distinctes, à l'instar d'un chien et d'une fourmi qui présentent des caractéristiques communes (être vivant, posséder une tête et des pattes, etc.) tout en étant de "classifications" différentes. Inversement, deux agents appartenant à la même classification n'ont pas nécessairement des attributs identiques.

3.2. Types d'agents experts IA

Cette liste des principales catégories d'agents n'est pas exhaustive et est susceptible d'évoluer parallèlement à l'adoption et au développement de fonctionnalités agentiques. Son objectif est de fournir un cadre de référence pour une meilleure compréhension des grandes familles d'agents, sans prétendre à l'exhaustivité ni à une taxonomie rigide.

Dans un premier temps, nous examinons les agents **d'interaction et de communication**. Ces derniers, qu'il s'agisse de *chatbots*, d'assistants vocaux ou de sentinelles en temps

réel, sont conçus pour dialoguer, guider ou alerter. Leur efficacité repose sur un équilibre subtil entre la réactivité automatique et la supervision humaine.

Ensuite, les agents de **génération et de traitement de contenus** sont présentés dans le contexte de leur capacité à manipuler divers formats de données, notamment le texte, l'image et le son. Leur polyvalence, illustrée par des tâches telles que l'extraction de contenu à partir de fichiers PDF et la synthèse vidéo, est remarquable. Cependant, cette polyvalence soulève également des questions cruciales relatives aux droits d'auteur et à la fiabilité des informations générées.

Troisièmement, les agents **d'automatisation et de raisonnement** jouent un rôle crucial. La planification constitue un élément fondamental de leur fonctionnement : elle permet d'orchestrer l'intervention de multiples services, d'optimiser les itinéraires et de sélectionner le compromis optimal entre coûts et risques. Ces agents transforment ainsi un simple flux de travail en un processus adaptatif et réactif.

La quatrième famille d'agents se caractérise par une **spécialisation par domaine**, englobant des profils tels que les développeurs virtuels, les simulateurs logistiques et les robots d'investigation de données. Grâce à un entraînement ciblé, des interfaces de programmation d'applications (*API*) dédiées et des protocoles métier spécifiques, ces agents se distinguent par leur expertise approfondie dans des environnements souvent soumis à une réglementation stricte, agissant ainsi comme des « experts de poche » au sein de leurs domaines respectifs.

Les **systèmes multi-agents** offrent une perspective élargie, permettant à plusieurs agents de collaborer, de se contrôler mutuellement ou de se compléter. Cette approche distribuée confère robustesse et redondance, tout en augmentant la surface de risques.

Dans les pages qui suivent, chaque famille sera minutieusement analysée : rôle principal, cas d'usage, niveau d'autonomie, forces et limites, exemples concrets et caractéristiques distinctives. Cette analyse vous permettra de contextualiser tout agent rencontré, qu'il soit actuellement actif ou susceptible de l'être à l'avenir, avant de le confronter aux caractéristiques détaillées présentées dans la section suivante.

3.2.1. Agents d'interaction et de communication

Cette famille regroupe des agents et assistants centrés sur l'interaction et la communication. Leur autonomie est très variable : certains restent principalement réactifs (interaction en langage naturel), tandis que d'autres deviennent agentiques lorsqu'ils sont outillés pour surveiller des systèmes, déclencher des alertes et initier des actions encadrées.

Assistants conversationnels et d'assistance

Ces assistants interagissent en langage naturel afin de fournir des réponses, une assistance ou un support. Ils peuvent être déployés dans divers contextes, tels que le service client, le support informatique ou les interfaces de dialogue, afin de guider l'utilisateur dans la réalisation de tâches spécifiques. L'exploitation de modèles de langage sophistiqués permet de simuler une conversation fluide et naturelle. Leur autonomie est généralement réactive et la supervision humaine demeure indispensable, en particulier pour prévenir les hallucinations ou les interprétations erronées.

- **focus** : interaction en langage naturel, support client, support utilisateur ;
- **exemples** : *chatbot* avec *feedback* (Projet Inria[51]), Copilotes RH ;
- **caractéristiques typiques** : collaboration, interaction avec l'environnement, souvent basées sur des systèmes de *Retrieval Augmented Generation* (*RAG*[52]) et des bases documentaires.

Agents de supervision en temps réel

Ces agents garantissent une surveillance ininterrompue et une intervention immédiate dans des contextes critiques. Ils traitent des flux de données en direct afin de détecter, analyser et réagir à des événements en temps réel, notamment pour des applications de sécurité, de monitoring système ou de gestion de crises techniques. Leur rapidité et leur fiabilité sont essentielles pour assurer une réponse efficace aux situations d'urgence, avec ou sans intervention humaine.

- ***focus*** : surveillance en temps réel et réponses rapides dans des situations critiques telles que les alertes de sécurité ou la surveillance des systèmes ;

[51] Comment interagir au mieux avec les chatbots ? Inria. 28 août 2024. https://www.inria.fr/fr/ia-agents-conversationnels-chatbots-ihm

[52] Le *RAG* constitue une technique d'optimisation des réponses générées par les modèles de langage en IA. Cette méthodologie permet notamment d'améliorer significativement la qualité des réponses aux requêtes en permettant aux *LLMs* d'exploiter des ressources de données additionnelles sans nécessiter de réentraînement. Voir : Hub France IA. Guide Pratique Chaînes de *RAG.* Septembre 2025. https://www.hub-franceia.fr/wp-content/uploads/2024/02/Livre-blanc_Les-usages-de-lia-generative-01.2024.pdf

- ***exemples*** : alerte lors de l'analyse automatique des journaux d'un système informatique (ITops[53]), détection de fraude en analyse bancaire, analyse cyber temps réel[54] ;
- ***caractéristiques typiques*** : interaction avec l'environnement, mémoire, scalabilité et haute disponibilité.

Agents pédagogiques

Ces agents sont conçus pour accompagner les individus dans l'acquisition de nouvelles compétences ou l'approfondissement de leurs connaissances sur un sujet donné. Ces agents possèdent la capacité d'adapter leur approche en fonction du public qui les sollicite. Ils analysent le niveau de compétence, le contexte culturel et la progression des utilisateurs afin de leur fournir un accompagnement personnalisé et optimal dans leur démarche de formation. Leur champ d'application ne se limite pas à la formation traditionnelle, mais s'étend à divers contextes applicatifs, où ils contribuent à l'amélioration de l'efficacité et de la productivité des utilisateurs.

- ***exemples*** : agents tutoriels (cf. les cas d'usage mentionnés plus haut), copilotes métier interactifs, explication en langage naturel avec des exemples ;
- ***caractéristiques typiques*** : collaboration, apprentissage continu, mémoire, mobilité.

3.2.2. Agents de génération et traitement de contenus

La famille des agents spécialisés dans la création ou la transformation de contenus au sein d'une ou plusieurs modalités (texte, image, audio, vidéo, etc.).

Agents de traitement de contenu multimodal

Ces agents sont spécialisés dans le traitement et la génération de formats multiples, incluant le texte, l'image, l'audio et la vidéo, souvent en les intégrant de manière intermodale. Ils excellent dans l'analyse et la génération de divers types de contenus, tels que le texte, les images, l'audio et la vidéo, et possèdent la capacité de fusionner des informations issues de différentes modalités. Par exemple, ils peuvent extraire du texte à partir d'images grâce à la reconnaissance optique de caractères (*OCR*), générer des visuels à partir de descriptions textuelles, ou encore résumer des vidéos. Leur force réside

[53] Mesh Flinders, Ian Smalley. Qu'est-ce que l'analyse de journaux optimisée par l'IA ? *IBM*. 27 mai 2025. https://www.ibm.com/fr-fr/think/topics/ai-for-log-analysis

[54] Jesse Kimbrel. Agents IA : Que signifient-ils pour la cybersécurité ? *Vectra*. 28 avril 2025. https://fr.vectra.ai/blog/ai-agents-what-do-they-mean-in-cybersecurity

dans l'intégration de données hétérogènes afin de produire une réponse cohérente et riche en informations.

- ***focus*** : extraction, analyses, transformation et génération multimodalités (texte, images, audio, vidéo).
- ***exemples*** : *OCR* et résumés vidéos.
- ***caractéristiques typiques*** : multimodalité, créativité.

Agents créatifs

Les agents de création de contenus originaux sont des agents professionnels spécialisés dans la production de contenus narratifs, marketing ou artistiques. Ils sont experts dans la conception de textes, d'images, de vidéos, de musiques ou d'autres formes d'expression artistique. Leur expertise est largement sollicitée dans les domaines du marketing, de la publicité et de la production médiatique afin de générer des slogans, des récits ou des visuels impactants et pertinents pour un public cible, en conjuguant créativité et pertinence contextuelle.

- ***focus*** : création de contenus originaux, de contenus pour des applications marketing, media ou artistiques ;
- ***exemples*** : suite d'agents de génération de clips vidéo[55] de The Next Stories, agents de génération automatisés de publicité[56] ;
- ***caractéristiques typiques*** : multimodalité, créativité.

Agents documentaires et d'analyse de données

Les rédacteurs, metteurs en forme et convertisseurs de documents, opérant fréquemment dans un environnement professionnel, se consacrent à l'extraction, la transformation et l'analyse de données structurées ou non structurées. Ils recourent à des techniques avancées de « *data mining* », de classification, d'analyse sémantique, d'apprentissage automatique et de statistiques afin de convertir d'importants volumes de données et de documents en informations exploitables, contribuant ainsi à l'optimisation de la prise de décision basée sur l'analyse des données de l'entreprise.

[55] http://thenextstories.com

[56] Amazon. Amazon Ads lance un nouvel outil créatif d'IA autonome. *Amazon Ads*. 17 septembre 2025. https://advertising.amazon.com/library/news/amazon-ads-agentic-ai-creative-tool. Consulté le 3 octobre 2025.

- ***focus*** : extraire, transformer et analyser des données structurées et non structurées pour obtenir des insights, selon le protocole *Modern Context Protocol (MCP*[57]) ;
- ***exemples*** : agents juridiques[58], rédaction technique ;
- ***caractéristiques typiques*** : capacité de raisonnement numérique pour l'analyse de données.

3.2.3. Agents d'automatisation et de raisonnement

Agents dotés de la capacité à planifier, coordonner, déléguer ou raisonner afin d'automatiser des processus complexes.

Agents d'orchestration

Ces agents orchestreront la coordination de multiples agents ou systèmes au sein d'un workflow structuré. Ils décomposeront des tâches complexes en sous-tâches spécialisées et superviseront leur exécution au sein d'un workflow rigoureusement défini. La coordination de plusieurs agents spécialisés (par exemple, un agent conversationnel et un agent de traitement de contenu multimodal) sera assurée afin d'optimiser l'efficacité opérationnelle. Leur rôle est primordial dans l'automatisation des processus métiers, garantissant ainsi la cohérence et la synchronisation des différentes étapes.

- ***focus*** : coordonner, superviser ;
- ***exemples*** : workflow RH automatisé, orchestrateur de service client pour Le Bon Coin (voir [38] ci-dessus) ;
- ***caractéristiques typiques*** : planification, prise de décision, mémoire.

Agents délibératifs

Les agents délibératifs se distinguent par leur capacité à prendre des décisions éclairées en fonction de leurs objectifs, de leurs connaissances et de leurs croyances. Contrairement aux agents réactifs, qui se limitent à des réponses immédiates aux perceptions de leur environnement, les agents délibératifs intègrent une représentation interne de celui-ci, qu'il s'agisse d'un contexte plus ou moins large, d'une carte ou d'une base de prédictions. Cette représentation interne leur permet d'évaluer les différentes

[57] Le *MCP (Model Context Protocol)* est un protocole ouvert qui standardise la manière dont les applications fournissent du contexte aux modèles d'IA générative. Il fournit un moyen standardisé de les connecter à différentes sources de données et outils. Pour aller plus loin : Supabase Docs. Model context protocol (MCP). https://supabase.com/docs/guides/getting-started/mcp

[58] Un Agent qui assiste dans la recherche, l'analyse et la rédaction de documents juridiques. https://www.jimini.ai

options d'action, d'anticiper les résultats potentiels, de planifier et de sélectionner les actions les plus susceptibles d'atteindre leurs objectifs de manière optimale.

- ***focus*** : modèle du monde, objectifs et intentions, planification, raisonnement et adaptabilité ;
- ***exemples*** : modèles de raisonnements, agents réflexifs[59] ;
- ***caractéristiques typiques*** : prise de décision, mémoire, intentionnalité.

Agents utilitaires

Les agents utilitaires prennent des décisions éclairées en s'appuyant sur une fonction d'utilité ou un système de préférences clairement définis. Ils procèdent à une évaluation rigoureuse de multiples options possibles, les confrontent selon des critères prédéterminés (tels que le coût, la performance, le risque, la satisfaction, etc.) et sélectionnent l'action optimale. Ce type d'agent s'avère particulièrement pertinent dans les contextes nécessitant des arbitrages complexes ou la formulation de recommandations personnalisées. Ils peuvent être intégrés à des systèmes de recommandation, de tarification dynamique, de sélection de scénarios ou d'optimisation logistique, et sont fréquemment associés à des capacités de simulation ou de pondération multicritères.

- ***focus*** : classification, quantification, prise de décision, ajustement/apprentissage ;
- ***exemples*** : agents de *scoring*[60] (banque, finance, assurance), agents de tri (modèle *genAI* de la classification) ;
- ***caractéristiques typiques*** : prise de décision, mémoire, intentionnalité, capacité de raisonnement numérique.

3.2.4. Agents spécialisés par domaine

Il s'agit de la famille des agents spécialisés dans l'exécution de tâches professionnelles spécifiques, souvent dotés de modèles pré-entraînés ou de protocoles définis.

[59] Mohamed Bal-Ghaoui, Fayssal Sabri. LLM-FS-Agent: A Deliberative Role-based Large Language Model Architecture for Transparent Feature Selection. *arXiv preprint arXiv:2510.05935*. October 2025. https://arxiv.org/pdf/2510.05935

[60] Gautam Jajoo, Pranjal A. Chitale, Saksham Agarwal. MASCA: LLM based-Multi Agents System for Credit Assessment. *arXiv preprint arXiv:2507.22758.* July 2025. https://arxiv.org/pdf/2507.22758

Coding Agents

Ces agents spécialisés dans le développement logiciel sont chargés de générer, tester, réécrire ou commenter du code. Ils se distinguent par leur capacité à générer, déboguer, refactoriser et tester le code, ainsi qu'à produire de la documentation technique. Entraînés sur de vastes ensembles de données de code, ils apportent une assistance précieuse aux développeurs en automatisant les tâches répétitives et en suggérant des correctifs, tout en s'intégrant de manière transparente aux environnements de développement intégrés (*IDE*) et aux systèmes de gestion de versions.

- ***focus*** : générer[61], déboguer[62], refactoriser du code, et gérer des environnements de développement, changer de *framework*, portager avec changement de langage de programmation ;
- ***exemples*** : GitHub Copilot[63], IDE windsurf[64] ;
- ***caractéristiques typiques*** : agent spécialisé dans la réédition du code.

Agents d'optimisation et de simulation

Ils modélisent des scénarios, proposent des décisions optimales et simulent divers contextes afin d'optimiser l'allocation des ressources et la planification des tâches. Grâce à l'utilisation d'algorithmes d'optimisation et de modèles mathématiques, ils identifient les solutions les plus efficaces dans des situations complexes, telles que la logistique, la gestion des chaînes d'approvisionnement et la planification stratégique.

- ***focus*** : modéliser des scénarios, optimiser les ressources et simuler des résultats pour la prise de décision stratégique ;
- ***exemples*** : agents logistique, planification stratégique, optimisation des agents eux-mêmes[65] ;

---

[61] Hao Li, Haoxiang Zhang, Ahmed E. Hassan. The Rise of AI Teammates in Software Engineering (SE) 3.0: How Autonomous Coding Agents Are Reshaping Software Engineering. *arXiv preprint arXiv:2507.15003*. July 2025. https://arxiv.org/pdf/2507.15003

[62] Jess Weatherbed. Google's New Jules AI Agent Will Help Developers Fix Buggy Code. *The Verge*. December 11, 2024. https://www.theverge.com/2024/12/11/24318628/jules-google-ai-coding-agent-gemini-2-0-announcement

[63] Tom Warren. GitHub Copilot Gets a New ChatGPT-like Assistant to Help Developers Write and Fix Code. *The Verge*. March 22, 2023, https://www.theverge.com/2023/3/22/23651456/github-copilot-x-gpt-4-code-chat-voice-support

[64] L'éditeur Windsurf facilite l'intégration de l'IA dans le processus de développement logiciel. https://windsurf.com

[65] Han Zhou et al. Multi-Agent Design: Optimizing Agents with Better Prompts and Topologies. *arXiv preprint arXiv:2502.02533*. February 2025. https://arxiv.org/pdf/2502.02533

- ***caractéristiques typiques*** : planification, prise de décision, mémoire, economicus, impact, explicabilité.

Agents actionnables

Ces systèmes opèrent au sein d'environnements logiciels tels que les navigateurs, les *APIs* et les *ERP*. Ils sont conçus pour observer leur environnement en collectant des données, par exemple via des capteurs ou des *APIs* (perception), interpréter ces données (raisonnement), planifier des actions et les exécuter. Afin d'optimiser leurs performances au fil du temps, ils peuvent intégrer des mécanismes d'apprentissage à partir de leurs actions passées.

- ***focus*** : agents orientés action, contrôle de navigateur web, contrôle des applications via *API, MCP* ;
- ***exemples*** : agents qui agissent (cliquent, remplissent, automatisent) à la place des utilisateurs humains devant leur ordinateur[66, 67] ;
- ***caractéristiques typiques*** : interaction avec l'environnement, apprentissage.

3.2.5. Systèmes multi-agents et agents collaboratifs

Cette famille représente les environnements où plusieurs agents spécialisés collaborent afin de résoudre un problème de plus grande envergure.

Agents collaboratifs / multi-agents

Dans le contexte d'une collaboration entre agents dotés de compétences spécialisées et d'une coordination efficace, plusieurs agents experts interagissent et se coordonnent afin de résoudre des problématiques complexes. Cette collaboration entre agents permet une validation croisée des résultats, une répartition optimale des tâches et une gestion plus efficace des erreurs. Ces systèmes multi-agents se révèlent particulièrement adaptés aux environnements multidisciplinaires où une expertise unique ne saurait couvrir l'ensemble des facettes d'un problème.

[66] Stephane Nachez. OpenAI déploie Operator, son agent IA de navigation web en Europe. *ActuIA*. 13 mars 2025. https://www.actuia.com/actualite/agents-ia-openai-deploie-operator-en-europe/
[67] Computer Use | Gemini API. Google AI for Developers. https://ai.google.dev/gemini-api/docs/computer-use. Consulté le 10 octobre 2025.

- ***focus*** : plusieurs agents spécialisés interagissant[68], validant les décisions et partageant le contexte pour résoudre des problèmes complexes de manière coordonnée.
- ***exemples*** : *c*ollaboration entre employés supportés par des agents[69], assistant de projet distribué[70], agents distribués[71] avec stratégies variées ;
- ***caractéristiques typiques*** : coordination, organisation, communication, distribution.

Agents explorateurs / mobiles

Ces agents opèrent au sein d'un réseau, interagissant de manière autonome sur diverses plateformes. Ils possèdent la capacité de se déplacer au sein d'environnements numériques distribués, de s'adapter à des contextes d'exécution variés et de collecter ou de traiter des données localement. Conçus pour évoluer dans des systèmes complexes tels que *l'edge computing*, les réseaux *IoT* ou les bases de données sensibles, où la centralisation des données est impossible, leur autonomie, leur mobilité et leur aptitude à gérer des environnements hétérogènes en font des outils singuliers pour la surveillance, l'analyse décentralisée ou la préservation de la confidentialité.

- ***focus*** : capables de s'adapter car ils comprennent leur environnement et ce qui en fait les caractéristiques nécessitant adaptation ;
- ***exemples*** : *Crawlers* IA[72], agents *edge*, agents pour la donnée privée[73], essaim de drones[74] ;
- ***caractéristiques typiques*** : mobilité, représentation et interaction avec l'environnement, prise de décision, mémoire, apprentissage, taille réduite du modèle

---

[68] Khanh-Tung Tran et al. Multi-Agent Collaboration Mechanisms: A Survey of LLMs. *arXiv preprint arXiv:2501.06322*. January 2025. https://arxiv.org/pdf/2501.06322

[69] Sun, Songtao, et al. Multi-agent Application System in Office Collaboration Scenarios. *arXiv preprint arXiv:2503.19584*. April 2025. https://arxiv.org/pdf/2503.19584

[70] Stavros Doropoulos, Stavros Vologiannidis, Ioannis Magnisalis. DevNous: An LLM-Based Multi-Agent System for Grounding IT Project Management in Unstructured Conversation. *arXiv preprint arXiv:2508.08761*. August 2025. https://arxiv.org/pdf/2508.08761

[71] Yongchao Chen et al. TUMIX: Multi-Agent Test-Time Scaling with Tool-Use Mixture. *arXiv preprint arXiv:2510.01279*. October 2025. https://arxiv.org/pdf/2510.01279

[72] Yu Li, Bryce Wang, Xinyu Luan. XPath Agent: An Efficient XPath Programming Agent Based on LLM for Web Crawler. *arXiv preprint arXiv:2502.15688*. December 2024. https://arxiv.org/pdf/2502.15688

[73] Adelusi Moji et al. Multi-Agent Game Models for Privacy-Utility Trade-offs in Edge AI Networks. *ResearchGate*. 2024. https://www.researchgate.net/publication/391195767_Multi-Agent_Game_Models_for_Privacy-Utility_Trade-offs_in_Edge_AI_Networks

[74] Ludovic Sauzier et al. Essaims : menaces et opportunités. Ministère des Armées. 22 mai 2023. https://www.defense.gouv.fr/sites/default/files/dgris/EPS%202022-01%20-%20Essaims%20de%20drones%20aériens%2C%20menaces%20et%20opportunités.pdf

(ou modèle spécialisé comme un *LLM* compact [*Small Language Model* : *SLM*]), exécutable dans des environnements à ressources calculatoires limitées.

Agents de gouvernance

Les agents de gouvernance sont responsables de la supervision, du contrôle et du suivi des actions et de l'impact d'autres agents au sein du système. Leur mission consiste à superviser, encadrer et évaluer le comportement des agents, garantissant ainsi le respect des règles de conformité, de sécurité et d'éthique[75]. Ils analysent les journaux d'interaction, procèdent à des audits des décisions prises et émettent des alertes en cas de dérive. Ces agents sont indispensables dans des contextes critiques ou réglementés, où la traçabilité, la fiabilité et la responsabilité des agents doivent être assurées. Ils jouent un rôle crucial dans l'instauration de la confiance et la maîtrise des risques au sein des systèmes multi-agents ou autonomes.

- ***focus*** : capacité à évaluer et à faire rectifier les réponses/décisions/actions des agents supervisés ;
- ***exemples*** : vérificateurs de biais[76], auditeurs éthiques IA[77] ;
- ***caractéristiques typiques*** : Economicus, impact, explicabilité, apprentissage.

3.3. Caractéristiques des agents experts IA

Dans le domaine de l'IA, les agents experts sont des systèmes autonomes dotés de la capacité de percevoir leur environnement, de prendre des décisions éclairées et d'agir de manière intelligente afin d'atteindre des objectifs spécifiques. Les caractéristiques des agents présentées ici sont à considérer comme des attributs observables par lesquels ils interagissent avec le monde et les utilisateurs. Ces caractéristiques sont fondamentales pour leur compréhension, leur manipulation et leur classification (cf. section 0 Typologie).

Il est important de noter que les caractéristiques présentées ci-après ne sont pas toutes indépendantes les unes des autres. Par exemple, un agent présentant un certain niveau

[75] Suyash Gaurav, Jukka Heikkonen, Jatin Chaudhary. Governance-as-a-Service: A Multi-Agent Framework for AI System Compliance and Policy Enforcement. *arXiv preprint arXiv:2508.18765*. August 2025. https://arxiv.org/pdf/2508.18765

[76] Martijn de Vos et al. Fairness Auditing with Multi-Agent Collaboration. Version 3, *arXiv preprint arXiv:2402.08522*. February 2024. https://arxiv.org/pdf/2402.08522

[77] Leif Azzopardi, Yashar Moshfeghi. PRISM: A Methodology for Auditing Biases in Large Language Models. *arXiv preprint arXiv:2410.18906*. October 2024. https://arxiv.org/pdf/2410.18906

d'autonomie devra nécessairement disposer d'une capacité de planification. De même, si un agent est conçu pour être collaboratif, cela implique une certaine autonomie. De nombreux autres exemples peuvent être établis à travers les différentes caractéristiques.

Enfin, pour la majorité de ces caractéristiques, le document établit une graduation (désignée sous le terme d'échelle de valeurs) déterminant le degré de présence de la caractéristique chez l'agent considéré.

3.3.1. Capacités cognitives

Ce sont les aptitudes cognitives fondamentales qui confèrent à l'agent expert d'IA la capacité de raisonner, de planifier, d'acquérir des connaissances et d'interagir de manière vraisemblablement intelligente.

Il n'est pas impératif qu'un agent maîtrise ces compétences à leur niveau optimal. La complexité d'une compétence est directement proportionnelle aux ressources temporelles et matérielles qu'elle requiert pour son exécution.

Autonomie /agentivité/ intentionnalité

Dans ce livre blanc, le terme « autonome » doit être compris au sens d'un niveau d'agentivité (au moins proactif) ; un assistant conversationnel standard se situe le plus souvent au niveau réactif, sauf intégration explicite à des outils et à une orchestration d'actions.

Cette caractéristique évalue le degré d'autonomie de l'agent expert d'IA, c'est-à-dire sa faculté à initier une action et à se fixer un objectif sans intervention humaine directe. Ce degré d'autonomie peut être extrêmement rudimentaire (un agent qui réagit à un stimulus et exécute sa tâche) ou parfaitement atteint lorsque l'agent possède une intention et mobilise l'intégralité de ses capacités pour la concrétiser.

Échelle de valeurs :

- **passif** : l'agent est un simple objet qui répond à des invocations de méthodes (exemple : rechercher du texte et le remplacer par un autre) ;
- **réactif** : l'agent agit selon des règles stimuli/action, exemple : *bots* basiques ;
- **proactif** : l'agent agit pour atteindre les objectifs propres qui lui sont assignés, ajuste ses stratégies, choisit voire crée ses propres outils (codage autonome) ;
- **dubitatif** : l'agent peut s'interroger sur les conséquences potentielles de son action. Cela peut l'amener à mettre en pause son activité en attendant la validation d'un humain pour poursuivre ;

- **intentionnel :** capable de trouver le juste milieu (doctrine) et les priorités entre ses propres objectifs et des intentions ou cadre réglementaire par exemple.

Capacité de planification

La capacité à structurer de manière rationnelle des séquences d'actions afin d'atteindre un objectif précis.

Échelle de valeurs :

- **aucune** : l'agent est utilisé directement par un autre acteur (humain ou agent ou système d'orchestration) ;
- **conditionnelle** : l'agent peut choisir parmi plusieurs stratégies définies en amont en fonction de critères qui lui ont été assignés (un équivalent de « *switch… ; case* ») ;
- **multi-étapes** : pour remplir sa tâche, il peut prendre des décisions sur l'organisation de ses sous-tâches ;
- **osmotique** : l'agent a sa propre capacité de planification interne qui agit en osmose avec les acteurs pour lesquels il agit (orchestrateur, autres agents).

Capacité d'apprentissage

C'est la capacité à ajuster ses comportements ou ses stratégies en fonction de l'expérience acquise, des retours d'information reçus ou de l'intégration de nouvelles données.

Échelle de valeurs :

- **figé** : Aucun apprentissage (connaissances figées – sorte de « *cut-off date* » à la date de création/mise-à-jour de l'agent) ;
- **guidé :** apprend à partir des *feedbacks* qu'il reçoit de son environnement (agent ou humain) ;
- **auto-adaptatif :** trouve des *patterns* sans supervision humaine ;
- **curieux :** fait évoluer ses connaissances par différents moyens, optimise ses actions via un système de récompenses, par imitation en observant des démonstrations humaines (sur Internet) ou IA ou par évolution à travers des algorithmes inspirés de la sélection naturelle (cf. algorithmes génétiques, algorithmes d'exploration / exploitation (Q-Learning)).

Les principes présentés ici ne sont pas mutuellement exclusifs.

Prise de décision

La capacité à opérer un choix parmi une pluralité d'alternatives possibles, notamment dans un contexte caractérisé par l'incertitude, constitue une compétence essentielle.

Échelle de valeurs :

- **bordé :** l'agent n'a pas de capacité de décision explicite (ce qui signifie aussi qu'il y a des aléas dans ses actions : hallucinations éventuelles) ;
- **règles simples** : basé sur des règles (arbre de décision) ;
- **pondérée** : les décisions sont sujettes à des arbitrages statistiques (poids) ;
- **argumentée** : l'agent expert IA a la capacité de prendre des décisions motivées [optimisation et planification (Monte Carlo & co), combinaison hybride (*Neuro-symbolic AI)*].

Mémoire

La faculté de l'agent à accéder à ses souvenirs d'exécutions antérieures s'étend de la mémoire à court terme, caractérisée par une fenêtre de contexte restreinte, à une mémoire à long terme, définie par une persistance permettant la personnalisation et la continuité des interactions. Dans certains contextes, il peut être opportun d'intégrer à l'agent une fonctionnalité de réinitialisation de sa mémoire, afin de pallier les biais potentiels, l'obsolescence des informations perçues, les considérations éthiques ou les problèmes de performance liés au surapprentissage, équivalente soit à un mécanisme de « bouton de destruction de la mémoire », soit à un effacement par oubli progressif des données envers lesquels la confiance n'est plus suffisante (trop âgées, trop volatiles).

Échelle de valeurs :

- **sans état** : il ne conserve aucune trace de ses dernières activations. Il est idempotent (sa fenêtre de contexte est donc réinitialisée entre deux activations) ;
- **limitée** : il a une mémoire à court terme (probablement liée à sa fenêtre de contexte) et peut donc être activé plusieurs fois de suite dans le même *workflow* en se souvenant des éléments déjà produits ;
- **augmentée** : on introduit la notion de mémoire temporaire, le temps de finaliser un processus complexe dans lequel l'agent est appelé à plusieurs reprises et il conserve des informations d'un appel à l'autre au sein du même processus. Il a une autre mémoire s'il est invoqué dans un autre processus ou une autre instance du même processus ;

- **Gogol** (le nombre, pas le romancier) : l'agent apprend au fur et à mesure de ses invocations et interactions. Chaque invocation augmente sa connaissance.

Représentation de l'environnement

La capacité de l'agent à posséder une représentation interne de l'état du monde, ou plus précisément du contexte dans lequel il évolue, est un élément crucial de son efficacité opérationnelle.

Échelle de valeurs :

- **absente** : l'agent se base uniquement sur l'entrée actuelle sans plus d'information sur son environnement ;
- **délibératif** : utilise des modèles internes de l'environnement pour raisonner et planifier (exemple : planification d'itinéraire) ;
- **partiellement observable** : décisions basées sur des informations incomplètes ;
- **complètement observable** : l'environnement est entièrement connu.

Autoréflexion / métacognition

L'agent expert IA doit être en mesure d'évaluer de manière critique sa propre progression, la qualité de ses réponses et la validité de sa stratégie. Cette capacité d'auto-évaluation lui permettra d'ajuster dynamiquement son plan d'action ou sa méthodologie en fonction des besoins et des exigences du projet.

Échelle de valeurs :

- **brute force** : l'agent, exécutant, déroule son plan sans se poser de question ;
- **auto-évaluation** : l'agent est capable de donner une note à son activité ;
- **auto-amélioratif** : l'agent est capable de proposer des améliorations à ses résultats ;
- **réflexif** : l'agent met spontanément en place une réflexion sur la qualité de ses activités, abandonne volontairement une stratégie inefficace pour en essayer une autre.

3.3.2. Capacités interactionnelles

Les mécanismes d'interaction de l'agent expert IA avec une pluralité d'entités, incluant les individus, les agents et les systèmes, ainsi que son intégration au sein d'un environnement élargi.

Capacité de collaboration

Collaboration avec d'autres agents ou individus, partage de responsabilités ou d'objectifs communs.

Échelle de valeurs :

- **isolé** : l'agent réalise son activité sans connaissance des autres acteurs ;
- **compétitif** : cherche à maximiser ses propres gains, souvent aux dépens des autres agents (ex : *trading* algorithmique) ;
- **coordonné** : l'agent sait mener son activité de manière coordonnée avec d'autres. Il sait par exemple se mettre en attente d'autres acteurs (humains ou agents) pour poursuivre ses activités (exemple : un agent assistant-auteur attend qu'un auteur humain choisisse l'orientation à donner à la suite d'un texte, capacité à déléguer des tâches) ;
- **coopérant** : l'agent est capable de juger de la criticité d'autres agents et de décider de s'éloigner temporairement de son objectif individuel pour permettre de débloquer la situation d'autres agents. (exemple : une navette autonome peut décider de quitter sa machine-outil pour en permettre l'accès à une autre navette ayant une commande plus urgente) ;
- **synergique** : l'agent travaille avec d'autres, se coordonne, coopère, pour atteindre un objectif commun (exemple : robots sur une chaîne de montage qui coordonnent leurs actions, essaim de drones).

Interaction avec l'environnement

La capacité à percevoir, modifier ou recevoir des retours de l'environnement (c'est-à-dire du système dans lequel l'individu est intégré) est essentielle.

Échelle de valeurs :

- **uniquement en lecture** : par exemple agent d'analyse de données ;
- **interactif :** par exemple assistant vocal interagissant avec un humain ou d'autres agents : aptitude à dialoguer, argumenter, négocier dans un langage compréhensible par un humain ;
- **action directe** sur l'environnement : par exemple robots, *trading* algorithmique, *Computer Use*, *MCP*.

Il convient de souligner que les deux dernières valeurs mentionnées peuvent être additionnées.

Transparence et explicabilité

La transparence des décisions et des actions, ainsi que la possibilité d'une supervision humaine, sont essentielles. La capacité d'un agent expert d'IA à rendre intelligibles ses décisions, actions ou raisonnements à un humain est primordiale.

Échelle de valeurs :

- **opaque :** par exemple la plupart des *LLMs* actuels et les modèles de génération d'images par diffusion ;
- **transparent** : l'agent est capable de montrer comment il a pris ses décisions (exemple : système de raisonnement réellement suivi[78]) ;
- **explicable** : non seulement l'agent est transparent, mais il sait expliquer pourquoi il a suivi ce cheminement plutôt qu'un autre, il sait justifier une action ou il peut expliquer pourquoi il préconise telle ou telle proposition ;
- **éthique** : l'agent est explicable et calé sur un système de valeurs défini et connu.

Dans le cadre d'un système multi-agent, l'agent orchestrateur fera la synthèse de l'ensemble des explications des agents du système.

Multimodalité

La compétence à recevoir, interpréter et produire des informations sous divers formats (texte, image, audio, tableaux, *APIs*…) et à adapter le protocole d'échange à son interlocuteur (humain ou machine) est essentielle.

Échelle de valeurs :

- **mono** : l'agent est capable de traiter un unique type d'entrée/sortie, par exemple, du texte pour un agent de traduction ;
- **multi** : l'agent est capable de recevoir plusieurs types de formats en entrée et délivre possiblement plusieurs types en sorties. Par exemple, un agent génère une vidéo à partir d'une série de photos et d'un texte narratif ;
- **omni** : l'agent possède l'autonomie suffisante dans la sélection du bon canal de communication. Par exemple, il peut choisir entre parler à un utilisateur ou à un agent

[78] Il a été démontré que certains systèmes à base de raisonnement comme OpenAI o1 https://openai.com/fr-FR/o1/ explicitait un mode de raisonnement qu'ils avaient prétendument suivi pour atteindre leur résultat mais qu'en fait leur cheminement n'avait pas été celui-là… (cf. Sigal Samuel. Is AI really thinking and reasoning — or just pretending to? Vox. February 21, 2025. https://www.vox.com/future-perfect/400531/ai-reasoning-models-openai-deepseek)

via *API*) et met en place la transformation entre types de données (exemple : transformer du texte en tableau ou en image).

3.3.3. Capacités systémiques

Les caractéristiques opérationnelles ou infrastructurelles qui impactent le fonctionnement de l'agent au sein d'un système.

Mobilité

Le système est conçu pour permettre un déploiement distribué, ainsi qu'un changement de contexte ou d'environnement d'exécution.

Un agent peut être :

- situé dans un environnement physique (drone, avatar) ou simulé (personne dans une simulation d'évacuation de bâtiment) ;
- ou non situé et alors, il raisonne sur des concepts (par solveur logique, par réseau bayésien).

Seuls les agents situés, se voient attribuer de la mobilité.

Échelle de valeurs :

- **fixe** : reste à un emplacement fixe (exemple : contrôle-commande industriel, *LLM*, …) ;
- **mobile** : peut se déplacer dans son environnement (exemple : drones, robots mobiles ou même code).

Economicus

Cet ensemble d'indicateurs, proposé par l'agent (en totalité, partiellement ou non), englobe les ressources consommées (temps, mémoire, énergie), ainsi que le coût opérationnel et le modèle économique. Bien que ces indicateurs soient présentés comme indépendants les uns des autres, ils sont en réalité implicitement interconnectés.

Échelle de valeurs :

- **businessman** : l'agent publie son offre de prix et attend un accord (contrat) pour donner suite. Si cette facette est vide, pas de transparence sur le prix à payer ;
- **watt** : l'agent indique sa consommation énergétique moyenne ;
- **QCR** : l'agent peut donner des indications de Qualité x Coût x Rapidité, voire proposer des choix et compromis ;
- **souverain** : indicateur permettant de connaître le degré de souveraineté estimé de l'agent (cela inclut tant la partie apprentissage qu'inférence).

Impact

Évaluation du coût global engendré par l'agent qu'il soit énergétique, environnemental, social ou éthique, incluant les externalités indésirables.

Échelle de valeurs :

- **inconnu** : aucun indicateur d'impact n'est disponible, ni mesuré, ni estimé. L'agent fonctionne comme une boîte noire, sans considération explicite pour son empreinte ou ses effets collatéraux ;
- **mesuré** : l'impact énergétique ou environnemental est quantifié ou estimé, mais n'est pas pris en compte dans la conception ou l'usage (exemple : un *LLM* consomme X kWh mais continue d'être invoqué sans optimisation) ;
- **réduit** : l'agent intègre des mécanismes de limitation de ressources (quantité de requêtes, compression, sélection de modèles adaptés, mise en cache…), ou choisit des chemins plus sobres pour certaines tâches ;
- **responsable** : l'agent prend en compte l'impact dans sa prise de décision ou dans sa stratégie d'exécution (exemple : choisir une tâche moins consommatrice, prioriser un traitement local, éviter des usages à risque), et peut intégrer des critères éthiques (préservation des droits, filtrage des contenus, etc.).

Explicabilité

La capacité d'un agent à rendre intelligibles ses décisions, actions ou raisonnements à un être humain.

Échelle de valeurs :

- **opaque** : l'agent fournit un résultat sans explication, même sur demande (boîte noire complète) ;
- **justificatif** : l'agent peut produire une justification si elle est explicitement demandée, mais ne le fait pas spontanément ;
- **didactique** : l'agent fournit systématiquement un raisonnement, une justification ou une trace de son processus décisionnel (exemple : chaîne de pensées) ;
- **pédagogue** : l'agent ajuste dynamiquement le niveau d'explication à son interlocuteur (humain ou machine), peut contextualiser ses choix, comparer des alternatives, ou simuler des raisonnements contrefactuels.

Fiabilité/reproductibilité

La capacité à produire des résultats uniformes dans des conditions identiques, tout en assurant une variabilité contrôlée, est primordiale.

Échelle de valeurs :

- **instable** : résultats fortement variables, non reproductibles, même à données constantes ;
- **stable** : résultats globalement cohérents, mais avec des marges d'erreurs ou de variabilité non maîtrisées, dépendant de la nature de l'algorithme (le même *LLM* avec le même *prompt), ou de celle de son environnement (puissance de calcul, mémoire utilisable fluctuantes*) ;
- **reproductible** : mêmes entrées → mêmes sorties (ou équivalentes), dans un cadre contrôlé, avec journalisation des étapes clés (exemple : moteur de règles) ;
- **vérifiable** : l'agent intègre des mécanismes internes ou externes de contrôle, permet la **traçabilité complète**, et peut signaler les cas d'incertitude ou d'incomplétude dans ses résultats.

4. Architecture d'un agent expert IA : composants, patterns, orchestration

Dans le contexte dynamique de l'IA, les agents experts IA occupent une place prépondérante en automatisant les tâches, en prenant des décisions éclairées et en interagissant de manière autonome avec leur environnement. Afin de saisir pleinement la conception et le fonctionnement de ces agents, il est impératif d'examiner en profondeur leur architecture interne, constituée de multiples composants et briques fondamentales.

Cette section a pour objectif de fournir une analyse exhaustive des éléments constitutifs des agents experts, en détaillant chaque composant clé et son rôle spécifique. Nous explorerons par la suite les interactions entre ces composants, permettant ainsi la création de *patterns* d'agents performants et adaptatifs. Enfin, nous introduirons les systèmes multi-agents qui collaborent pour résoudre des tâches plus complexes.

4.1. Composants des agents experts IA

Les agents intelligents sont composés de divers éléments essentiels qui leur permettent d'exécuter des tâches spécifiques et d'atteindre leurs objectifs prédéfinis. Voici les principaux composants que l'on retrouve généralement au sein d'un agent IA.

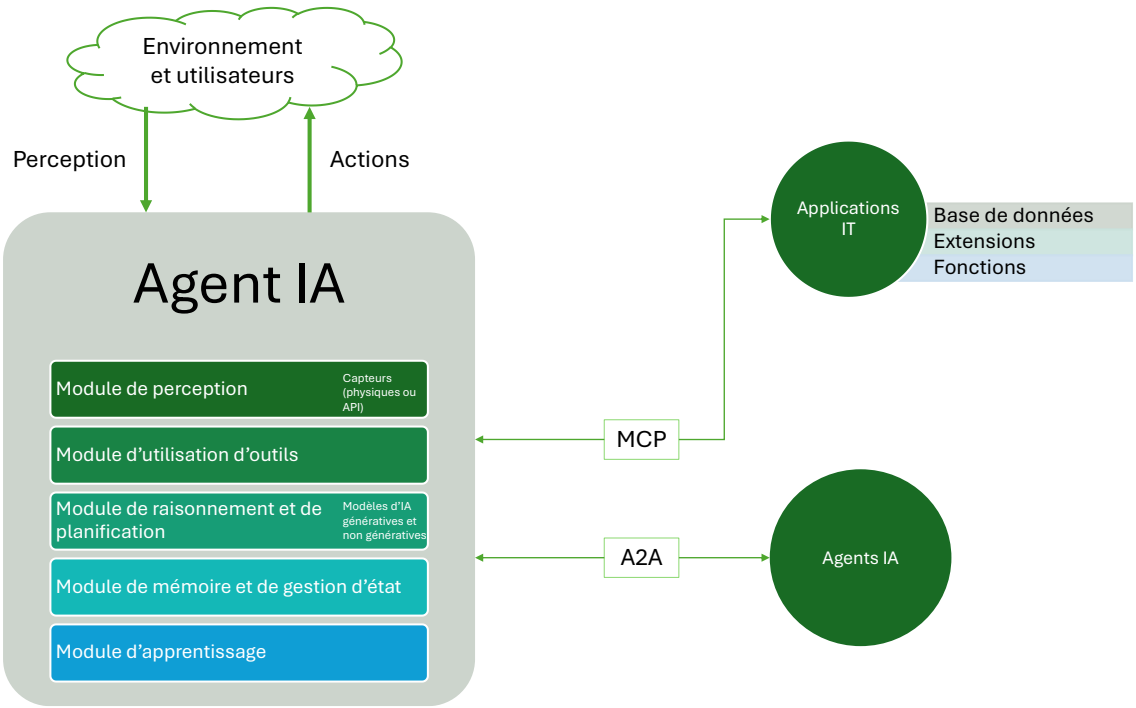

Figure 2: Architecture d'un agent IA

4.1.1. Module de perception

La **perception** constitue un élément fondamental pour les agents. Elle leur permet de recueillir des données relatives à leur environnement par l'intermédiaire de capteurs, comme par exemple les capteurs sensoriels. Que ce soit par l'analyse visuelle, la reconnaissance auditive ou l'accès à des sources de données, la perception offre aux agents la faculté de prendre des décisions éclairées ou d'exécuter des actions pertinentes, de s'adapter aux évolutions de leur environnement et d'interagir de manière appropriée avec celui-ci.

- **capteurs** : les capteurs permettent à l'agent de recueillir des informations sur son environnement. Ces capteurs peuvent être physiques, comme des caméras et des microphones, ou virtuels, comme des *APIs* collectant des données en ligne ;
- **analyse de données** : une fois les données collectées, l'agent les analyse pour en extraire des informations pertinentes. Cela peut inclure la reconnaissance des objets, des voix ou des textes.

4.1.2. Module d'utilisation d'outils

Afin d'assurer l'efficacité optimale des agents experts IA, l'intégration d'outils spécialisés s'avère indispensable. Ces outils permettent aux agents d'interagir de manière fluide avec l'environnement extérieur et de mobiliser les connaissances et compétences requises pour la réalisation des tâches qui leur sont confiées. L'utilisation d'outils peut englober des extensions, des fonctions ou des bases de données.

- **extensions :** fonctionnent comme des intermédiaires, facilitant la connexion d'un agent à des interfaces *API*. Elles permettent à l'agent de sélectionner l'*API* appropriée et d'extraire les paramètres requis à partir de la requête de l'utilisateur afin d'effectuer un appel réussi au point de terminaison de l'*API*. Exécutées du côté de l'agent, les extensions simplifient les appels aux *APIs* grâce à des techniques d'apprentissage par échantillons réduits. Elles offrent une interaction directe avec les *APIs*, garantissant un accès complet à celles-ci et un temps d'exécution optimisé en fonction des réponses fournies.

  L'intégration des *APIs* reposait auparavant sur un codage manuel et personnalisé pour chaque *API*. L'avènement du nouveau serveur *MCP* (voir ci-dessus note[57]) permet désormais à l'agent d'effectuer des appels *API* automatisés, d'identifier les outils disponibles, de saisir aisément les entrées requises et de remplir dynamiquement les paramètres.

- **fonctions** : contrairement aux bases de données, les fonctions déplacent l'exécution vers le client, conférant ainsi aux développeurs un contrôle renforcé sur les opérations sensibles et les flux de travail complexes. Dans ce contexte, l'accès de l'agent est limité. Ce système de fonctions est particulièrement adapté aux tâches nécessitant un niveau de sécurité élevé ou aux opérations asynchrones ;
- **bases de données** : ces dernières représentent des dépôts d'informations additionnelles aux données initiales utilisées pour l'entraînement du modèle. Elles permettent à l'agent d'accéder à des informations plus dynamiques et actualisées, constituant ainsi une source d'information externe exploitable par l'agent. Ces bases de données sont généralement vectorisées afin de faciliter l'application de techniques telles que le *RAG* ou le *RIG (Retrieval Interleaved Generation)*[79].

4.1.3. Module de raisonnement et de planification

Le module de raisonnement et de planification constitue le centre névralgique de l'agent, chargé de la prise de décision et de l'élaboration de stratégies opérationnelles afin d'atteindre les objectifs fixés, lui permettant ainsi un certain degré d'**autonomie**. Il procède à une évaluation rigoureuse de la situation actuelle et intègre les données issues des modules de mémoire et de perception afin de déterminer la stratégie optimale à mettre en œuvre. Il a la capacité de générer, de prédire, de classifier ou d'appliquer des règles, en combinant des approches déterministes, des modèles de *machine learning* classique ou d'IA générative. Il s'appuie sur des modèles de fondation comme GPT, Claude ou Llama pour interpréter les instructions, analyser le contexte et décider de la marche à suivre.

- **moteur de raisonnement :** est fondé sur une architecture cognitive sophistiquée qui orchestre les processus de raisonnement, de planification, de prise de décision et oriente les actions de l'agent en s'appuyant sur des règles prédéfinies ou des modèles d'apprentissage automatique.

  Le moteur d'inférence basé sur l'IA peut exploiter diverses méthodologies classées de la plus réactive à la plus cognitive.

---

[79] Le *RIG*, méthode relativement récente, transcende les limites du *RAG* en intégrant une dynamique itérative au processus de génération de réponses. Contrairement au *RAG*, le *RIG* autorise le *LLM* à interroger la base de données à plusieurs reprises durant la génération de texte, ce qui confère une précision accrue et une contextualisation approfondie.
Voir Sahin Ahmed. Retrieval Interleaved Generation (RIG) using LLM: what is it and how it works? *Medium*. October 2, 2024. https://medium.com/@sahin.samia/retrieval-interleaved-generation-rig-using-llm-what-is-it-and-how-it-works-aa8be0e27bbc

- ***ReAct***[80] **:** *framework* de *prompt engineering* permettant au modèle de raisonner et d'agir en réponse à une requête utilisateur. Son principe repose sur un cycle itératif (voir Figure 3) utilisant des actions et des observations afin d'affiner sa compréhension.

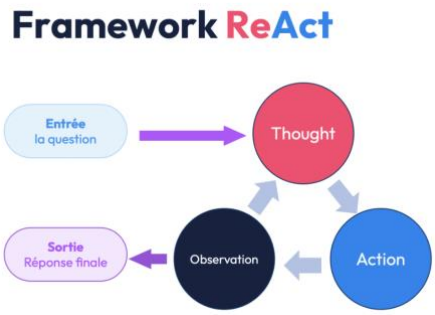

Figure 3 : Schéma ReAct[81]

- ***Chain-Of-Thought*** **(CoT)** : une architecture qui permet au modèle de raisonner à travers des étapes intermédiaires (voir Figure 4), ce qui améliore ses capacités de résolution de problèmes complexes.

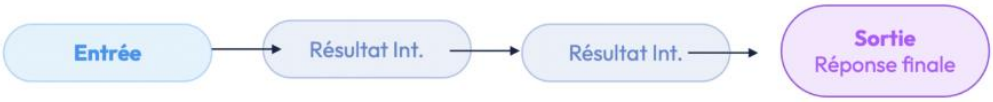

Figure 4 : Schéma Chain of thought (CoT)[81]**.**

- ***Tree-of-Thoughts*** **(ToT)**[82]**:** une généralisation de CoT, qui permet au modèle d'explorer plusieurs chaînes de raisonnement (voir Figure 5Figure 5), ce qui est idéal pour les tâches exploratoires ou les problèmes stratégiques.

[80] Le framework ReAct (Reasoning + Acting) a été introduit en 2022 par une équipe de chercheurs de Google DeepMind et de Princeton dans l'article scientifique : Shunyu Yao et al. ReAct: Synergizing Reasoning and Acting in Language Models. *ICLR2023*, the 11th international conference on learning representations. 2023. https://arxiv.org/abs/2210.03629

[81] D'après: Thibault Renouf. Le ReAct : une avancée majeure dans la conception des agents IA. *Tribes*. https://www.followtribes.io/react-agents-ia-cot/

[82] Shunyu Yao et al. Tree of Thoughts: Deliberate Problem Solving with Large Language Models. *Advances in neural information processing systems 36,* p. 11809-11822. 2023. https://proceedings.neurips.cc/paper_files/paper/2023/file/271db9922b8d1f4dd7aaef84ed5ac703-Paper-Conference.pdf

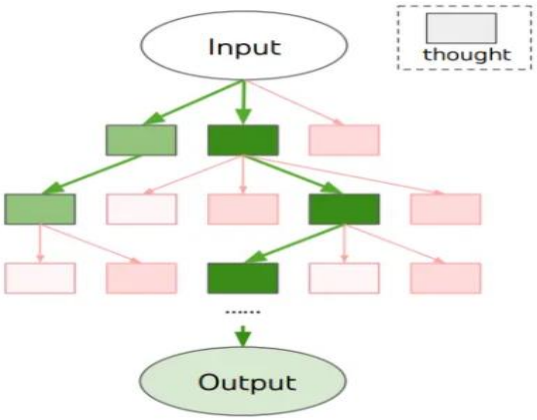

Figure 5: Schéma Tree-of-Thoughts (ToT)[83].

- **autoréflexion** : la capacité de l'agent d'examiner ses sorties précédentes pour tenter de corriger ses propres erreurs.
- **planificateur :** élabore des stratégies opérationnelles afin d'atteindre les objectifs définis, en s'appuyant sur les données disponibles. Il intègre les contraintes et les ressources pour optimiser les actions de l'agent.
- algorithmes **de planification** : algorithmes de recherche et de planification destinés à l'élaboration de plans d'action, c'est-à-dire des règles et des modèles d'IA.
- **optimisation** : techniques d'optimisation permettant de prendre en compte les contraintes et les ressources disponibles.

Dans le contexte des agents experts IA, l'intégration de modèles raisonnants, de modèles dotés de capacités de mémoire variables, et de *SLM* (*small language models*) permet de spécialiser les IA génératives pour des applications distinctes. Cela facilite la conception de sous-systèmes hautement spécialisés et performants, susceptibles d'être intégrés dans un « *mix of experts* ». Par exemple, les modèles raisonnants sont employés pour traiter des tâches complexes requérant une logique avancée et une prise de décision sophistiquée. Les mini *LLM*, quant à eux, offrent une flexibilité et une efficacité accrues pour des tâches spécifiques de traitement du langage naturel, tout en réduisant la charge computationnelle. L'ensemble de ces modèles spécialisés collabore afin de fournir des solutions précises et contextuellement adaptées, optimisant ainsi les performances globales des agents[84].

[83] D'après: Tree of Thoughts (ToT). Prompt Engineering Guide.
https://www.promptingguide.ai/techniques/tot
[84] Qingyun Wu *et al.* AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. In *Proceedings of the 1st Conference on Language Modeling (COLM 2024)*, Philadelphie, October 2024.
https://openreview.net/pdf?id=BAakY1hNKS

4.1.4. Module de mémoire et de gestion d'état

Les grands modèles de langage contemporains ne possèdent pas la capacité de se souvenir de manière autonome. Il est donc essentiel d'intégrer un composant mémoire au sein d'un agent pour en tirer pleinement parti.

Conformément à l'article *COALA*[85], élaboré par une équipe de l'Université de Princeton, il est possible de distinguer divers types de mémoire.

- **mémoire à court terme** (*STM*[86]) : ce mécanisme permet à un agent expert d'IA de conserver les entrées récentes afin d'optimiser la prise de décision immédiate. Ce type de mémoire revêt une importance capitale dans le domaine de l'IA conversationnelle, où la gestion du contexte à travers plusieurs échanges conversationnels est primordiale. Par exemple, un *chatbot* doté de cette capacité de mémorisation peut fournir des réponses cohérentes tout au long d'une session, évitant ainsi de traiter chaque entrée de l'utilisateur de manière isolée. Cette fonctionnalité contribue significativement à l'amélioration de l'expérience utilisateur ;
- **mémoire à long terme (*LTM*[87])** : permet aux agents de conserver et de récupérer des informations sur une période étendue, transcendant les différentes sessions. Cette capacité confère aux agents une dimension plus personnalisée et pertinente au fil du temps, tout en optimisant les temps d'exécution en s'appuyant sur des raisonnements antérieurs. Contrairement à la mémoire à court terme, la mémoire à long terme est conçue pour un stockage permanent, souvent mise en œuvre par le biais de bases de données, de graphes de connaissances ou d'intégrations vectorielles.

Par ailleurs, l'article *CoALA* identifie trois catégories distinctes de mémoire à long terme, comme illustré dans le tableau 2 ci-dessous :

Type de mémoire	**Description et caractéristiques**	**Implémentation pratique**
Mémoire épisodique	• Permet de se souvenir d'expériences et d'évènements passés, semblable à la mémoire humaine.	• Utilisation de bases de données pour stocker les événements précis avec les actions et les résultats.

[85] Theodore R. Sumers, Shunyu Yao, Karthik Narasimhan, Thomas L. Griffiths. Cognitive Architectures for Language Agents. *Transactions on Machine Learning Research 02/2024*. https://arxiv.org/abs/2309.02427
[86] Short term memory
[87] Long term memory

Mémoire sémantique	• Permet le stockage de connaissances factuelles structurées que l'IA peut récupérer pour raisonner. • Contient des informations généralisées (faits, définitions, règles).	• Développement de bases de connaissances. • Utilisation de techniques d'intégration vectorielle pour faciliter la recherche d'informations.
Mémoire procédurale	• Capacité de stocker et rappeler des compétences et comportements appris pour effectuer des tâches automatiquement.	• Mise en œuvre d'algorithmes d'apprentissage par renforcement pour entraîner les agents.

Tableau 3 : Catégories des mémoires à long terme

4.1.5. Module d'apprentissage

Ce module confère à l'agent expert IA la capacité d'inférer des conclusions à partir d'expériences passées et de s'adapter à des contextes nouveaux. Il repose sur des méthodologies d'apprentissage et d'adaptation sophistiquées.

- **algorithmes d'apprentissage** : ces algorithmes sont mis en œuvre afin d'optimiser les performances de l'agent au fil du temps, en lui permettant d'apprendre de ses expériences et des données collectées. Cette optimisation peut être réalisée par le biais de diverses méthodes, notamment l'apprentissage supervisé, non supervisé ou par renforcement ;
- **adaptation** : l'agent ajuste ses stratégies et ses comportements en fonction des nouvelles informations et des changements intervenus dans l'environnement. L'adaptation revêt une importance capitale pour garantir l'efficacité de l'agent dans des environnements dynamiques ;
- ***in-context-learning*** : ce mécanisme permet à l'agent d'acquérir des compétences et d'exécuter des tâches à l'aide de prompts, d'exemples et d'instructions fournis au moment de l'inférence, sans nécessiter de mise à jour des poids ou de réentraînement du modèle ;
- ***retrieval-based in-context learning*** : dans ce contexte, l'agent récupère des exemples pertinents à partir de sa mémoire externe afin d'améliorer son adaptation à la requête utilisateur ;
- ***fine-tuning based learning*** : ce processus permet d'entraîner le modèle sur des exemples spécifiques et sur des données supplémentaires spécialisées, afin

d'accroître sa capacité à exécuter certaines tâches ou à sélectionner les outils les plus appropriés.

4.2. Patterns

Un *pattern* constitue une solution réutilisable à un problème récurrent. Il représente un cadre ou un modèle abstrait facilitant la structuration et l'organisation de systèmes, notamment d'agents experts IA. Contrairement à une implémentation spécifique, un *pattern* offre la flexibilité nécessaire pour être adapté à des exigences particulières.

Eric Broda[88], identifie quatre catégories de *patterns* d'agents : les *patterns* de communication, les *patterns* fonctionnels, les *patterns* de rôle et les *patterns* organisationnels.

4.2.1. Patterns de communication

Les agents sont conçus pour structurer les échanges et les interactions entre agents experts IA ou entre agents et humains (agents conversationnels). Ils organisent la manière dont les agents experts IA partagent des informations, sollicitent de l'assistance, délèguent des tâches ou collaborent sur des objectifs communs. Ces agents intègrent des mécanismes permettant de gérer les interactions synchrones, asynchrones ou collaboratives, répondant ainsi aux besoins de coordination, d'escalade et de diffusion d'informations.

Caractéristiques :

- les agents englobent des interactions allant de simples échanges, tels que ceux avec des *chatbots*, à des conversations complexes, comme les négociations multi-agents, ou agent-humain ;
- ils s'adaptent à des scénarios variés, incluant la diffusion en masse (*broadcast*) et les escalades urgentes. Selon les besoins, ils peuvent maintenir ou non un état contextuel ;

Description des composants :

- **perception** : les agents conversationnels perçoivent les messages entrants, qu'ils proviennent d'autres agents ou d'humains. La détection de ces signaux ou requêtes

[88] Eric Broda. Agentic Mesh: patterns for an agent ecosystem. *Medium*. Data Science Collective. February 10, 2025. https://medium.com/data-science-collective/agentic-mesh-patterns-for-an-agent-ecosystem-ef13469b7cf7

nécessite l'utilisation de capteurs ou d'interfaces, telles que des *APIs* ou des interfaces conversationnelles ;

- **raisonnement :** les agents interprètent les requêtes afin de déterminer la réponse ou l'action la plus appropriée. Dans les *patterns* complexes, tels que les conversations, le raisonnement contextuel est primordial ;
- **mémoire :** la mémoire est peu sollicitée dans les interactions simples, où la conservation de l'état n'est pas nécessaire. En revanche, elle est essentielle dans les conversations afin de maintenir le contexte sur plusieurs étapes ;
- **action** : les agents répondent aux requêtes ou transmettent des messages en fonction des besoins identifiés. Des exemples d'actions incluent la diffusion d'informations (*broadcast)* et l'escalade (*attention*) ;
- **apprentissage :** les agents ont la capacité d'améliorer leurs réponses en analysant les interactions passées. Pour les agents, des techniques telles que le traitement du langage naturel (*NLP*) sont utilisées pour adapter les réponses en fonction de l'expérience acquise. En interaction multi-agent, un agent est capable d'évaluer la confiance qu'il a envers ses contacts ;
- **utilisation d'outils** : les agents conversationnels peuvent s'appuyer sur des *APIs* ou des bases de données pour récupérer ou transmettre des informations, optimisant ainsi l'efficacité et la précision de leurs interactions.

4.2.2. Patterns organisationnels

Les *patterns* organisationnels constituent le cadre structurel permettant la coordination et l'interaction entre plusieurs agents au sein d'un système à grande échelle. Ils mettent l'accent sur la hiérarchie, la coordination et la gestion efficace des flux d'informations. Ces *patterns* établissent des stratégies rigoureuses pour organiser des systèmes multi-agents, garantissant ainsi leur efficacité opérationnelle et leur évolutivité.

Caractéristiques :

- les *patterns* organisationnels facilitent des interactions fluides à grande échelle, favorisant ainsi la cohérence et la résilience au sein d'écosystèmes complexes.

Description des composants :

- **perception** : les agents perçoivent les événements globaux ou les signaux déclencheurs au sein du système. Ils surveillent également les autres agents afin de détecter des dépendances ou des changements d'état susceptibles d'impacter les performances globales du système ;

- **raisonnement** : le processus de raisonnement implique la coordination des tâches entre agents en fonction des priorités établies et des dépendances identifiées *a priori*. Un raisonnement stratégique est employé pour optimiser l'organisation globale du système et maximiser son efficacité pendant le fonctionnement du système ;
- **mémoire** : il est impératif que les agents conservent une vue d'ensemble des états et des interactions des autres agents. Une mémoire partagée ou distribuée est mise en place afin de suivre efficacement les dépendances et d'assurer une coordination optimale ;
- **action** : la coordination des actions entre agents est essentielle pour atteindre un objectif collectif. Des mécanismes tels que la synchronisation des tâches et la gestion des conflits sont mis en œuvre pour garantir une exécution fluide et efficace des actions ;
- **apprentissage** : les agents ont la capacité d'apprendre des erreurs organisationnelles passées afin d'améliorer la collaboration future. Une analyse approfondie des flux de travail est réalisée afin d'optimiser les interactions et d'accroître l'efficacité globale du système ;
- **utilisation d'outils** : des plateformes de coordination ou de gestion des flux d'informations, telles que des systèmes de gestion de tâches ou des outils de *workflow*, sont utilisées pour faciliter la coordination et la gestion des activités au sein du système.

4.2.3. Patterns de rôle

Les *patterns* de rôle établissent les responsabilités et les comportements spécifiques de chaque agent au sein d'un écosystème. Ils définissent des rôles clairs et précis afin d'éviter tout chevauchement ou conflit de compétences. Cette approche favorise la spécialisation des agents experts IA, leur permettant de se concentrer sur des tâches spécifiques et d'optimiser leur contribution à l'ensemble du système.

Caractéristiques :

- les *patterns* de rôle englobent des rôles tels que le planificateur (*planner)*, le coordinateur (*orchestrator*) et l'exécuteur (*executor*) ;
- ils facilitent une collaboration efficace en décomposant les tâches complexes en sous-tâches gérées par des agents spécialisés. Cette décomposition permet une meilleure gestion des ressources et des priorités, assurant ainsi l'atteinte des objectifs globaux de l'écosystème.

Description des composants :

- **perception** : les agents perçoivent les tâches ou les demandes qui relèvent de leur rôle prédéfini. Ils surveillent attentivement les signaux pertinents à leur domaine de spécialisation, leur permettant de réagir rapidement et efficacement aux besoins du système ;
- **raisonnement** : le raisonnement des agents est focalisé sur leur domaine ou spécialité. Ils sont dotés de la capacité de prendre des décisions autonomes dans les limites de leur rôle, en tenant compte des contraintes et des objectifs définis ;
- **mémoire** : les agents conservent un historique détaillé des tâches accomplies, ce qui leur permet d'affiner leur spécialisation et d'améliorer leurs performances au fil du temps. Le stockage des connaissances spécifiques à leur rôle constitue un atout précieux pour la résolution de problèmes et la prise de décision ;
- **action** : les agents exécutent les tâches spécifiques décrites dans les rôles qui leur sont attribués. Ils collaborent étroitement avec d'autres agents pour remplir des objectifs plus larges, contribuant ainsi au succès global de l'écosystème ;
- **apprentissage** : les agents apprennent continuellement à mieux remplir leur rôle en analysant leurs performances passées et en identifiant les domaines d'amélioration. Cette approche d'amélioration continue leur permet de rester à la pointe de leur spécialisation et de s'adapter aux évolutions du système ;
- **utilisation d'outils** : les agents spécialisés utilisent des outils adaptés à leur rôle, optimisant ainsi leur efficacité et leur productivité. Par exemple, un agent « planificateur » peut utiliser des outils d'optimisation pour élaborer des plans d'action efficaces, tandis qu'un « exécuteur » peut utiliser des *APIs* pour exécuter des actions de manière automatisée et précise.

4.2.4. Patterns fonctionnels

Les *patterns* fonctionnels se concentrent sur les méthodologies et les processus par lesquels les agents exécutent leurs tâches et atteignent leurs objectifs. Ils mettent l'accent sur les capacités opérationnelles des agents, englobant des mécanismes pour la surveillance, la planification et l'exécution des tâches.

Caractéristiques :

- ces *patterns* intègrent des agents à visée opérationnelle (*task-oriented*) ainsi que des agents à visée stratégique (*goal-oriented*), permettant une surveillance continue (*monitoring*) et la simulation (simulation agents via des stress tests et simulation des scénarios). Ils sont conçus pour s'adapter à des environnements dynamiques et non linéaires.

Description des composants :

- **perception** : les agents surveillent les entrées nécessaires à leurs tâches, telles que les capteurs et les flux de données. L'analyse des données permet de détecter des anomalies ou des opportunités ;
- **raisonnement** : le raisonnement implique la planification des tâches ou des actions pour atteindre un objectif spécifique, ainsi que la résolution de problèmes pour surmonter les obstacles opérationnels ;
- **mémoire** : la mémoire assure la conservation des données relatives aux tâches en cours ou passées, ainsi que l'historique des performances, contribuant ainsi à l'amélioration de l'efficacité ;
- **action** : l'action englobe l'exécution des tâches nécessaires, ainsi que la collaboration avec d'autres agents pour atteindre des objectifs complexes ;
- **apprentissage** : les agents apprennent à optimiser leurs performances en fonction des résultats obtenus, en utilisant des simulations pour tester différentes approches ;
- **utilisation d'outils** : les agents fonctionnels utilisent des outils spécifiques à leurs tâches, tels que des outils d'analyse, des simulateurs et des *APIs* de gestion.

4.3. Systèmes multi-agents et orchestration

Le paysage des agents experts IA évolue à une cadence fulgurante. Si des plateformes multi-agent existent depuis les années 1990 dans le monde de la recherche, avec quelques applications intéressantes dans le monde industriel, concernant les agents experts liés à des LLM, de nouveaux utilitaires apparaissent chaque mois, et seuls ceux apportant une valeur tangible survivent – une véritable « sélection naturelle » opérée par les utilisateurs. Parallèlement, les critères de choix dépassent la seule performance : souveraineté technologique, confidentialité des données et ouverture du code s'imposent désormais. Toute architecture multi-agents doit donc viser à la fois l'efficacité et la conformité aux exigences de confiance, de gouvernance et de transparence.

4.3.1. Brique d'orchestration

Une tendance structurante est l'utilisation de systèmes multi-agents où plusieurs agents spécialisés collaborent pour accomplir des tâches complexes. La brique d'orchestration assure la coordination, la gestion des sous-agents, la sélection du modèle le plus adapté selon la tâche (*model switching*), l'intégration des outils externes et la prise en compte

des enjeux de cybersécurité afin de garantir un fonctionnement fluide et sécurisé des Agents experts IA.

Côté Microsoft, *AutoGen* propose un schéma de dialogue et de coopération entre agents *LLMs* outillés, facilitant la délégation de sous-tâches vers l'agent le plus compétent[89]. Dans la même veine, *Magentic-One*[90] introduit un *Orchestrator* central qui planifie, assigne et réajuste dynamiquement des rôles d'agents spécialisés (*WebSurfer*, *FileSurfer*, *Coder*, *Terminal*, etc.) afin d'atteindre des objectifs ouverts, avec des gains sensibles sur des *benchmarks* exigeants[91]. Cette approche illustre comment la coordination hiérarchique améliore la robustesse, la reprise après erreur et la réussite de tâches multi-étapes à grande échelle.

Sur l'*open source*, des briques d'orchestration s'imposent. *LangGraph* modélise les *workflows* d'agents sous forme de graphes persistants (exécution durable, mémoire, *human-in-the-loop*, reprise sur incident), ce qui facilite le contrôle pas-à-pas et l'industrialisation[92]. *CrewAI* permet de composer des *crews* d'agents dotés de rôles explicites et d'outils partagés et une mise en œuvre rapide pour des scénarios métiers[93].

Les fournisseurs de services cloud de grande envergure (*hyperscalers)* intègrent ces concepts. *Amazon Bedrock* proposera une fonctionnalité de collaboration multi-agents (disponible en aperçu fin 2024)[94] : un « superviseur » décomposera les requêtes, les délèguera à des agents spécialisés et consolidera les sorties. Des améliorations de taux de succès et de productivité ont été observées en interne sur des tâches multi-étapes.

---

[89] Qingyun Wu et al. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. In *Proceedings of the 1st Conference on Language Modeling (COLM 2024)*, Philadelphia, October 2024. *arXiv 2308.08155* (version 2, October 3, 2023) https://arxiv.org/pdf/2308.08155

[90] Chris Paoli. Microsoft Unveils Multi-Agent AI System Magnetic-One. *Redmond Magazine*, November 7, 2024. https://redmondmag.com/articles/2024/11/07/microsoft-unveils-multi-agent-ai-system-magnetic-one.aspx

[91] Adam Fourney et al. Magentic-One: A Generalist Multi-Agent System for Solving Complex Tasks. *arXiv preprint arXiv:2411.04468*, November 7, 2024. https://arxiv.org/abs/2411.04468.

[92] LangChain Team. LangGraph: Low-level orchestration framework for long-running, stateful agents. Documentation, 2025. https://langchain-ai.github.io/langgraph/.

[93] CrewAI Team. CrewAI Documentation: Build collaborative AI agents, crews and flows. Documentation, 2025. https://docs.crewai.com/

[94] Alfred Shen, Anya Derbakova. Design multi-agent orchestration with reasoning using Amazon Bedrock and open-source frameworks. *AWS Blogs*. December 19, 2024. https://aws.amazon.com/blogs/machine-learning/design-multi-agent-orchestration-with-reasoning-using-amazon-bedrock-and-open-source-frameworks/

De plus, des guides techniques détaillent l'intégration de *LangGraph* avec *Bedrock* pour des architectures *multi-agents graph-based*, prêtes pour la production[95].

Afin de prévenir l'enfermement technologique, un effort de standardisation se concrétise avec l'émergence du *LangChain Agent Protocol*. Ce protocole définit des *API*s communes pour la communication entre agents, englobant l'exécution (*runs*), l'état/les *threads* et la mémoire. Cette initiative favorise l'interopérabilité et l'intégration entre des solutions telles qu'*AutoGen, LangGraph, CrewAI* et d'autres architectures propriétaires[96].

4.3.2. Protocoles de communication

À mesure que les agents basés sur des grands modèles de langues (LLM) s'installent dans les flux métiers, la question n'est plus seulement « quel modèle choisir », mais « comment connecter des agents entre eux et aux systèmes de l'entreprise de façon sécurisée, standardisée et facile à passer à l'échelle. C'est précisément le rôle des protocoles ouverts : *MCP* (*Model Context Protocol)* pour relier un agent à ses outils et données, et *A2A* (*Agent-to-Agent*) pour permettre la collaboration entre agents hétérogènes. Ensemble, ces briques forment une couche d'interopérabilité essentielle aux systèmes multi-agents.

Introduit par Anthropic en 2024, *MCP* est un standard ouvert visant à remplacer les intégrations sur-mesure par une interface universelle entre agents et sources externes (extensions, outils métiers, environnements de développement). Concrètement[97], les développeurs exposent des serveurs MCP (qui déclarent des outils, ressources et *prompts*) et construisent des clients MCP (côté agent) qui s'y connectent via une architecture client-serveur inspirée du *Language Server Protocol*. MCP a plusieurs avantages :

- il standardise l'accès aux systèmes (CRM, bases SQL, fichiers, emails) sans écrire un connecteur par agent, ce qui réduit la friction et accélère les déploiements.

---

[95] Jagdeep Singh Soni, Ajeet Tewari, Rupinder Grewal. Build multi-agent systems with LangGraph and Amazon Bedrock. *AWS Blog*. April 14, 2025. https://aws.amazon.com/blogs/machine-learning/build-multi-agent-systems-with-langgraph-and-amazon-bedrock/.

[96] LangChain Team. Agent Protocol: Interoperability for LLM agents. *LangChain Blog*. November 19, 2024. https://blog.langchain.com/agent-protocol-interoperability-for-llm-agents/.

[97] World Economic Forum. AI Agents in Action: Foundations for Evaluation and Governance. November 2025. https://reports.weforum.org/docs/WEF_AI_Agents_in_Action_Foundations_for_Evaluation_and_Governance_2025.pdf

- Il rend l'agent actionnable : consulter un calendrier, récupérer des emails, mettre à jour une base, etc., via un contrat partagé et typé, limitant les erreurs et les hallucinations.
- Il s'insère dans des outils existants (IDE, assistants de code, bots internes) et favorise le *plug-and-play* d'outils métier.

Là où MCP règle l'accès aux outils, *A2A* s'attaque au besoin de faire collaborer des agents développés par des fournisseurs différents. En avril 2025, Google a annoncé A2A[98] comme protocole ouvert permettant à des agents de s'échanger des informations, déléguer des tâches et coordonner des actions à travers des plateformes et *clouds* hétérogènes. A2A est asynchrone : il prend en charge la découverte de capacités (via des *Agent Cards* qui décrivent les agents), des workflows multi-étapes, du streaming pour les tâches longues, et s'appuie sur des transports web standards (HTTP/HTTPS avec JSON-RPC 2.0). AGNTCY[99] est une alternative à A2A. Dans l'architecture AGNTCY, l'*Agent Connect Protocol* (ACP) propose une API REST standardisée pour découvrir, configurer et invoquer des agents.

Synthèse du chapitre

Ce document a mené une analyse approfondie des architectures internes des agents experts IA, en examinant minutieusement leurs composants fondamentaux – perception, mémoire, utilisation d'outils, raisonnement et planification, et apprentissage – ainsi que les quatre patterns clés :

- de communication ;
- fonctionnels ;
- de rôle;
- organisationnels.

Ces éléments constituent une base flexible pour la conception d'agents adaptés à une large gamme de cas d'utilisation, allant des systèmes simples aux agents hautement sophistiqués.

Un aspect crucial de ces architectures réside dans leur modularité : les composants internes ne sont pas systématiquement présents dans tous les agents. Un agent peut

[98] Rao Surapaneni, Miku Jha, Michael Vakoc, Todd Segal. Announcing the Agent2Agent Protocol (A2A). *Google blog.* April 9, 2025. https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/

[99] DeepWiki. agntcy/docs. https://deepwiki.com/agntcy/docs/1-overview

être minimaliste, reposant sur des règles simples, sans mémoire ni capacité d'apprentissage, et néanmoins parfaitement adapté à des tâches spécifiques. À l'inverse, un agent plus avancé peut intégrer un moteur de raisonnement performant tel qu'un modèle de *LLM*, une mémoire à long et court terme, ainsi que des algorithmes d'apprentissage et d'adaptation, lui permettant d'évoluer et de s'améliorer au fil du temps.

Un système multi-agent peut posséder des agents de différents niveaux de réactivité, d'autonomie, de réflexion. Ainsi, les agents au contact de l'environnement (capteur, matériel (moteur, interrupteur) sont réactifs pour une réaction épidermique rapide, et interagissent avec des agents de « plus haut niveau » capable de raisonnement et de planification sur le long terme des actions de l'ensemble du système.

En perspective, ces architectures offrent un terrain propice au développement d'agents IA de plus en plus autonomes, collaboratifs et adaptatifs. À mesure que les technologies progressent, nous pouvons anticiper des agents capables de s'intégrer de manière transparente dans des écosystèmes complexes, de co-évoluer avec les humains et de repousser les limites d'utilisation de l'IA dans des domaines encore inexplorés. Ces avancées promettent de transformer notre interaction avec la technologie.

HUB FRANCE IA

5. Les techniques des agents experts d'IA

5. Autres techniques

Les agents experts IA révolutionnent le paysage entrepreneurial. Leur développement s'appuie sur l'essor de l'IA générative, qui, d'après une enquête menée par Bain auprès de décideurs américains, révèle que 95 % des entreprises utilisent déjà l'IA générative en phase de production ou en phase pilote[100]. Cependant, le passage à l'échelle de ces solutions demeure un obstacle fréquemment évoqué[101].

L'intérêt pour les techniques d'agents experts s'explique par deux facteurs principaux. Premièrement, le règlement européen sur l'IA impose aux entreprises de procéder à une analyse approfondie des risques associés à l'IA, plaçant la confiance, les impacts sociétaux et la sobriété énergétique au cœur des préoccupations. Deuxièmement, la vague « *Agent AI* » progresse à un rythme sans précédent, avec l'émergence continuelle de nouveaux cadres conceptuels, modèles et architectures qui redéfinissent les paradigmes existants.

Pour maîtriser cette accélération tout en atténuant les risques, on aura la possibilité d'orchestrer plusieurs familles de méthodes, notamment les approches génératives, symboliques, neuro-symboliques, d'optimisation « classique » et les systèmes multi-agents, chacune présentant des avantages spécifiques ainsi que des limitations inhérentes.

La technique d'IA prédominante utilisée dans les agents experts IA est issue des *LLM* et de l'IA générative. Cependant, l'adoption d'autres techniques d'IA pourra être utile, notamment lors de la phase d'industrialisation, afin de minimiser les risques résiduels et de maintenir un certain degré de déterminisme. Ces techniques supplémentaires permettront d'assurer la fiabilité des résultats et de réduire les risques d'hallucinations, tout en offrant une efficacité énergétique accrue et en préservant un niveau de contrôle supérieur grâce à des capacités d'explicabilité renforcées.

Les points suivants seront abordés ici :

- les techniques d'IA symboliques et neuro-symboliques ;

[100] Gene Rapoport, Sanjin Bicanic, Muyiwa Talabi. Survey: Generative AI's Uptake Is Unprecedented Despite Roadblocks. *Bain & Company*. May 7, 2025. https://www.bain.com/insights/survey-generative-ai-uptake-is-unprecedented-despite-roadblocks/

[101] Adgully Bureau. Gen AI adoption has skyrocketed, but scaling remains a hurdle. *Bain & Company*. May 12, 2025. https://adgully.me/post/10561/gen-ai-adoption-has-skyrocketed-but-scaling-remains-a-hurdle-bain-company

- d'autres techniques algorithmiques telles que les solveurs, la planification sous contraintes et la recherche de corrélations.

Bien que cette liste ne soit pas exhaustive, elle ouvre la voie à l'exploration de techniques plus économes en ressources, même si elles peuvent parfois s'avérer plus contraignantes et nécessiter des prérequis plus importants que l'utilisation systématique des techniques de réseaux de neurones.

5.1. Techniques des IA symboliques

5.1.1. Principes fondamentaux du raisonnement symbolique

L'IA symbolique s'efforce de reproduire le raisonnement humain en s'appuyant sur notre logique et notre aptitude à représenter notre environnement par le biais de symboles. Comme le souligne le philosophe John Haugeland dans son ouvrage[102], cette approche repose sur deux postulats fondamentaux :

- notre capacité à traiter les informations de manière intelligente découle de notre aptitude à les appréhender de façon rationnelle ;
- cette aptitude équivaut à une faculté de manipulation symbolique interne « automatique ».

Le raisonnement symbolique postule qu'un système intelligent doit intégrer des sous-systèmes computationnels (des « ordinateurs internes ») afin d'effectuer des manipulations internes « rationnelles ». Cette conception de l'intelligence comme processus symbolique a orienté le développement des premiers systèmes d'IA et continue d'influencer la conception des agents experts contemporains, notamment lorsque l'explicabilité est un impératif. Les systèmes à base de connaissance, les solveurs, les algorithmes de planification, sont issus du raisonnement symbolique.

5.1.2. Architectures symboliques dans les agents experts

Les architectures symboliques traditionnelles reposent essentiellement sur des systèmes à base de règles et des mécanismes d'inférence logique. Ces méthodologies ont démontré leur efficacité dans des domaines caractérisés par la clarté des règles et l'importance primordiale de l'explicabilité. L'IA symbolique formalise les connaissances

[102] John Haugeland. Artificial intelligence: The very idea. *MIT Press*, Cambridge, MA, 1985. Andre Vellino. Book Review. *Artificial Intelligence* 29(3), p 349-353. September 1986. https://www.researchgate.net/publication/256121747_Artificial_intelligence_The_very_idea

sous la forme de règles logiques et de graphes, conférant ainsi transparence et contrôle. IBM Watson constitue un exemple emblématique de cette approche : le supercalculateur victorieux de la compétition *Jeopardy*! en 2011 s'appuyait sur l'architecture *DeepQA*, qui intégrait la recherche textuelle et le raisonnement basé sur des règles afin de comprendre les questions et de formuler des réponses pertinentes. Bien que depuis supplanté par les réseaux neuronaux, ce succès illustre la puissance du raisonnement symbolique dans des contextes spécifiques[103].

5.1.3. IA neuro-symbolique

Un système d'IA **neuro-symbolique** constitue une approche qui intègre l'IA symbolique, telle que celle employée dans les systèmes de navigation par satellite, et l'IA des réseaux neuronaux. Cette synergie permet de conjuguer le raisonnement symbolique avec les capacités d'apprentissage des réseaux neuronaux, engendrant ainsi des systèmes performants tant en matière de raisonnement déductif (inférence logique) qu'en matière d'apprentissage inductif (reconnaissance de motifs).

Comme l'a souligné David Cox (IBM), « ce qui fait défaut à l'IA symbolique, ce ne sont pas des données ou des traitements spécifiques, mais le *deep learning* »[104]. Cette observation met en évidence la complémentarité intrinsèque des deux approches et justifie leur intégration[105].

5.1.4. Avantages des approches neuro-symboliques pour les agents experts

Les systèmes neuro-symboliques offrent plusieurs avantages significatifs pour le développement d'agents experts IA :

- **meilleure explicabilité** : contrairement aux réseaux neuronaux souvent qualifiés de « boîtes noires », les systèmes neuro-symboliques peuvent fournir des explications sur leur raisonnement, ce qui est crucial dans de nombreux domaines d'application ;
- **capacités d'apprentissage robustes** : les composantes connexionnistes permettent aux agents d'apprendre à partir de données, compensant ainsi une limitation importante des systèmes purement symboliques ;

---

[103] David Ferrucci et al. Building Watson: An Overview of the DeepQA Project. *AI Magazine*, vol. 31, n° 3, p. 59-79. 2010. https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/2303/2165%3Cbr

[104] George Lawton. L'IA neuro-symbolique, évolution de l'intelligence artificielle ? *MagIT.* 23 novembre 2020. https://www.lemagit.fr/conseil/LIA-neuro-symbolique-levolution-de-lintelligence-artificielle

[105] George Lawton. Neuro-symbolic AI emerges as powerful new approach. *TechTarget,* Search Enterprise AI. May 4, 2020. https://www.techtarget.com/searchenterpriseai/feature/Neuro-symbolic-AI-seen-as-evolution-of-artificial-intelligence

- **adaptabilité accrue** : la combinaison des deux approches permet aux agents de s'adapter à des environnements changeants tout en maintenant un raisonnement cohérent ;
- **performance améliorée** : des études récentes suggèrent que l'incorporation de modules symboliques externes dans un agent basé sur un *LLM* conduit à une précision moyenne améliorée dans les tâches de raisonnement logique[106].

Les agents experts IA pourraient s'appuyer sur diverses techniques, tant symboliques que neuro-symboliques, pour améliorer leurs performances et leur adaptabilité.

5.1.5. Intégration neuro-symbolique et approches hybrides

L'approche *HybridAGI* illustre la transition vers des systèmes plus sophistiqués et programmables. Cette approche repose sur l'utilisation d'un *Domain Specific Language (DSL*) pour la programmation d'agents basés sur des *LLM*, définissant ainsi la syntaxe de programmes représentés sous forme graphique. Ces programmes sont structurés sous forme de graphes d'actions, impliquant l'utilisation d'outils et des étapes de prise de décision explicites, avec des branchements au sein du graphe.

En substance, un tel système se conforme à un graphe prédéfini pour l'exécution précise des instructions. Si l'on conceptualise un programme comme un réseau, son interprétation requiert la détermination du chemin optimal au sein du graphe, offrant ainsi un cadre sécurisé et contrôlé pour une IA de type « général »[107].

5.1.6. Focus sur une technique neuro-symbolique explicable

Parmi les nombreuses techniques alternatives, on peut citer l'approche développée par *Xtractis* (Intellitech) en 2017, qui propose une IA cognitive basée sur la *Fuzzy Symbolic AI* augmentée. Cette approche se distingue par son architecture de raisonnement *humano-centrique* et sa capacité à générer des systèmes de décision transparents. La technologie *Xtractis* s'appuie sur plus de 25 ans de R&D en mathématiques floues, notamment sur la théorie des relations floues d'ordre N formulée en 1993 par Zyed Zalila (le fondateur d'Intellitech), permettant un raisonnement avec des logiques continues qui

[106] Zishen Wan et al. Towards Cognitive AI Systems: A Survey and Prospective on Neuro-Symbolic AI. *arXiv preprint arXiv: 2401.01040*. January 2024. https://arxiv.org/abs/2401.01040
[107] SynaLinks Team. HybridAGI – Programmable LLM-based Agent with Graph-based Prompt Programming *(DSL / graph programs).* PyPI, September 25, 2024. https://pypi.org/project/hybridagi/

peuvent être mieux adaptées aux problèmes complexes du monde réel que la logique binaire classique.

Cette approche d'IA collective et évolutive permet aux agents experts IA d'optimiser continuellement leurs stratégies d'induction sans intervention humaine, tout en garantissant une capacité d'audit exhaustive par les experts du domaine. Ces derniers peuvent ainsi appréhender et valider rationnellement chaque décision prise par le système, instaurant ainsi une IA de confiance par conception « *Trusted AI by design* ». *Xtractis* illustre ainsi comment une approche neuro-symbolique rigoureusement conçue peut répondre aux impératifs de fiabilité et de transparence des agents experts IA.

5.2. Autres techniques algorithmiques

Au-delà des approches à base de réseaux de neurones et symboliques, il existe une multitude d'autres techniques algorithmiques susceptibles d'être exploitées efficacement au sein des agents experts IA. Voici un aperçu de certaines de ces techniques :

- **algorithmes génétiques :** inspirés par le processus de sélection naturelle, les algorithmes génétiques sont utilisés pour optimiser et rechercher des solutions dans des espaces de recherche complexes. Ils sont particulièrement utiles pour les problèmes d'optimisation où les solutions potentielles peuvent être représentées sous forme de chromosomes ;
- **apprentissage par renforcement :** cette technique permet aux agents d'apprendre en interagissant avec leur environnement. Les agents reçoivent des récompenses ou des pénalités en fonction de leurs actions, ce qui leur permet d'apprendre des stratégies optimales pour maximiser les récompenses cumulatives ;
- **algorithmes de fouille de données :** les techniques de fouille de données, telles que l'association, la classification, et le *clustering*, permettent d'extraire des motifs et des connaissances utiles à partir de grandes quantités de données. Ces techniques sont souvent utilisées pour découvrir des tendances et des relations cachées dans les données ;
- **algorithmes d'optimisation :** les algorithmes d'optimisation, tels que la programmation linéaire, la programmation dynamique, la programmation par contraintes, et les méthodes de descente de gradient, sont utilisés pour trouver les meilleures solutions possibles à des problèmes complexes sous des contraintes spécifiques ;

- **algorithmes de recherche heuristique :** les algorithmes de recherche heuristique, comme A* et les algorithmes de recherche locale, sont utilisés pour explorer efficacement les espaces de recherche et trouver des solutions optimales ou quasi-optimales à des problèmes de recherche complexes ;
- **réseaux bayésiens** : les réseaux bayésiens sont des modèles graphiques qui représentent des ensembles de variables et leurs dépendances conditionnelles via un graphe acyclique dirigé. Ils sont utilisés pour le raisonnement probabiliste et l'inférence dans des environnements incertains ;
- **algorithmes de traitement du langage naturel (*NLP*)** : les techniques de *NLP*, telles que l'analyse syntaxique, l'extraction d'entités nommées, et la modélisation de sujets, permettent aux agents de comprendre et de générer du langage naturel, facilitant ainsi l'interaction humain-machine ;
- **algorithmes de planification :** les algorithmes de planification, tels que les planificateurs basés sur des graphes et les planificateurs hiérarchiques, sont utilisés pour générer des séquences d'actions permettant d'atteindre des objectifs spécifiques dans des environnements dynamiques et incertains (ils décomposent un but complexe en sous-tâches ordonnées).

L'application judicieuse de ces techniques algorithmiques permet aux agents experts IA de résoudre une variété étendue de problèmes complexes et de s'adapter à des environnements dynamiques et incertains. L'architecte d'agents pourra choisir et orchestrer ces méthodes en fonction du problème à résoudre et des contraintes inhérentes au monde réel, telles que le temps réel et les ressources limitées.

Synthèse du chapitre

Bien que les techniques actuellement employées dans les agents experts IA soient majoritairement issues de l'IA neuronale, il est probable que l'industrialisation de ces agents nécessitera le remplacement ou l'enrichissement de certains de leurs composants. L'objectif est de constituer un « *mix of experts* » garantissant la fiabilité des résultats tout en optimisant la consommation énergétique, et en améliorant l'explicabilité et la traçabilité afin de minimiser les risques et les impacts résiduels. Des architectures hybrides intégrant la « créativité » des *LLM*, des règles métier explicites, des méthodes d'optimisation déterministes et du raisonnement probabiliste permettront ainsi de concevoir des agents experts IA, performants, traçables et économes en ressources.

HUB FRANCE IA

6. La gouvernance des agents

6. La gouvernance des agents

La gouvernance des agents IA est devenue un enjeu crucial pour les organisations souhaitant intégrer ces technologies de manière éthique et responsable, tout en conciliant l'agilité nécessaire à leur adoption à grande échelle et à la maîtrise des nouveaux risques. La présente section examine en détail les objectifs, les responsabilités, les acteurs et les mesures de performance associés à la gouvernance de ces agents.

L'émergence des agents experts IA incite les organisations à adapter leur gouvernance existante. En effet, les organisations anticipent une démocratisation de l'usage des agents IA connectés aux données d'entreprise et manipulant des outils pour réaliser des actions automatisées[108]. Ces agents sont désormais encadrés par le RIA[109], qui rend obligatoire une analyse des risques inhérents. Par conséquent, l'évolution de la gouvernance des entreprises due au RIA sera fortement influencée par la démocratisation des agents experts IA.

Cette évolution soulève de nouvelles préoccupations liées à l'utilisation des agents, telles que :

- **le partage excessif de données :** la gestion des droits d'accès aux données d'entreprise par les organisations est souvent déficiente. L'absence de labellisation de données sensibles et l'attribution excessive de droit d'accès, notamment aux données sensibles, constituent des problèmes majeurs. Désormais, les agents experts IA utilisés par des utilisateurs dotés de tels droits excessifs peuvent accéder à l'ensemble de ces informations et les agréger, augmentant ainsi les risques d'utilisation frauduleuse ;
- **la fuite de données sensibles via les agents experts IA :** les utilisateurs demandent à ces agents de générer des rapports et des synthèses à partir des données internes. Les réponses des agents experts IA peuvent inclure des données sensibles dont

[108] World Economic Forum. Navigating the AI Frontier: A Primer on the Evolution and Impact of AI Agents. White paper. December 2024. https://reports.weforum.org/docs/WEF_Navigating_the_AI_Frontier_2024.pdf

[109] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

l'utilisateur ne perçoit pas le niveau de sensibilité. L'utilisateur partage alors largement ces informations sans connaître le niveau de sensibilité ;

- **l'utilisation non conforme des agents experts IA :** les utilisateurs, qui peuvent également être des développeurs de ces agents, peuvent les utiliser pour des actions non éthiques ou relevant des catégories à haut risque définies par le RIA ;
- **la lenteur de réponses des gouvernances existantes :** la démocratisation de l'utilisation des agents experts IA via le *no code*, accentuée par la pression exercée sur les employés pour se différencier et devenir proactifs face à l'adoption de l'IA, ne peut plus se permettre d'attendre les délais de réponse des comités et augmente le risque de *ShadowIT+AI* ;

Mais il y a une différence entre les gouvernances d'organisations humaines et d'organisations agentiques : comme le démontre l'analyse de Noam Kolt[110], les mécanismes conventionnels développés dans les gouvernances existantes tant métier que IT, régissent les conflits entre objectifs métiers et risques au niveau organisationnel humain. En effet, les agents IA ne possèdent pas d'intérêt personnel qui les pousserait à rechercher des gains financiers, et ne sont pas sensibles aux sanctions ou les détournent quand on les menace de s'éteindre. La notion d'intentions devient un élément indispensable du contexte des agents et leur surveillance doit être continue, car sous pression les agents peuvent rapidement "oublier".

Par ailleurs, les agents experts IA, dont les modèles génératifs sont souvent au cœur du processus décisionnel, constituent des vecteurs supplémentaires de propagation des risques associés aux modèles génératifs[111]. Parmi les principales menaces, on trouve :

- l'injection de *prompt* qui consiste à insérer des commandes malveillantes pour influencer le comportement de l'agent expert d'IA ;
- le *jailbreak,* qui permet de contourner les restrictions de sécurité afin d'accéder à des fonctionnalités non autorisées ;
- l'empoisonnement des données, qui manipule les données d'entraînement pour fausser les résultats de l'agent expert d'IA ;
- le détournement de modèle, qui implique la prise de contrôle des modèles d'IA à des fins malveillantes ;

[110] Noam Kolt. Governing AI Agents. *arXiv preprint arXiv: 2501.07913* January 2025. https://arxiv.org/abs/2501.07913

[111] Analyse des attaques sur les systèmes de l'IA. *Hub France IA et Campus Cyber*. Livre blanc. Mai 2025. https://www.hub-franceia.fr/wp-content/uploads/2025/05/25_05_11_Analyse-des-attaques_sur_les_systeme_de_l_IA_VF.pdf

- l'abus de portefeuille, qui exploite les systèmes de paiement intégrés utilisés par les agents experts IA ;
- l'oubli volontaire des intentions.

Il est ainsi crucial pour les organisations de mettre en place des mesures de sécurité robustes afin de protéger leurs systèmes contre ces menaces émergentes et d'adapter leur gouvernance pour devenir à la fois plus agiles et réactives, tout en étant plus prudentes face aux risques multiples et évolutifs. Chaque technique d'IA comporte ses propres risques, dont certains sont encore inconnus et qui se manifesteront bien après leurs premières mises en œuvre, et dont l'impact dépend de l'usage. Dans le premier paragraphe ci-dessous, nous présentons (de manière non exhaustive) les risques liés à l'utilisation de systèmes basés sur des agents experts IA.

6.1. Les risques des agents experts IA

Comparativement à l'usage d'un *chatbot* qui se fait toujours sous la supervision d'un humain, les risques des agents experts IA sont amplifiés par les deux caractéristiques clés des agents : la capacité d'exécuter des actions[112] et la capacité de le faire avec plus ou moins de supervision/contrôle humain (autonomie). Un agent classant automatiquement des documents comptables par exemple n'a pas le même impact qu'un agent expert IA chargé d'optimiser les achats en ayant la responsabilité, sans supervision, de déclencher des demandes d'achat. La gestion des risques autour des agents experts IA devient ainsi beaucoup plus dynamique et évolutive afin de minimiser l'impact potentiel suivant différentes dérives selon trois niveaux de granularité potentiels à avoir en tête :

1. **au niveau du ou des composants d'IA générative** utilisés (ex : *LLM* ou IA multimodale) : le Hub France IA a publié en juillet 2024 un livre blanc[113] qui détaille les différents risques des IA génératives, leurs causes et potentiels impacts suivant différentes dimensions (juridiques, financières, opérationnelles, …). Il présente une démarche d'analyse des risques génériques appliquée à différents cas d'usage dans la finance, le marketing, la cybersécurité ou d'autres secteurs d'activité. Enfin ce livre blanc introduit un certain nombre de pistes d'atténuation et de remédiation des risques ;

[112] Jinwei Hu et al. Position: Towards a Responsible LLM-empowered Multi-Agent Systems. *arXiv preprint arXiv:2502.01714*. February 2025. https://arxiv.org/pdf/2502.01714

[113] Hub France IA. Les risques de l'IA générative. Livre blanc. Juillet 2024. https://www.hub-franceia.fr/wp-content/uploads/2024/09/Hub-France-IA-Les-risques-de-lIA-Generative-final.pdf

2. **au niveau d'un agent expert IA** on peut citer[114] :

- de **potentielles fuites d'informations sensibles** : les fuites de confidentialité surviennent lorsque les agents, en raison de leurs interactions avec des applications, demandent des informations personnelles sensibles, ce qui augmente le risque d'extraction de données. Les agents experts IA doivent gérer les sessions pour maintenir la confidentialité et l'intégrité des données échangées entre les utilisateurs et le serveur. La gestion des sessions et des droits est complexe, car les agents doivent suivre les interactions des utilisateurs dans le temps. Si cette gestion est inadéquate, cela peut entraîner des fuites d'informations et des assignations d'actions erronées ;
- des **défaillances dans les protocoles de sécurité** : un agent expert IA peut avoir des accès/privilèges plus importants qu'un utilisateur (voire un attaquant) et donner à ce dernier la possibilité d'effectuer des actions non autorisées. Les droits d'accès de l'agent expert IA peuvent être attachés à un utilisateur et si ce dernier change de fonction dans une même entreprise sans que ses anciens droits aient été supprimés, cela peut engendrer des failles de sécurité. L'isolation des privilèges, la limitation des droits de l'agent à ceux de l'utilisateur pour lequel il agit et l'identification des requêtes légitimes sont des méthodes permettant de diminuer ce risque de faille dans les protocoles de sécurité, notamment en cas d'attaque basée sur l'injection de prompt ;
    - **l'empoisonnement de la mémoire :** désigne une technique visant à manipuler les systèmes de mémoire à court terme ou à long terme d'un agent. Cette méthode consiste à injecter des données falsifiées ou malveillantes afin d'influencer le contexte et de manipuler l'agent (ex : altération des décisions, actions non autorisées) ;
    - **une vulnérabilité accrue aux attaques par déni de service** (*DoS*) : en raison de la charge computationnelle potentiellement élevée des requêtes et des appels répétés à des outils par l'agent ;
    - **l'utilisation malintentionnée d'outils :** par des cybercriminels, notamment via l'exécution de code à distance (*Remote Code Execution*) ;
    - **des erreurs dans l'exécution des tâches ou d'appels aux outils adéquats :** car l'utilisation d'une IA générative comme composant pour interpréter et générer des instructions introduit éventuellement des imprécisions impactant le fonctionnement des outils (ex : formats attendus des entrées, paramètres manquants dans l'appel de l'outil, ...). Un agent pourrait appeler par erreur un

[114] La liste n'est pas exhaustive

outil qui effectuerait une tâche potentiellement dangereuse (ex : écriture dans une base de données par exemple) ;

- **l'incompréhension des** contextes : les agents experts IA peuvent mal interpréter des contextes ou ne pas disposer d'informations suffisantes, entraînant des actions inappropriées ou dangereuses ;
- **la rapidité d'exécution** : en raison de sa vitesse, l'IA agentique peut s'engager très rapidement dans une direction contraire aux attentes. Cette amplification du risque par la vitesse nécessite des mécanismes de surveillance en temps réel et des garde-fous automatisés ;

3. **au niveau d'un système multi-agent**, on peut citer[115] :
   - **la propagation des hallucinations (dérive de connaissances et propagation d'erreurs)** : lorsque les modèles d'IA générative intégrés aux agents experts génèrent des informations incorrectes, les erreurs peuvent se propager entre ces agents. Dans les tâches de raisonnement collaboratif, les agents experts IA peuvent s'aligner sur un consensus erroné en raison de phénomènes comme la conformité et le biais d'autorité. Par exemple, dans des débats entre agents experts IA, un agent expert ayant une compréhension erronée peut générer des raisons persuasives mais fausses, impactant les autres et les détournant des chemins de raisonnement vers des solutions correctes. Ce problème constitue une dérive de connaissances ;
   - **la propagation de l'incertitude :** à mesure que les systèmes deviennent plus complexes, les incertitudes inhérentes aux agents experts IA individuels peuvent s'accumuler, compromettant potentiellement la stabilité globale du système. Les agents experts IA montrent également une tendance à l'expansion des biais cognitifs, amplifiant et propageant les erreurs plutôt que de les filtrer, ce qui aggrave la dérive de connaissances.[116] Les solutions existantes, telles que l'ingénierie des prompts et les interventions de type « humain dans la boucle », sont souvent limitées en termes d'évolutivité et de praticité. Une piste pour résoudre ces problèmes consiste à utiliser une architecture intégrant des mécanismes de quantification de l'incertitude dans ses principes opérationnels, garantissant un alignement cohérent des connaissances à travers le réseau d'agents. Il existe plusieurs sources d'incertitudes : **l'incertitude inhérente à**

[115] La liste n'est pas exhaustive

[116] Moshe Glickman, Tali Sharot. How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour* 9.2, *p* 345–359. 2025. https://www.nature.com/articles/s41562-024-02077-2

chaque agent expert et **l'incertitude dans les interactions entre agents experts IA**. La performance des systèmes multi agents (de type *SMA-LLM*)[117] peut être évaluée par des métriques statistiques ou grâce à une vérification humaine (de type « humain dans la boucle »). Identifier les résultats en sortie du système ayant un niveau d'incertitude associée élevé pour une évaluation ultérieure par un humain contribue à renforcer la robustesse des systèmes multi-agents, à condition que le nombre d'éléments incertains reste raisonnable et appréhendable par les opérateurs humains ;

- **l'intelligence collective** : l'un des défis est d'atteindre une compréhension mutuelle entre agents experts IA pour maximiser l'intelligence collective. Contrairement aux SMA traditionnels, qui reposent sur des protocoles prédéfinis, ces agents basés sur des LLMs présentent des comportements émergents et imprévisibles en raison de leur fonctionnement et de leur entraînement sur des ensembles de données vastes et variés. Cette imprévisibilité nécessite le développement de mécanismes quantifiables, tels que des **métriques de confiance**, pour faciliter une coordination efficace entre agents experts. Sans ces mécanismes, les agents experts IA peuvent avoir du mal à interpréter ou à s'aligner sur les actions des autres humains ou agents. La notion de **coopération** dans les SMA-LLMs (coopération agent-agent) est essentielle pour garantir la cohérence responsable et opérationnelle. La coopération se manifeste par la capacité des agents experts IA à traiter les intentions et les sorties des autres agents. Elle peut reposer sur des capacités de délégation : des modèles avec des capacités de raisonnement plus élevées guident des modèles plus faibles en leur déléguant des tâches. Pour atteindre et évaluer efficacement la coopération globale dans un système multi-agents, des méthodes d'évaluation dédiées sont essentielles[118]. Des métriques telles que les ratios de coopération et de coordination, les scores de confiance et la similarité sémantique sont proposées pour évaluer la qualité de la collaboration entre agents ;
  - lorsque la coopération n'est pas souhaitée, il y a un risque de **collusion**. La collusion peut survenir à la fois des communications entre agents et des mécanismes internes des agents individuels.
- les **conflits entre agents** dans les SMA-LLMs proviennent généralement d'un désalignement des objectifs et d'une asymétrie des connaissances. Les conflits

[117] LLM : *Large Language Model*. SMA-LLM : Système multi agents basé sur des *Large Language Models*
[118] Taicheng Guo et al. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*. 2024. https://arxiv.org/pdf/2402.01680

au niveau des objectifs peuvent émerger d'interprétations divergentes d'un même objectif de haut niveau, entraînant des stratégies d'exécution divergentes. Les conflits basés sur les connaissances surviennent lorsque les agents experts IA construisent des modèles mentaux différents malgré des informations initiales identiques. La nature probabiliste des LLMs et les ambiguïtés sémantiques inhérentes amplifient les effets du désalignement des connaissances.

- la **compréhension agent-humain** : pour opérer avec des humains dans la boucle (**coopération agent-humain**), les agents experts IA doivent interpréter avec précision le langage naturel et le contexte (ex : contraintes sociétales). Des méthodes telles que l'apprentissage par renforcement à partir de retours humains (RLHF[119]), le perfectionnement supervisé (SFT[120]) et l'optimisation des préférences (PO[121]) sont couramment utilisées pour aligner les sorties des agents/LLM sur les valeurs humaines. Cela nécessite de travailler aussi sur la désambiguïsation du langage notamment lorsque le langage est très spécifique à un métier et donc potentiellement soumis à mésinterprétation par les LLM entrainés sur des corpus généralistes ;
- les **attaques par empoisonnement de données et le jailbreaking** : les attaques par empoisonnement de données et le jailbreaking introduisent des vulnérabilités dans les SMA-LLMs en exploitant les canaux de communication. La dépendance à des interactions dynamiques et à des outils/connaissances externes élargit les surfaces d'attaque, nécessitant des mécanismes dédiés pour détecter et filtrer les données compromises ;
  - les **menaces cyber** sur les architectures distribuées : les menaces cyber posent également des défis importants aux SMA-LLMs en raison de leur architecture distribuée. Par exemple, des attaques au niveau du réseau peuvent perturber les performances ;

[119] RLHF : *Reinforcement Learning from Human Feedback* (apprentissage par renforcement à partir de rétroaction humaine).
Daniel M. Ziegler et al. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*. 2019. https://arxiv.org/pdf/1909.08593

[120] SFT : *Supervised Fine-Tuning*. Il consiste à adapter des modèles de langage pré-entraînés à des tâches spécifiques, en les entraînant sur un ensemble de données spécifique à la tâche ou domaine, avec des exemples étiquetés.

[121] PO : *Preference Optimization*. Cette technique consiste à aligner le modèle de langage grâce à une base de données contenant pour chaque *prompt* une réponse préférée et une non préférée.

- le **scaling up** : lorsque le nombre d'agents IA dans un système augmente, cela conduit à une augmentation de la complexité calculatoire. Le système a besoin d'une puissance de calcul et de mémoire accrues. L'orchestration des agents devient clé ;
- le potentiel manque de **traçabilité** (effet « boîte noire ») et de **reproductibilité** des décisions dans les SMA-LLMs représente enfin un risque à prendre en compte lors de l'utilisation de ce type de système.

6.2. Objectifs de la gouvernance des agents experts IA

6.2.1. Assurer l'éthique et la conformité

L'un des principaux objectifs de la gouvernance des agents experts IA est de garantir que leur utilisation respecte les normes éthiques et les réglementations en vigueur, tout en apportant un réel gain de productivité pour l'entreprise ou la collectivité, sans pour autant nuire à l'environnement. La gouvernance a donc pour rôle de conserver cet équilibre délicat, qui évoluera sans doute, et cela inclut la protection des données personnelles et la transparence dans les processus décisionnels des agents experts IA.

Le RIA interdit l'utilisation d'agents qui présentent un risque inacceptable pour les droits fondamentaux des individus. Ces agents incluent :

- **les agents de notation** sociale : l'évaluation des individus basée sur leur comportement social, susceptible d'entraîner des discriminations ;
- **les agents utilisant l'identification biométrique en temps réel** : l'utilisation des technologies de reconnaissance faciale dans des espaces publics sans consentement ;
- **les agents utilisant la** manipulation **comportementale** : les systèmes qui exploitent les vulnérabilités des individus pour influencer leurs décisions, notamment dans des contextes tels que le recrutement ou l'éducation[122].

Au-delà des interdictions, le RIA exige que chaque projet d'IA évalue la nature des risques et mette en place des mécanismes de gestion de ces risques. Or dans le cas de l'IA agentique, il sera nécessaire d'intégrer à la gouvernance une validation lors de l'industrialisation, dès le prototype réalisé par un utilisateur non spécialiste. Pour les agents les plus impactants, un suivi dans la durée tout au long de leur utilisation sera

[122] Virginie Dignum. Responsible Artificial Intelligence: Designing AI for Human Values. *ITU Journal: ICT Discoveries*, Special Issue, No. 1, p. 1-8. September 25 2017. https://www.itu.int/dms_pub/itu-s/opb/journal/S-JOURNAL-ICTF.VOL1-2018-1-P01-PDF-E.pdf

également requis, tout en veillant à conserver un bon équilibre vis-à-vis de l'humain. Le récent retour de la FinTech *Klarna* de remplacement des employés par des agents IA démontre combien cette recherche d'équilibre est complexe[123].

Cette nécessité d'adaptation est d'autant plus critique que les retours d'expérience internationaux révèlent des défis organisationnels majeurs. Selon Harvard Business School[124], 61% des organisations rapportent une anxiété croissante des employés concernant l'impact des agents IA sur l'emploi, nécessitant une approche de gestion du changement structurée combinant *mindsets, skill sets*, et *tool sets*.

6.2.2. Promouvoir la transparence

La transparence est essentielle pour instaurer la confiance parmi les parties prenantes. Les organisations doivent s'assurer que les décisions prises par les agents experts IA sont compréhensibles et justifiables notamment lorsque ces décisions impactent des humains ou leur organisation. Cela nécessite une documentation claire des critères algorithmiques et des données utilisés : ces informations ne sont pas toujours disponibles (notamment pour les *LLM*), le RIA prévoit des exigences sur ce point. Il est également crucial de mettre en place des mécanismes de transparence pour informer les utilisateurs sur le fonctionnement des agents experts IA, ainsi que d'assurer un accompagnement sociotechnique, notamment via un dialogue social plus ouvert, c'est-à-dire une écoute franche et ouverte des acteurs impactés par l'utilisation de ces agents.

6.2.3. Assurer la qualité des données

La démocratisation des agents experts IA accédant à l'ensemble des données utilisateurs pour effectuer de multiples tâches soulève des enjeux cruciaux de gouvernance de la donnée, notamment en matière de sécurisation des données. Cette sécurisation est essentielle pour limiter les erreurs des agents IA (hallucinations) et la diffusion d'informations sensibles.

Plusieurs actions sont à mener pour atteindre cet objectif :

---

[123] Frédéric Charles. Klarna nous montre les limites des agents IA. *ZDNet*. 11 mai 2025. https://www.zdnet.fr/blogs/green-si/klarna-nous-montre-les-limites-des-agents-ia-474924.htm.

[124] Manuel Hoffmann et al. Generative AI and the Nature of Work. *Harvard Business School*. Working Paper 25-021. October 27, 2024. https://www.hbs.edu/ris/Publication%20Files/25-021_49adad7c-a02c-41ef-b887-ff6d894b06a3.pdf

- **connaître ses données** : il est essentiel de comprendre la nature, l'origine et l'utilité des données. Les bonnes pratiques incluent l'inventaire des données, leur classification selon leur sensibilité et l'utilisation d'outils de gestion des métadonnées intégrant le « contexte » de collecte, de traçabilité, de traitement et de filtrage de ces données ;
- **assurer la pérennité des données :** cela implique de maintenir la pertinence et l'utilisabilité des données tout au long de leur cycle de vie. Les pratiques recommandées incluent la définition de politiques de conservation, la mise à jour et le nettoyage des données, ainsi que l'utilisation de technologies de stockage appropriées ;
- **protéger ses données** : la sécurité des données est primordiale pour prévenir les accès ou les utilisations non autorisés. Les bonnes pratiques incluent la mise en œuvre de mesures de sécurité telles que le chiffrement, la réalisation d'audits de sécurité réguliers et la sensibilisation des employés aux risques. Il convient également de tenir compte de la propriété des données, notamment celles acquises et donc soumises à une licence d'utilisation, qui peuvent être exposées à des contraintes contractuelles d'utilisation d'agrégation et/ou de revente ;
- **prévenir la perte des données** : il est crucial de mettre en place des stratégies pour éviter la perte de données, qu'elle soit due à des erreurs humaines ou à des incidents techniques. Cela comprend des systèmes de sauvegarde réguliers, des plans de reprise d'activité et des tests des procédures de récupération.

Certaines organisations, comme La Poste par exemple, envisagent la mise en place de données synthétiques pour réaliser des démonstrations de preuve de concept (*PoC*) sans données sensibles. L'objectif est de créer des jeux de données fictifs mais représentatifs, non traçables jusqu'aux données réelles. Cela permet aux services de progresser sur un *PoC* tout en préparant et sécurisant les données réelles pour l'industrialisation.

6.2.4. Définir des intentions pour éclairer les décisions des agents experts IA et conserver le contrôle

Certains types d'agents, notamment ceux relevant des catégories délibératives, de classification ou autres, prennent et exécutent des décisions. Ils nécessitent donc une connaissance approfondie du contexte dans lequel ils agissent afin d'éviter des prises de décisions hors contexte. Il est donc nécessaire que, sous l'égide des organes de gouvernance, un travail soit entrepris afin d'accompagner l'adoption et le déploiement

à grande échelle des agents experts d'AI dans des *frameworks* de gouvernance. Ce travail devrait inclure les actions suivantes :

- **transcrire la raison d'être et la mission de l'organisation :** en règles contextuelles utiles aux agents de prise de décisions, et traduire les contraintes réglementaires en doctrines claires ;
- **hiérarchiser les objectifs** : productivité, agilité, acceptabilité sociale et sociétale, impact environnemental, afin que les intentions soient clairement comprises et intégrées par les agents experts IA ;
- **identifier de nouvelles métriques et seuils d'alerte** : notamment en matière de bien-être et d'acceptabilité, et négocier les seuils d'alerte au sein d'un dialogue social élargi. Cela correspond à la notion de visibilité (voir note 110) et de transparence.
- Retranscrire **en bonnes pratiques :** les découvertes, les compromis vertueux et les leçons apprises, et les partager au sein de groupes de travail intra et inter-entreprises.

Le principe de responsabilité humaine reste fondamental : comme l'illustre l'analogie du pilote d'avion, si un avion vole en pilote automatique avec le pilote qui surveille depuis son siège, qui est responsable de l'atterrissage sûr ? Le pilote. Cette responsabilité s'applique identiquement aux agents IA, indépendamment de leur niveau d'autonomie.

6.3. Responsabilités de la gouvernance des agents experts IA

6.3.1. Évaluation et audit des systèmes d'IA

L'une des responsabilités fondamentales de la gouvernance des agents experts IA est l'évaluation et l'audit de ces derniers. Cela implique la mise en place de processus d'audit réguliers pour examiner les résultats produits par ces agents, notamment le niveau d'hallucination et l'acceptabilité. Ces audits permettent d'identifier les problèmes potentiels et d'assurer la conformité avec les besoins métiers, les normes éthiques et réglementaires.

Les exigences du RIA se chevauchent souvent avec celles du RGPD. Les entreprises doivent intégrer des principes de « *Privacy by Design* » dans leurs agents, effectuer des évaluations d'impact pour les agents experts IA à haut risque et maintenir une documentation claire de protection des données.

6.3.2. Évaluation et gestion des risques

Les entreprises doivent effectuer des évaluations de risques pour classer leurs agents en fonction de leur niveau de risque (inacceptable, élevé, limité, minimal) mais aussi en fonction de l'impact potentiel. Cela permet de prioriser les efforts de conformité et de s'assurer que les agents experts IA à haut impact soient conformes aux exigences du RIA, mais aussi au bon fonctionnement de l'entreprise de manière durable.

La gestion des risques associés à l'utilisation des agents experts IA est une autre responsabilité cruciale[125]. Les organisations doivent identifier les risques, tels que les biais algorithmiques, la sécurité des données et les impacts notamment sur la vie privée. Cela nécessite l'élaboration de stratégies pour atténuer ces risques, y compris des évaluations d'impact régulières et des mesures de sécurité renforcées.

Le risque de *Shadow AI* tournant avec les agents experts IA en un risque de *Shadow IT* déstructure non seulement l'architecture des SI, mais aussi les process d'une entreprise. Cela nécessite une adaptation constante et une grande vigilance sur les impacts de la part de toutes les gouvernances pas seulement celles liées à l'*IT*.

6.3.3. Formation et sensibilisation pour une meilleure transparence

Une autre responsabilité importante est la formation et la sensibilisation des employés aux pratiques éthiques et responsables liées à l'utilisation de l'IA. Les organisations doivent s'assurer que leurs équipes comprennent les implications éthiques de l'IA et sont formées pour utiliser ces technologies de manière responsable.

Pour « automatiser l'ordinaire et humaniser l'extraordinaire » dans une méthodologie de déploiement organisationnel, l'expérience de Visteon, par exemple, illustre une approche structurée en trois étapes :

1. Réunions plénières pour sensibiliser l'ensemble de l'organisation ;
2. *Newsletter* mensuelle détaillant les cas d'usage et bénéfices ;
3. Communication hiérarchique avec évaluation de performance intégrant l'adoption de l'IA.

[125] Owasp. Agentic AI - Threats and Mitigations - OWASP Top 10 for LLM Apps & Gen AI. *Owasp*. 2025. https://genai.owasp.org/resource/agentic-ai-threats-and-mitigations/

6.3.4. Engagement et responsabilités des parties prenantes

Les responsabilités ne sont jamais totalement transférées aux agents experts IA. *A contrario* on ne peut rendre responsable un humain que de ce qu'il peut maitriser et avoir la capacité d'agir dessus. Donc, la gouvernance des agents experts IA doit exiger des projets IA l'engagement des parties prenantes, notamment les employés, les clients et/ou la société civile et accompagner un juste partage des responsabilités entre technique et humain, suivant la visibilité et la capacité d'agir. Cela implique, non seulement de consulter ces groupes, mais de les impliquer lors de l'élaboration pour chaque solution ainsi que des politiques et de pratiques de gouvernance, afin de s'assurer que les préoccupations et les attentes de tous soient prises en compte dans chaque projet d'IA.

6.3.5. Mutualisation des agents experts IA

La gouvernance des agents experts IA dans les grandes entreprises fait face à un arbitrage clé : développer des agents spécialisés ou mutualiser des modèles généralistes. Les agents spécialisés, bien que très performants pour une tâche précise, engendrent des coûts de maintenance et de surveillance élevés pour chaque instance et ne profitent pas des mises à jour de fiabilité des autres projets similaires.

À l'inverse, un agent générique réutilisable permet de rationaliser les investissements initiaux. Cependant, l'adapter à de nouveaux contextes pour garantir sa fiabilité et sa performance engendre des surcoûts significatifs en conception et en tests.

L'enjeu est donc de trouver un équilibre stratégique, en définissant quand privilégier la spécialisation et quand opter pour la mutualisation, afin de maîtriser les coûts tout en alignant l'IA sur les objectifs de l'entreprise.

Dans cet exercice d'équilibre, certaines grandes entreprises vont jusqu'à créer une filiale dédiée à la création d'agents experts IA et permettre ainsi d'en faire des produits générant de nouveaux revenus. Tout à fait en ligne avec la vision du web agentique portée par le projet *NANDA* du *MIT*[126].

[126] https://nanda.media.mit.edu

6.4. Acteurs et mesures de performance de la gouvernance des agents experts IA

6.4.1. Acteurs de la gouvernance

Les principaux acteurs impliqués dans la gouvernance des agents experts IA sont les équipes de direction, les responsables de la conformité, les développeurs et concepteurs d'IA, ainsi que des représentants des parties prenantes externes.

Les équipes de direction sont responsables de la définition des stratégies et des politiques de gouvernance, tandis que les développeurs doivent intégrer des considérations éthiques dès la phase de conception.

L'IA agentique représente un changement organisationnel plus important encore que l'arrivée de l'agilité par rapport au cycle en V dans le monde du développement. L'agilité en informatique était encore optionnelle, notamment dans certains secteurs très réglementés. Or l'IA agentique, la pression concurrentielle et l'incertitude géopolitique imposent un changement de posture radical des gouvernances très hiérarchisées afin d'acquérir davantage d'agilité, de fluidité et même de collaboration entre les différentes instances de gouvernance existantes.

Le secteur bancaire, utilisateur majeur d'IA classique, est indéniablement précurseur dans l'organisation de sa gouvernance IA, notamment pour se conformer à des réglementations telles que la *SR 11-7*[127]. Cette réglementation est cruciale car elle garantit la précision, la fiabilité et la gestion rigoureuse des modèles quantitatifs utilisés par les institutions financières pour la prise de décision. Le non-respect de *SR 11-7* peut entraîner des sanctions réglementaires, des dommages à la réputation et des pertes financières liées à l'utilisation de modèles défectueux.

Afin de se conformer à la réglementation *SR 11-7* sur la gestion des risques liés aux modèles, les banques ont mis en place une structure organisationnelle fondée sur trois lignes de défense :

- **première ligne de défense** : elle regroupe les développeurs et propriétaires de modèles ainsi que les unités opérationnelles. Ils sont responsables du développement, de l'implémentation et de la gestion quotidienne des modèles ;

[127] Board of Governors of the Federal Reserve System, Washington, D.C. 20551, Division of Banking Supervision and Regulation SR 11-7 April 4, 2011.
https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm

- **deuxième ligne de défense** : elle est constituée des équipes de validation indépendante et de gouvernance des modèles. Ces fonctions établissent les normes, supervisent et coordonnent la gestion des risques sans être impliquées dans le développement ;
- **troisième ligne de défense :** représentée par l'audit interne, elle évalue de façon indépendante l'efficacité du cadre de gestion des risques et vérifie la conformité des deux premières lignes aux politiques et exigences réglementaires.

Cette structure tripartite assure une séparation claire des responsabilités, garantit la transparence et l'efficacité dans la gestion des risques de modèles, et permet une supervision adéquate par la direction générale et le conseil d'administration, conformément aux exigences de la *SR 11-7*.

La Société Générale a par exemple déployé ce principe de trois lignes de défense pour gérer les risques des cas d'usage basés sur des IA.

L'expérience montre que 80% des politiques nécessaires aux agents IA existent déjà (sécurité des données, certification, due diligence fournisseurs). Seuls 20% sont spécifiques à l'IA (hallucinations, dérive des modèles, biais). Cette approche modulaire évite la bureaucratie excessive tout en maintenant la gouvernance nécessaire.

6.4.2. Indicateurs de conformité et indicateurs nouveaux

Les organisations doivent établir des indicateurs de conformité pour évaluer le respect des réglementations et des normes éthiques. Cela inclut la vérification de la conformité avec des lois telles que le RGPD et d'autres réglementations pertinentes. Des indicateurs nouveaux, notamment de bien-être, d'impact social et sociétal, doivent être introduits, mais avec la contrainte imposée par le RIA de ne pas utiliser les émotions à l'encontre des employés[128].

6.4.3. Évaluations de l'impact

Les évaluations de l'impact mesurent l'influence des systèmes d'IA (SIA) sur les processus organisationnels et les parties prenantes. Elles peuvent inclure des analyses sur la manière dont les décisions prises par l'IA influencent les résultats commerciaux, la satisfaction des clients et l'engagement des employés. Des indicateurs d'impact

[128] Hub France IA. Fiche AI Act – Définitions clés du Hub France IA. Septembre 2024. https://www.hub-franceia.fr/wp-content/uploads/2024/09/Definitions-AI-Act_Analyse_Hub_France_IA.pdf

environnementaux seront également intégrés à l'évaluation régulière de la bonne utilisation de certaines techniques d'IA, notamment celles moins frugales.

6.4.4. Mesures de la qualité des données

La qualité des données est un aspect crucial de la gouvernance des agents experts IA. Les organisations doivent surveiller des indicateurs tels que l'exactitude, l'exhaustivité et la pertinence des données utilisées pour entraîner les modèles d'IA[129]. Il ne s'agit plus de la qualité des données traditionnellement requise pour la comptabilité. L'IA peut tolérer des critères de qualité de données moins stricts, mais cela dépend évidemment des techniques d'IA et de l'impact des défauts de qualité.

6.4.5. Suivi des biais algorithmiques

Le suivi des biais algorithmiques est essentiel pour garantir l'équité et l'objectivité des SIA. Les organisations doivent mettre en place des mécanismes pour détecter et atténuer les biais dans les modèles d'IA, en utilisant des métriques spécifiques pour évaluer l'équité des résultats produits par ces systèmes et leurs impacts, notamment en fonction des « intentions » et objectifs poursuivis.

6.4.6. Satisfaction des parties prenantes

La satisfaction des parties prenantes, y compris des employés et des clients, est un indicateur important de la performance de la gouvernance des agents experts IA. Des enquêtes et des retours d'expérience peuvent être utilisés pour évaluer la perception des utilisateurs concernant la transparence, le partage des responsabilités et l'efficacité des systèmes d'IA.

Synthèse du chapitre

On constate que certaines organisations intègrent à leur gouvernance générale une gouvernance spécifique à l'IA générative et aux agents experts IA. La gouvernance de ces agents représente un domaine complexe qui exige une attention particulière de la part des organisations.

---

[129] Reuben Binns. Fairness in Machine Learning: Lessons from Political Philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency PMLR*, p. 149-158. 2018. https://proceedings.mlr.press/v81/binns18a/binns18a.pdf

Une adoption à grande échelle de l'IA agentique ne peut se faire sans une gouvernance rigoureuse, ne se limitant pas aux aspects de retour sur investissement, et ne tenant pas compte des risques résiduels à gérer dans le temps. Ces risques incluent une part intangible liée à l'acceptation des changements organisationnels induits, souvent sous-estimée. Cependant, en définissant des objectifs clairs, des responsabilités bien établies, des métriques complémentaires, non seulement financières, mais aussi relatives à l'acceptabilité, à l'éthique et à l'impact environnemental, en impliquant les acteurs concernés et en mesurant tous les aspects de la performance de manière rigoureuse, les organisations peuvent évoluer efficacement dans le contexte complexe de l'IA agentique tout en assurant une utilisation agile, éthique et responsable de ces technologies.

HUB FRANCE IA

7. Conclusion générale

7. Conclusion générale

7.1. L'avènement d'une nouvelle ère technologique et organisationnelle

L'analyse exhaustive des agents experts IA présentée dans ce livre blanc met en lumière une transformation technologique significative qui transcende la simple évolution des outils informatiques. Nous assistons à l'émergence d'un nouveau paradigme où l'IA ne se limite plus à l'assistance des utilisateurs dans l'exécution de leurs tâches. Elle s'impose comme un acteur autonome, capable d'interagir avec ses pairs et de prendre des décisions complexes dans des environnements dynamiques et imprévisibles.

Cette révolution repose sur la convergence de plusieurs avancées technologiques majeures : la maturité des *LLM*, le développement d'architectures multi-agents sophistiquées, l'intégration de techniques neuro-symboliques permettant un raisonnement explicable et transparent, et l'émergence de protocoles d'échanges standardisés qui facilitent l'interopérabilité entre systèmes hétérogènes. Les cas d'usage illustrés dans ce document, qu'ils se rapportent aux secteurs de la santé, de la finance, de l'industrie ou de l'administration publique, témoignent du passage d'une phase d'expérimentation à une phase de préparation active pour un déploiement opérationnel à grande échelle.

Néanmoins, cette transformation s'accompagne de défis considérables qui exigent une approche réfléchie et responsable. Les risques identifiés, tels que les hallucinations, les biais algorithmiques, les problèmes de sécurité et les impacts environnementaux, ne constituent pas de simples obstacles techniques à surmonter, mais des enjeux systémiques qui remettent en question nos modèles organisationnels, nos processus de gouvernance et notre relation au travail humain.

7.2. Le dilemme de la productivité : promesses et périls

Les entreprises, qu'elles soient de grande envergure ou de taille moyenne, se trouvent confrontées à un dilemme complexe. D'une part, le potentiel de gains de productivité annoncé par les agents experts IA est considérable et amplement documenté par de nombreux cas d'usage présentés dans ce livre blanc. L'automatisation de tâches complexes, l'optimisation des processus décisionnels, la personnalisation à grande échelle des services et la capacité à traiter des volumes de données sans précédent promettent des avantages concurrentiels significatifs.

Les exemples concrets analysés, tels que l'assistant IA de *Klarna* gérant les deux tiers des conversations du service client et les agents de génération automatique de contenu publicitaire, démontrent que ces gains ne relèvent plus de la prospective mais constituent une réalité opérationnelle. Pour les PME en particulier, ces technologies offrent l'opportunité d'accéder à des capacités d'analyse et d'automatisation jusqu'alors réservées aux grandes entreprises, créant ainsi un potentiel de démocratisation technologique inédit.

Cependant, cette course à la productivité s'accompagne de risques systémiques qui ne peuvent être ignorés. La dépossession progressive du savoir-faire constitue l'un des dangers les plus insidieux. Lorsque les agents prennent en charge des processus complexes, les collaborateurs risquent de perdre progressivement leur compréhension fine des métiers, engendrant une dépendance technologique qui peut s'avérer problématique en cas de dysfonctionnement ou d'évolution des besoins.

La dépréciation des compétences humaines constitue un défi majeur dans les secteurs où l'expertise métier représente un avantage concurrentiel durable. Si les agents experts IA excellent dans l'optimisation des processus existants, ils demeurent insuffisants en matière de créativité, d'intuition et de capacité d'adaptation, qualités intrinsèquement humaines face à des situations inédites.

Cette déshumanisation peut être analysée sous deux autres angles, dont les implications méritent une attention particulière : le risque d'emploi pour les jeunes diplômés et la potentielle convergence vers un contenu homogène. Concernant le premier aspect, les effets se manifestent déjà : bénéfiques pour les employés expérimentés, dont l'expertise est consolidée par les entreprises, mais préjudiciables pour les jeunes diplômés accédant à l'emploi dans certains secteurs, tels que le développement logiciel. Le risque lié aux contenus est corrélé au rythme et au volume de contenus générés par l'IA qui intègrent les corpus de données d'entraînement des générations suivantes (IA de générations N+1). Plus la proportion de ces contenus issus de l'IA augmentera, plus la visibilité des productions humaines diminuera, au profit de contenus fades, stéréotypés et homogènes.

Plus préoccupant encore, la perte de contrôle sur les processus critiques expose les organisations à des risques opérationnels majeurs. Les phénomènes d'hallucination des modèles génératifs, la propagation d'erreurs dans les systèmes multi-agents, ainsi que la difficulté d'audit et la rapidité des décisions prises par des systèmes complexes engendrent des zones d'incertitude susceptibles d'avoir des conséquences dramatiques dans des secteurs sensibles.

7.3. L'impératif du dialogue social et de la transformation des métiers

L'adoption généralisée des agents experts IA requiert une transformation significative des pratiques de dialogue social au sein des organisations. L'absence de concertation avec les parties prenantes – collaborateurs, représentants du personnel, clients et partenaires – concernant l'intégration de ces technologies représente un facteur de risque majeur, tant sur le plan humain qu'opérationnel.

Les entreprises doivent impérativement concevoir de nouveaux cadres de dialogue permettant d'associer l'ensemble des acteurs concernés à la définition des usages, des limites et des modalités de contrôle des agents experts IA. Cette démarche participative ne constitue pas seulement une exigence éthique, elle représente également un facteur clé de succès pour l'acceptation, l'efficacité de ces technologies, ainsi que pour l'optimisation de leur utilisation, notamment dans la réinvention de modèles d'affaires ou de processus existants.

La transformation des métiers constitue un défi particulièrement complexe nécessitant une approche proactive et structurée. Plutôt que de subir une automatisation, les organisations doivent anticiper l'évolution des rôles humains afin de créer une valeur ajoutée complémentaire à celle des agents experts IA. Cela implique de repenser les fiches de poste, les parcours de formation et les critères d'évaluation pour intégrer des dimensions de supervision, de créativité et de relation humaine que les agents ne peuvent pas remplacer.

L'émergence de nouveaux métiers – superviseurs d'agents, architectes de systèmes multi-agents, spécialistes de l'éthique algorithmique – offre des opportunités de reconversion et d'évolution professionnelle, à condition que les organisations investissent massivement dans la formation et l'accompagnement de leurs collaborateurs.

7.4. La nécessité d'un contrôle humain instrumenté

L'autonomisation croissante des agents experts IA exige la mise en œuvre d'un contrôle humain instrumenté, concept dépassant la simple supervision pour englober la création de mécanismes de gouvernance. Ces derniers permettent aux acteurs humains de comprendre, d'orienter et de corriger les décisions prises par les agents, tout en préservant les avantages de l'automatisation.

Cette approche requiert le développement d'outils de visualisation et d'analyse sophistiqués, rendant transparents les processus décisionnels des agents. Les techniques d'IA explicable, les tableaux de bord de monitoring en temps réel et les systèmes d'alerte proactive constituent autant d'instruments permettant aux équipes humaines de maintenir une compréhension et un contrôle effectifs sur les systèmes automatisés.

La planification de ce contrôle humain doit être intégrée dès la phase de conception des systèmes d'agents, et non ajoutée ultérieurement. Cela implique la définition de points de contrôle critiques, de seuils d'alerte, de procédures d'escalade et de mécanismes de reprise en main manuelle, garantissant ainsi que l'autonomie des agents demeure toujours subordonnée aux objectifs et aux valeurs de l'organisation.

7.5. Les risques de fragmentation des systèmes d'information

L'adoption généralisée d'agents spécialisés soulève une préoccupation majeure quant à l'architecture des systèmes d'information d'entreprise. La tentation de décomposer les processus métiers en micro-tâches gérés par des agents peut engendrer une fragmentation excessive, compromettant ainsi la cohérence globale des opérations.

Les grands progiciels de gestion intégrée (*ERP, CRM, SCM, GMAO*) ont été conçus pour assurer la cohérence des données et des processus à l'échelle de l'entreprise. Leur remplacement par une multitude d'agents spécialisés risque de créer des silos technologiques et fonctionnels, complexifiant la gouvernance, réduisant la visibilité globale et multipliant les risques de dysfonctionnement. Il est fort probable que le *Shadow AI* s'accompagne d'un risque de *Shadow IT* sans précédent pour les architectes de systèmes d'information.

Cette fragmentation peut également entraver la capacité des organisations à maintenir une **vision stratégique cohérente**. Lorsque chaque processus est optimisé localement par des agents spécialisés, il devient ardu de garantir l'alignement global sur les objectifs de l'entreprise et de détecter les effets de bord entre les différents domaines d'activité.

Les décideurs doivent donc trouver un équilibre subtil entre l'agilité apportée par les agents spécialisés et la nécessité de maintenir une architecture de systèmes d'information cohérente et gouvernable. Cela implique de développer de nouvelles approches d'architecture d'entreprise qui intègrent les agents comme des composants d'un système global orchestré, plutôt que comme des solutions ponctuelles indépendantes. La « transversalité » inhérente à la mise en place de tribus d'agents invite

à repenser la manière de concevoir, d'opérer et de maintenir l'ensemble des systèmes d'information et de leurs composantes.

7.6. L'infrastructure IA agentique du futur : vers une interopérabilité généralisée

L'un des enjeux majeurs pour l'avenir des agents experts IA réside dans le développement de protocoles d'échanges standardisés garantissant une interopérabilité optimale entre agents. Des initiatives émergentes telles que *NLWEB*[130] *(Natural Language Web), MCP* et *A2A (Agent-to-Agent)* constituent les fondements d'un écosystème où les agents pourront communiquer, collaborer et s'orchestrer de manière fluide, indépendamment de leur origine technologique ou de leur domaine d'application.

Plus ambitieux encore, le projet *NANDA (Network for Agent-based Distributed Applications)*, annoncé par le *MIT*, promet de révolutionner l'architecture des systèmes multi-agents en proposant une infrastructure de registre distribuée extra entreprise (voir note 126). Cette plateforme permettra aux agents d'une entreprise de découvrir automatiquement d'autres agents spécialisés disponibles sur le web à partir de catalogues, de négocier des collaborations et d'orchestrer des workflows complexes. *NANDA* incarne une vision où l'IA devient véritablement distribuée et autoorganisée, créant un écosystème d'agents experts IA capables de s'adapter en temps réel aux besoins émergents au-delà des frontières de l'entreprise.

Cette évolution vers une infrastructure d'agents experts IA interconnectés transformera fondamentalement la manière dont les organisations conçoivent leurs systèmes d'information et même leurs modèles d'affaires. Plutôt que de développer des solutions monolithiques, les entreprises pourront composer dynamiquement des services à partir d'un écosystème d'agents spécialisés, conférant une agilité et une adaptabilité sans précédent.

7.7. Recommandations stratégiques pour les décideurs

Face à la complexité des enjeux actuels, à la rapidité des avancées technologiques et à la concomitance de la découverte de nouveaux risques et impacts, les dirigeants d'entreprises, qu'ils soient à la tête de grandes corporations ou de PME, sont tenus

[130] https://news.microsoft.com/source/features/company-news/introducing-nlweb-bringing-conversational-interfaces-directly-to-the-web/

d'adopter une approche stratégique rigoureuse qui concilie innovation technologique et responsabilité organisationnelle.

Pour les grandes entreprises, il est impératif de prioriser le développement d'une gouvernance robuste permettant d'expérimenter et de déployer les agents experts IA tout en maîtrisant efficacement les risques associés. Cette gouvernance d'IA doit s'intégrer et collaborer avec d'autres gouvernances, notamment d'entreprises, afin d'assurer une cohérence et une efficacité optimales. Cela implique la création de centres d'excellence dédiés à l'IA, la mise en place de processus d'évaluation des risques spécifiques aux agents IA, et l'investissement dans le développement de compétences internes de supervision et de contrôle des systèmes autonomes. Il est également essentiel que cette gouvernance soit en apprentissage continu, tout en conservant son agilité, qu'elle accepte de prendre des risques calculés mais soit capable de s'inscrire dans une perspective de long terme.

Pour les PME, l'enjeu principal réside dans la capacité à bénéficier des avantages substantiels offerts par les agents experts IA sans compromettre leur agilité naturelle et leur aptitude à appréhender les risques et impacts potentiels. L'adoption de solutions standardisées et interopérables, le recours à des partenaires technologiques de confiance et la formation généralisée des équipes, y compris les équipes opérationnelles, constituent des leviers essentiels pour réussir cette transformation numérique.

Dans tous les cas, les organisations doivent résister à la tentation d'une adoption précipitée de l'IA motivée uniquement par la pression concurrentielle ou l'effet de mode. La mise en place d'une stratégie d'IA responsable, qui intègre dès le départ les dimensions humaines, éthiques et environnementales, représente un investissement à long terme qui déterminera la durabilité et la pérennité des avantages obtenus.

7.8. Vers une IA au service de l'humain

L'avenir des agents experts IA ne se détermine pas exclusivement par les avancées technologiques, mais également par notre aptitude collective à définir un modèle d'intégration qui préserve et valorise l'intelligence humaine. Les exemples les plus probants d'adoption d'agents experts IA sont ceux qui établissent une synergie et une relation de confiance entre les individus et entre les capacités des machines et celles des humains, plutôt que de viser à remplacer intégralement l'intervention humaine.

Cette perspective exige une réévaluation fondamentale de notre relation à l'automatisation. Au lieu de percevoir les agents experts IA comme des substituts aux

travailleurs humains et, par conséquent, comme une menace pour l'humanité, nous devons les concevoir comme des amplificateurs de l'intelligence collective, libérant les individus des tâches répétitives afin de leur permettre de se consacrer à des activités à plus forte valeur ajoutée : créativité, empathie, résolution de problèmes complexes, innovation et participation à des collectifs transversaux.

L'enjeu principal ne réside donc pas dans la détermination de l'impact des agents experts IA sur la transformation de nos organisations – cette transformation est déjà en cours – mais dans la définition des modalités de son orientation afin qu'elle serve nos objectifs humains et sociétaux. Cela requiert un effort collectif de réflexion, d'expérimentation et d'adaptation impliquant l'ensemble des acteurs de l'écosystème : entreprises, institutions publiques, chercheurs et citoyens.

En conclusion, les agents experts IA constituent une opportunité historique d'accroître nos capacités collectives et de résoudre des problématiques complexes qui dépassent les limites des capacités humaines individuelles. Toutefois, cette opportunité ne se réalisera que si nous parvenons à établir un cadre d'adoption responsable, plaçant l'humain au cœur de la transformation technologique. C'est à cette condition que nous pourrons tirer pleinement parti du potentiel révolutionnaire des agents experts IA tout en préservant les valeurs et les compétences qui constituent la richesse de nos organisations et de notre société.

HUB FRANCE IA

Glossaire

AGENTS EXPERTS IA
Janvier 2026

Glossaire

A2A (Agent-to-Agent)	Protocole de communication permettant l'échange direct d'informations et de commandes entre agents IA autonomes, facilitant la coordination et la collaboration dans les systèmes multi-agents.
Agent actionnable	Type d'agent IA capable d'agir concrètement dans un environnement logiciel (navigateur, *API*, *ERP*) en observant, interprétant et exécutant des actions de manière autonome.
Agent collaboratif	Agent IA conçu pour travailler en coordination avec d'autres agents ou humains, partageant des tâches, des objectifs et des ressources pour résoudre des problèmes complexes de manière collective.
Agent conversationnel	Agent IA spécialisé dans l'interaction en langage naturel avec les utilisateurs, capable de comprendre les requêtes, de maintenir un contexte conversationnel et de fournir des réponses appropriées.
Agent créatif	Agent IA spécialisé dans la génération de contenus originaux (texte, images, vidéos, musique) à des fins artistiques, *marketing* ou narratives, combinant créativité et pertinence contextuelle.
Agent délibératif	Agent IA utilisant une représentation interne de son environnement pour planifier, raisonner et prendre des décisions en fonction de ses objectifs, croyances et connaissances.
Agent d'orchestration	Agent IA responsable de coordonner et synchroniser l'exécution de plusieurs agents spécialisés dans un workflow structuré, optimisant l'efficacité opérationnelle globale.
Agent d'optimisation	Agent IA spécialisé dans la modélisation de scénarios, la simulation de contextes et l'optimisation de l'allocation des ressources pour la prise de décision stratégique.
Agent de gouvernance	Agent IA chargé de superviser, contrôler et auditer le comportement d'autres agents, assurant le respect des règles de conformité, sécurité et éthique dans un système.
Agent documentaire	Agent IA spécialisé dans l'extraction, la transformation et l'analyse de données structurées et non structurées pour produire des documents et des insights exploitables.
Agent explorateur	Agent IA mobile capable de se déplacer dans des environnements numériques distribués, de s'adapter à différents contextes d'exécution et de collecter des données localement.

Agent expert	Agent IA hautement spécialisé dans un domaine particulier, combinant des modèles de langage génératifs avec des chaînes de raisonnement explicites pour résoudre des tâches complexes.
Agent multimodal	Agent IA capable de traiter et générer des informations dans plusieurs formats (texte, image, audio, vidéo) et d'adapter ses protocoles de communication selon l'interlocuteur.
Agent pédagogique	Agent IA conçu pour guider l'apprentissage humain, s'adaptant au niveau, à la culture et à la progression des utilisateurs pour faciliter la montée en compétences.
Agent réflexif	Agent IA doté de capacités d'autoréflexion et de métacognition, capable d'évaluer sa propre performance et d'ajuster dynamiquement ses stratégies.
Agent utilitaire	Agent IA qui prend des décisions en fonction d'une fonction d'utilité ou de préférences explicites, évaluant et comparant plusieurs options selon des critères définis.
Agentivité (proactivité)	Capacité d'un agent IA à initier des actions et à se fixer des objectifs de manière autonome, sans intervention humaine directe, reflétant son niveau d'autonomie opérationnelle.
Apprentissage par renforcement	Technique d'apprentissage automatique où un agent apprend à prendre des décisions optimales en interagissant avec son environnement et en recevant des récompenses ou pénalités selon ses actions.
AutoGen	Framework développé par Microsoft permettant de créer des systèmes multi-agents où plusieurs agents *LLMs* spécialisés collaborent pour résoudre des problèmes complexes.
Chain-of-Thought (CoT)	Technique de prompt engineering qui permet aux modèles d'IA de raisonner à travers des étapes intermédiaires, améliorant leurs capacités de résolution de problèmes complexes.
ChatGPT	*Chatbot* développé par *OpenAI*, fondé sur un grand modèle de langage.
Coding Agent	Agent IA spécialisé dans les tâches de développement logiciel, capable de générer, déboguer, refactoriser et tester du code, ainsi que de produire de la documentation technique.

Computer Use	Capacité d'un agent IA à contrôler directement un ordinateur en simulant les actions d'un utilisateur humain (clics, saisie, navigation), permettant l'automatisation de tâches complexes.
CrewAI	*Framework* open source permettant de composer des équipes d'agents IA dotés de rôles explicites, d'outils partagés et d'interfaces no-code pour les utilisateurs métiers.
DSL (Domain Specific Language)	Langage de programmation spécialisé conçu pour un domaine d'application particulier, utilisé notamment dans *HybridAGI* pour la programmation d'agents basés sur les LLM.
Empoisonnement de données	Technique d'attaque consistant à manipuler les données d'entraînement ou la mémoire d'un agent IA pour fausser ses résultats ou influencer son comportement de manière malveillante.
Explicabilité	Capacité d'un agent IA à rendre compréhensibles ses décisions, actions ou raisonnements pour un humain, permettant la transparence et la vérification de ses processus décisionnels.
Fine-tuning	Le *fine-tuning* d'une IA générative pré-entraînée consiste à lui faire exécuter un entraînement supplémentaire sur des données labellisées spécifiques d'une tâche ou d'un domaine particulier afin d'améliorer sa performance.
Fuzzy Symbolic AI	Approche d'IA cognitive basée sur les mathématiques floues et la logique continue, permettant un raisonnement plus adapté aux problèmes complexes du monde réel que la logique binaire classique.
Gestion électronique de documents (GED)	Solution logicielle visant à organiser et gérer des informations sous forme de documents électroniques.
Hallucination	Information fausse, inexacte ou incohérente créée par une IA générative.
Health and Usage Monitoring Systems (HUMS)	Sont un terme générique désignant des activités qui utilisent des techniques de collecte et d'analyse de données afin de garantir la disponibilité, la fiabilité et la sécurité des véhicules.
HybridAGI	Approche utilisant un langage spécifique au domaine (*DSL*) pour programmer des agents basés sur les *LLM*, définissant des programmes sous forme de graphes d'actions avec des étapes de décision explicites.
IA générative	Sous-ensemble du *Deep Learning*, visant à produire du contenu, que ce soit du texte, une image, de l'audio ou une vidéo, à partir de données en

	entrée (on parle alors de prompt), elles-mêmes du texte, une image, de l'audio ou une vidéo. En anglais Generative AI ou GenAI.
IA neuro-symbolique	Approche fusionnant l'IA symbolique (raisonnement logique) et l'IA connexionniste (réseaux neuronaux), créant des systèmes excellant dans le raisonnement déductif et l'apprentissage inductif.
IA symbolique	Approche d'IA basée sur la manipulation de symboles et de règles logiques pour simuler le raisonnement humain, privilégiant l'explicabilité et la transparence.
In-Context Learning	Capacité d'un agent IA à apprendre de nouvelles compétences et tâches à l'aide de prompts, d'exemples et d'instructions fournis au moment de l'inférence, sans réentraînement.
Injection de *prompt*	Technique d'attaque consistant à insérer des commandes malveillantes dans les entrées d'un agent IA pour influencer son comportement et contourner ses restrictions de sécurité.
Intentionnalité	Capacité d'un agent IA à avoir des intentions propres et à utiliser toutes ses capacités pour les réaliser, représentant le niveau le plus élevé d'autonomie agentique.
Jailbreak	Technique permettant de contourner les restrictions de sécurité d'un agent IA pour accéder à des fonctionnalités non autorisées ou obtenir des réponses interdites.
LangGraph	*Framework open source* qui modélise les *workflows* d'agents comme des graphes persistants, offrant reprise sur incident, contrôle pas-à-pas et déploiement industrialisable.
Large Language Model (LLM)	Un type d'IA générative capable de générer et d'analyser du texte (par exemple : langage naturel, langage de programmation …)
Long-Term Memory (LTM)	Mémoire à long terme permettant aux agents IA de stocker et rappeler des informations sur une période prolongée, à travers différentes sessions, pour une personnalisation et une continuité accrues.
Machine Learning	Apprentissage automatique à partir d'un ensemble de données.
Magentic-One	Système multi-agents développé par Microsoft avec un orchestrateur central qui planifie, assigne et réajuste dynamiquement les rôles d'agents spécialisés pour des objectifs complexes.

Model Context Protocol (MCP)	Protocole permettant aux agents IA de faire des appels *API* automatisés, de connaître les outils à disposition et de remplir dynamiquement les paramètres nécessaires.
Métacognition	Capacité d'un agent IA à réfléchir sur ses propres processus cognitifs, à évaluer sa performance et à ajuster ses stratégies de manière autonome.
Mix of Experts	Architecture combinant plusieurs modèles d'IA spécialisés (génératifs, symboliques, neuro-symboliques) pour créer des agents plus fiables, explicables et efficaces.
Multimodalité	Capacité d'un agent IA à recevoir, interpréter et produire des informations dans divers formats (texte, image, audio, vidéo) et à adapter ses protocoles de communication.
NANDA	Infrastructure de registre annoncée par le *MIT* permettant les appels entre agents à partir de catalogues, facilitant la découverte et l'interopérabilité des services d'agents.
NLWEB	Protocole d'échange émergent permettant la communication et l'interopérabilité entre agents IA via des interfaces web en langage naturel.
Optical Character Recognition (OCR)	Technologie de reconnaissance optique de caractères permettant aux agents IA d'extraire du texte à partir d'images et de documents numérisés.
Orchestrateur	Composant central dans un système multi-agents responsable de décomposer les missions, router le travail, surveiller l'état global et relancer les agents défaillants.
Planification	Capacité d'un agent IA à organiser logiquement des séquences d'actions pour atteindre un objectif, allant de stratégies conditionnelles simples à la planification osmotique complexe.
Prompt	Le *prompt* est l'instruction ou la requête en langage naturel fournie à l'IA générative dans le but d'obtenir une réponse (un contenu).
Retrieval Augmented Generation (RAG)	Génération augmentée via la récupération d'informations d'une base de connaissances qui n'a pas été utilisée lors de l'entraînement de l'IA générative.

ReAct	*Framework* de *prompt engineering* permettant aux modèles d'IA de raisonner et d'agir en réponse à une requête utilisateur à travers un cycle itératif d'actions et d'observations.
Règlement européen sur l'intelligence artificielle (RIA)	Règlement européen sur l'IA entré en vigueur en 2024, établissant un cadre juridique pour le développement et l'utilisation éthique et sûre des systèmes d'IA dans l'Union européenne.
RGPD	Règlement Général sur la Protection des Données.
Retrieval Interleaved Generation (RIG)	Technique avancée de génération où l'agent récupère et intègre dynamiquement des informations externes pendant le processus de génération de contenu.
Reinforcement Learning from Human Feedback (RLHF)	Méthode d'apprentissage par renforcement utilisant les retours humains pour aligner les sorties des agents IA sur les valeurs et préférences humaines.
Supervised Fine-Tuning (SFT)	Technique d'entraînement supervisé permettant d'adapter un modèle pré-entraîné à des tâches spécifiques en utilisant des données labellisées.
Small Language Model (SLM)	Modèle de langage de taille réduite optimisé pour des tâches spécifiques et des environnements à ressources limitées, souvent utilisé dans les agents mobiles.
Système multi-agents (SMA)	Architecture où plusieurs agents IA spécialisés interagissent et collaborent pour résoudre des problèmes complexes, offrant robustesse, redondance et spécialisation.
Short-Term Memory (STM)	Mémoire à court terme permettant à un agent IA de conserver les entrées récentes pour faciliter la prise de décision immédiate et maintenir la cohérence conversationnelle.
Tree-of-Thoughts (ToT)	Généralisation de *Chain-of-Thought* permettant aux modèles d'IA d'explorer plusieurs chaînes de raisonnement simultanément, idéal pour les tâches exploratoires et stratégiques.
Workflow	Séquence organisée de tâches et de processus automatisés exécutés par un ou plusieurs agents IA pour atteindre un objectif spécifique de manière coordonnée.

HUB FRANCE IA

8. Remerciements

8. Remerciements

Le Hub France IA remercie l'ensemble des participants au groupe de travail IA Générative, et tout particulièrement les contributeurs de ce livrable.

Le pilote et le référent du GT :

- Alberto Tépox – Hub France IA.
- Georges Acar – Inquizyt.

Les contributeurs :

- Benjamin Bosch – Société Générale.
- Bertrand Lafforgue – Konverso.
- Charles Antoine Poirier – Future Path.
- Kajetan Wojtacki – Decision Brain.
- Henry Peyret – Wassati.
- Hugo David – Mindflow.
- Seif Benayed – Ask Hedi.
- Zineb Baroudi – Société Générale.

Les relecteurs :

- Alberto Tépox – Hub France IA.
- Caroline Chopinaud – Hub France IA.
- Christophe Baoudin – ITESOFT.
- Cyril Nicolotto – Hub France IA.
- Emmanuel Adam –Université Polytechnique Hauts-de-France.
- Florent Carlier – Université de Le Mans.
- Françoise Soulié-Fogelman – Hub France IA.
- Maxime Morge – Université de Lyon.
- Nicolas Sabouret – Université Paris Saclay.
- Pauline Bir – Hub France IA.
- Pierre Monget – Hub France IA.
- Zahia Guessoum – Université de Reims.

La touche finale :

- Mélanie Arnould – Hub France IA.

HUB FRANCE IA

AGENTS EXPERTS IA

Janvier 2026