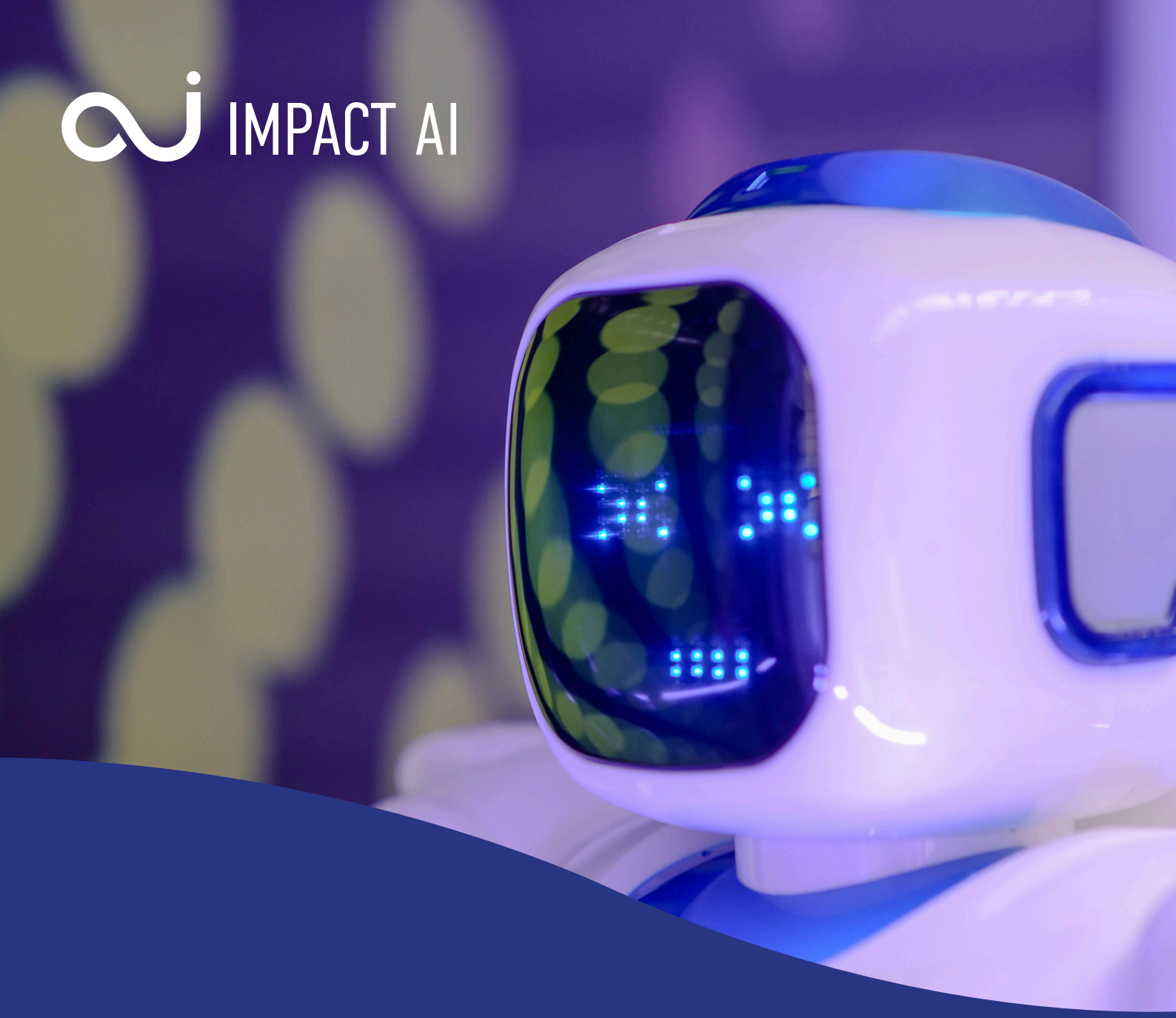
IMPACT AI



LES BRIEFS DE L'IA RESPONSABLE

10

IA agentique

Novembre 2025

IA AGENTIQUE : COMMENT RÉUSSIR L'INDUSTRIALISATION DES PROJETS PILOTES ?

L'IA agentique fascine mais engendre aussi beaucoup d'inquiétudes. Présentée dans la 10e édition des Briefs de l'IA Responsable d'Impact AI, notre communauté a coconstruit des recommandations concrètes pour le passage d'une intelligence artificielle simple assistante à une IA capable de décisions autonomes. Autonome, vraiment ?

Notre communauté d'experts et de managers de terrain reste partagée. Preuve en est, les décideurs en sont encore souvent au stade exploratoire, guidés par la quête de performance. Faut-il attendre pour en étendre les usages au cœur des métiers, au risque de manquer cette transformation majeure ?

Ce Brief livre les clés pour réussir un passage à l'échelle maîtrisé, éthique, sûr et sécurisé, en intégrant robustesse des systèmes, observabilité, contrôle renforcés et gouvernance responsable. S'appuyant sur l'expérience terrain de la communauté Impact AI, complétée par des études récentes, il propose une culture partagée de l'IA agentique pour une adoption maîtrisée, transparente et rentable. Ce nouvel eldorado des entreprises ne tiendra ses promesses que si ses erreurs sont anticipées et corrigées. Du PDG aux utilisateurs métiers. De la stratégie aux gestes du quotidien, le 10ième brief de l'IA responsable d'Impact AI nous révèle comment.

Autour des témoignages de nos experts et managers de terrain, ce premier volet retranscrit les impacts et enjeux de l'IA agentique au sein des organisations, qu'il s'agisse de pratiques avancées ou d'expérimentations en cours. Nous avons dégagé des thématiques communes susceptibles d'orienter vos futures décisions.

1/ L'IA agentique : l'émergence d'une nouvelle vague

Selon Forrester, « les systèmes d'**IA agentique** peuvent **planifier, décider et agir de façon autonome**, en orchestrant des flux de travail complexes avec une intervention humaine minimale ». Autrement dit, cette IA dite agentique ne se contente plus d'exécuter des tâches prédéfinies comme avec l'IA traditionnelle en recourant à l'apprentissage automatique, profond, et par transfert, elle répond à des questionnements, synthétise des informations, agit sur son environnement en vue de l'atteinte d'un objectif. Capable d'améliorer ses performances grâce à l'apprentissage ou à l'acquisition de nouvelles connaissances, elle agit comme une **véritable main-d'œuvre numérique**, venant compléter et renforcer les effectifs humains.

Cette nouvelle **collaboration entre humains et IA** ouvre des opportunités inédites pour les entreprises et transforme en profondeur leurs modes de fonctionnement. Elle accélère l'innovation et améliore la productivité à tous les niveaux de l'organisation, dans l'ensemble des fonctions.

Chaque domaine évolue à son propre rythme et selon ses besoins spécifiques : **automatisation** de bout en bout des processus complexes, **amélioration de l'agilité, optimisation de la prise de décision**, ou encore e**nrichissement des parcours clients**. Cette transformation dépasse les frontières de l'entreprise, en favorisant une collaboration étendue avec ses partenaires, fournisseurs et écosystèmes commerciaux.

Selon le dernier Work Trend Index, **82 % des dirigeants estiment que l'année en cours est charnière pour repenser les aspects clés de leur stratégie et de leurs opérations, et 81 % prévoient d'intégrer des agents d'IA dans leur stratégie d'ici 12 à 18 mois.** Le service client, le marketing et le développement produit figurent parmi les principaux domaines d'investissement identifiés.
Mais cette mutation profonde impose également de repenser nos approches de conception, de sécurisation et de gouvernance de l'IA, afin d'en garantir un usage responsable et durable :

- Le contrôle, la transparence, l'approbation et la responsabilité doivent être au cœur de tout déploiement pour que l'IA agentique devienne un atout fiable, éthique et digne de confiance.
- Le respect des obligations réglementaires et la mise en conformité avec la future loi européenne sur l'IA constituent désormais des priorités incontournables.

2/ L'IA agentique : une exigence stratégique

L'IA agentique n'est plus un simple concept futuriste : elle s'impose désormais comme un enjeu stratégique majeur pour les entreprises. Comme le soulignent Ethan Mollick et Lilach Mollick dans leur étude **Agentic AI : The Emerging Strategic Frontier**, cette nouvelle génération d'IA redéfinit les frontières de la stratégie d'entreprise et ouvre une ère où les agents intelligents deviennent des acteurs opérationnels à part entière.

Elle transforme ainsi la manière de concevoir, de traduire et d'opérer les processus métiers voire de les réinventer entièrement. Cette transformation est déjà tangible dans plusieurs domaines.

- **Dans le développement logiciel**, des agents comme Devin ou SWE-Agent commencent à automatiser tout ou partie du cycle de développement (revue de code, débogage, génération de tests). Les équipes peuvent ainsi analyser comment redistribuer les tâches et redéfinir les rôles entre développeurs, chefs de projet et testeurs.
- **Dans les métiers du support client**, des agents capables de résoudre de manière autonome 60 à 80 % des tickets simples permettent aux équipes humaines de se concentrer sur les cas complexes et les interactions à forte valeur ajoutée.
- **Et dans la gestion des ressources humaines**, des agents internes peuvent déjà simuler des scénarios de compétences ou suggérer des plans de montée en expertise, amorçant une **GPEC (Gestion Prévisionnelle des Emplois et des Compétences) avec l'IA augmentée**, sans attendre une refonte complète des processus.

Contrairement aux systèmes plus simples, qui réagissent à des stimuli immédiats, les IA agentiques incarnent une forme avancée d'agentivité numérique : de **véritables collaborateurs intelligents**, capables de raisonner, planifier, interagir avec des systèmes métiers ou d'autres agents, et même s'auto-évaluer selon les résultats obtenus. Ils se distinguent ainsi fondamentalement des programmes classiques à logique fixe.

Les progrès rapides de l'IA générative ouvrent des perspectives inédites : ils permettent de repenser les organisations en profondeur, en brisant les silos existants et en questionnant la conception même des systèmes. Selon la loi de Conway, les logiciels et processus métier reflètent la structure des équipes qui les conçoivent ; or, l'IA agentique offre désormais la possibilité de remodeler cette dynamique à la racine.

Ces avancées poussent à reconsidérer le « monde » des agents logiciels pour adresser des **problématiques plus complexes** : gestion d'objectifs, adaptation au contexte, autonomie ou collaboration inter-agents. Mais exploiter pleinement leur potentiel **exige des conditions de déploiement rigoureuses** : un développement fiable, maîtrisé avec un suivi continu pour prévenir les nouveaux risques et modes de défaillance. La mesure systématique de la performance des agents est également cruciale pour convaincre directions et équipes, et donner une ambition claire aux futurs déploiements.

Répondre à ces nouvelles exigences suppose des décideurs capables d'aligner l'impact des agents autonomes sur la vision globale de l'entreprise : stratégie, valeurs business et de marque, gestion des compétences et objectifs commerciaux. Cette vision doit être partagée avec l'ensemble des collaborateurs, **en levant les freins liés à l'anxiété, la défiance ou la méconnaissance de l'IA**. Pour garantir la rentabilité et l'alignement stratégique, la spécialisation des agents devient essentielle.

Le stade des agents universels touche à sa fin : **la valeur réside désormais dans la précision** de l'adaptation aux besoins métier. Cette spécialisation doit s'accompagner d'une gouvernance claire : chaque acteur — développeur, utilisateur, organisation — doit connaître ses responsabilités et conserver un contrôle sur les actions semi-automatisées.

Cependant, cette ambition se heurte à des réalités concrètes : lourdeurs administratives, complexité organisationnelle, coûts de coordination.

Les entreprises doivent intégrer l'IA dès la conception de leurs processus, en la considérant comme une véritable main-d'œuvre numérique appelée à collaborer avec les humains. « Les entreprises doivent modéliser un scénario de rentabilité intégrant non seulement le ROI financier, mais aussi les gains en agilité, résilience et réduction des risques. Le temps est un facteur clé : la performance d'un agent IA se construit par itérations et montée en maturité progressive », selon l'**Experte Data IA, Xdecisio, Delphine Chardon**.

Ainsi, « Chez Orange la stratégie agentique IA s'inscrit dans le prolongement de notre stratégie IA et IA générative. Un des piller important pour démarrer est de construire l'architecture technique de base avec des partenaires pour nous permettre d'être à la pointe technologique, développer nos usages. En même temps nous déployons une démarche responsable, frugale, sécurisée qui assure l'évolutivité et notre autonomie vis à vis des choix des fournisseurs de LLM. » comme témoigne **Olivier Simon, Orange VP Smart Networks & Data**.

3/ L'IA agentique : un défi managérial

Alors que les entreprises doivent désormais se mobiliser pour intégrer l'IA agentique dans un cadre conforme à la loi européenne sur l'IA, la mise en pratique révèle encore des défis qui dépassent le contenu strict du texte légal.

La culture d'entreprise doit évoluer vers une **maîtrise approfondie de l'IA literacy**, avec des formations obligatoires et une communication interne renforcée pour garantir que tous comprennent les enjeux éthiques et les risques liés à l'IA agentique. Autonomiser les collaborateurs est essentiel pour canaliser les usages de l'IA de manière sécurisée, limitant ainsi les **phénomènes de « Shadow AI »**.

L'organisation doit être **repensée de manière systémique** : cataloguer les agents autorisés, cartographier leurs instances dans des registres centralisés, et redéfinir les flux de travail et métiers pour intégrer les agents IA autonomes comme de véritables collaborateurs numériques, avec une responsabilité auditable pour chaque action, au même titre que les salariés.

« L'efficacité de l'IA dépend encore profondément de la présence humaine : il faut apprendre à devenir le "**Human in the loop**", celui qui guide, corrige et amplifie les capacités de l'agent. » (inspiré d'Ethan Mollick, Co-Intelligence: Living and Working with AI, 2024). Certains parlent même aujourd'hui d'expert in the loop pour accentuer le besoin de personnes formées, outillées pour assurer un contrôle pertinent.

Pour maximiser l'impact de ces équipes mixtes, les entreprises doivent réfléchir à un nouveau ratio agents-humains : Combien d'agents IA pour quels rôles et tâches ? Combien de collaborateurs humains restent indispensables pour les guider, superviser ou expertiser ?

Des études récentes montrent que pour des problématiques complexes, impliquant innovation, coordination ou raisonnement multi-étapes, les équipes mixtes surpassent toutes les autres configurations. Par exemple, un projet impliquant 791 professionnels chez P&G a montré que les équipes humaines + IA obtenaient un rendement supérieur à celui des équipes sans IA. (Harvard Digital Data Design Institute)

Cela suggère que le véritable ROI de l'IA agentique se trouve surtout dans les environnements complexes, et pas uniquement dans l'automatisation de tâches simples. Pour ces dernières, le ratio peut être élevé en agents IA seuls, mais pour des activités à forte valeur ajoutée, la collaboration humains + agents reste dominante.

Dans un monde où l'IA débloque de nouvelles compétences, appliquant le principe schumpétérien de destruction créatrice, les dirigeants devront faire confiance aux salariés pour évoluer au-delà des rôles traditionnels. L'émergence du **« patron des agents »** devient probable : un salarié qui construit, délègue et gère des agents IA pour amplifier son impact, travailler plus intelligemment, évoluer plus rapidement et prendre le contrôle de sa carrière.

La définition d'un **ratio optimal agents / humains** sera cruciale et spécifique à chaque tâche. Comme les RH gèrent les performances humaines et l'IT les systèmes, les entreprises devront développer de nouveaux modèles de gouvernance pour allouer et superviser la main-d'œuvre numérique. Certaines organisations pourraient fusionner RH et IT ou créer des rôles dédiés — par exemple un directeur de l'équilibre humain / IA, pour superviser la répartition optimale entre collaborateurs et agents.

Pour répondre aux exigences managériales :

- **Portefeuille IA centralisé** : établir une vue d'ensemble des projets d'agents IA autonomes par équipe, avec stratégie et feuille de route spécifiques.
- **Sensibilisation du middle management** : organiser ateliers et communication interne pour vulgariser la stratégie IA et fournir aux managers les outils pour la transmettre à leurs équipes. Adapter l'accompagnement selon la culture, le profil et le niveau technique des managers (formations segmentées non techniques).
- **Renforcement des managers** : leur donner les clés pour manager des équipes mixtes, conduire des team buildings humains/agents, clarifier l'évolution de carrière et les incitations des salariés.
- **Formation des salariés** : apprendre à collaborer avec les agents IA, les superviser, les améliorer, et développer des pratiques d'usage sécurisées et efficaces.

Observation terrain – Salesforce :
Le déploiement à grande échelle de l'IA agentique dans le service support client a franchi un cap d'une année. L'entreprise a récemment dressé un premier bilan, analysant à la fois les résultats obtenus et les pratiques éthiques encadrant cette technologie. Interrogé sur les performances de cette technologie, **Julie Ravillon, Directrice du développement durable, Salesforce France** déclare : « Plus d'1,8 million de requêtes clients ont été traitées par nos agents IA avec un taux de résolution des cas traités de 77%. Par ailleurs, nous avons réussi à réduire les coûts de support de 17%. … ce qui est crucial, c'est de bâtir cette confiance inhérente à l'ère de l'IA agentique en combinant l'expertise humaine à la puissance des agents. Pour cela, chaque décision doit être traçable, supervisée et parfaitement transparente, un cheminement que nous préconisons pour nos clients et mettons en œuvre. »

4/ L'IA agentique : une nécessaire gouvernance

L'autonomie croissante des agents IA oblige les organisations à repenser leur gouvernance et à redéfinir le rôle et le périmètre de la supervision humaine. Ces agents, capables de poursuivre des objectifs complexes, de prendre des décisions, de s'évaluer et de s'adapter en continu, soulèvent des questions cruciales en matière de responsabilité, de transparence et de conformité.

Comme le souligne Gunther Anders dans L'obsolescence de l'homme (1956) : « Les instruments eux-mêmes [...] ne sont pas de simples objets que l'on peut utiliser mais déterminent déjà, par leur structure et leur fonction, leur utilisation ainsi que le style de nos activités et de notre vie, bref, nous déterminent... »; cette observation rappelle que la technologie n'est jamais neutre sur le plan des valeurs et que les agents IA, par leur conception et leurs capacités, influencent profondément les pratiques et décisions organisationnelles.

Le succès du **déploiement de l'IA agentique repose donc sur une gouvernance intégrée**, portée au plus haut niveau. Le COMEX, les directions Data, IT, Juridique, RH et métiers doivent collaborer étroitement pour définir des politiques, normes et cadres clairs, permettant de piloter, auditer et garantir, autant que possible, la conformité légale et réglementaire.

Cela implique la mise en place de pratiques rigoureuses tout au long du cycle de vie des agents IA :

- **Analyses d'impact IA responsable**, revues de sécurité (analyse des vulnérabilités, modélisation des menaces, red teaming IA) et revues de confidentialité (impact sur les données, contrôles et mesures).
- **Traçabilité, observabilité et auditabilité** des actions et décisions semi- ou pleinement **automatisées**.
- **Plans opérationnels robustes**, incluant déploiement progressif, gestion des incidents et procédures de retour arrière.
- **L'éthique occupe une place centrale** : dès la conception (IA ethic by design), les grands et petits modèles de langage doivent être alignés sur les valeurs et les domaines d'application de l'entreprise. La mise en place de comités d'éthique, de « champions » IA et de relais organisationnels permet d'anticiper et de prévenir les dérives ou préjudices.
- **La question de la répartition des responsabilités est également cruciale** : qui est responsable en cas d'erreur ou de préjudice causé par un agent IA? Le développeur, l'utilisateur, ou l'entreprise? Des rôles clairs doivent être définis pour chaque étape : conception, déploiement, utilisation et exploitation opérationnelle. Sans cette clarté, la confiance des collaborateurs et des parties prenantes reste fragile, et l'acceptabilité de la technologie incertaine.

Pour répondre aux exigences de gouvernance, il est recommandé de :

- **Établir et opérationnaliser une gouvernance rigoureuse**, intégrée aux flux de travail et aux processus métiers.
- **Garantir une supervision humaine constante (Human in the Loop)**, avec traçabilité, observabilité, auditabilité et gestion proactive des risques.
- **Mettre en place des outils d'évaluation, de guardrail et d'observabilité** tout au long des projets
- **Assurer une documentation et des audits continus**, afin de répondre aux exigences légales et aux standards de qualité, tout en maintenant un suivi rigoureux des systèmes et modèles IA.

Le cadre réglementaire actuel souffre encore de manque de clarté opérationnelle et de méthodologies éprouvées, freinant la mise en œuvre concrète des obligations. La complexité technique et l'opacité de certains modèles compliquent la traçabilité et les audits.

Observation terrain – AG2R La Mondiale :
Avec sa plateforme Almia, AG2R lance plusieurs expérimentations autour de l'IA agentique, visant à déployer des agents intelligents pour assister les collaborateurs dans leurs tâches quotidiennes. Plusieurs de ces projets sont aujourd'hui opérationnels, accessibles via l'assistant Almia Bot ou intégrés directement aux processus métiers.
Autour de ce projet stratégique, Ludovic LETORT, Directeur IA chez AG2R confie : « Nous devons être conscients des défis organisationnels et humains à surmonter pour en tirer pleinement parti. Le déploiement de l'IA générative constituait une première phase relativement simple à déployer. L'agentique, avec la capacité à gérer des processus de bout en bout, ainsi qu'à interagir avec d'autres agents, internes ou externes, pose des questions encore plus complexes d'intégration et de constitution de certitude dans un univers non déterministe. Il reste essentiel d'appréhender le plus tôt possible les impacts RH et de gouvernance IA." »

5/ L’IA agentique : des exigences technologiques

Pas de modèle unique, mais une famille d’architectures évolutives
Force est de constater qu’il n'existe, à ce jour, aucune solution unique — ni pour les défis posés par ces technologies, ni, au fond, pour concevoir les architectures et orchestrations capables de relever les enjeux de l’IA agentique. Loin d’être un cadre figé, l’IA agentique représente une famille d’architectures en constante évolution, sans modèle universel.

À l’instar des services qui ont transformé le développement logiciel, les **architectures agentiques** s’appuient sur une **diversité de patterns** adaptés à des contextes, des environnements et des objectifs métiers spécifiques. Aucune approche standard ne s’impose : chaque configuration doit être pensée selon ses contraintes propres et les besoins de performance ou de gouvernance. Au cœur de cette évolution, **l’architecture multi-agents émerge comme modèle de référence**, où un agent superviseur coordonne des agents spécialisés pour accroître résilience, flexibilité et agilité, grâce à une **décomposition fine des fonctions et processus**.

Une pile agentique robuste n'est pas une technologie unique mais une composition de services qui, ensemble, constituent la base de systèmes multi-agents ouverts, sécurisés et de qualité professionnelle. Il s'agit d'un maillage d'agents hétérogènes qui peuvent encore être exploités de manière cohérente dans le cadre de processus distribués à plusieurs étapes. Au fur et à mesure de l'évolution des systèmes multi-agents, les entreprises auront de plus en plus besoin de flexibilité pour connecter des agents IA ayant des compétences et le cas échéant des technologies et des plateformes différentes afin de « faire le travail » pour et avec les utilisateurs finaux.

Plusieurs facteurs entrent en jeu : l'intégration, de divers outils et API avec des interfaces et des formats de données différents. **L'interopérabilité** reste un défi majeur, car les agents doivent fonctionner de manière transparente sur différentes plateformes entre protocoles ouverts et solutions propriétaires. Enfin, **l'évolutivité** est essentielle pour que les agents IA puissent gérer des volumes de données et des charges de travail croissantes en s’adaptant aux spécificités des écosystèmes concernés.

L'absence de **standards ouverts consolidés** freine encore la collaboration et l'adoption de l'IA agentique entre plateformes et organisations. À cela s'ajoute la **rareté de profils hybrides** — combinant data science, maîtrise des grands (et petits) modèles et compétences logicielles — indispensables pour traiter des données variées et déployer des systèmes à grande échelle.

Sur le terrain, cette complexité technologique se traduit par des **défis très concrets** : passer du prototype à l'industrialisation suppose de définir clairement les objectifs, tout en posant les bases de la **confiance, de la conformité et de l'amélioration continue**. Le **traitement en quasi-temps réel**, la **maintenance** et la **gestion des erreurs** ajoutent encore à la difficulté.

Dans ce contexte, la traçabilité des actions menées par les agents IA et une observabilité rigoureuse en quasi-temps réel deviennent des conditions essentielles pour garantir transparence, fiabilité et contrôle.

Il s'agit de surveiller en continu le comportement global des agents IA, de détecter les dérives éventuelles et de garantir leur alignement avec les objectifs fixés, tant sur les décisions (semi-)automatisées que sur leurs résultats concrets.
Cela implique :

- Des conventions sémantiques et une instrumentation orientée agent IA, permettant le suivi des conversation (entre agents et agents et entre agents et humains, la capture des performances et la détection d'anomalies en quasi-temps réel — par exemple avec OpenTelemetry ;
- La prise en charge des agents de longue durée, capables d'exécuter des tâches étendues et persistantes ;
- L'enregistrement systématique des décisions et actions associées à l'identité de chaque agent (ou collaborateur), selon des standards ouverts tels qu'OIDC (ID Connect) et OAuth2.
- Ces mécanismes assurent un contrôle d'accès fondé sur les rôles et attributs, garantissent que seuls les acteurs autorisés interviennent dans les processus, et établissent une responsabilité traçable pour chaque action. L'ensemble contribue à une posture de sécurité "Zero Trust" cohérente à l'échelle de toute la pile agentique.

Selon IDC, les entreprises orientent désormais une part croissante de leurs investissements vers la gouvernance et la responsabilité de l'IA : en 2024, 35 % des dépenses liées à l'IA concernent les outils de gouvernance, et 32 % les services professionnels associés à une IA responsable. Pourtant, le principal frein à l'adoption et à la mise à l'échelle de l'IA demeure le manque d'outils de gouvernance et de gestion des risques, cité par plus de 30 % des organisations interrogées.

Tous ces défis compliquent encore le déploiement rapide et la mise à l'échelle d'agents autonomes à la fois performants et sécurisés. Parmi les principaux obstacles :

- Interopérabilité entre outils et API hétérogènes ;
- Scalabilité et performance en environnement distribué ;
- Gouvernance et traçabilité, avec une observabilité en quasi-temps réel (ex. OpenTelemetry), identité vérifiée (OIDC, OAuth2) et approche Zero Trust ;
- Rareté des compétences mêlant data science, ingénierie logicielle et IA.
- Les entreprises doivent arbitrer entre innovation, contrôle et efficacité, en intégrant coûts, résilience, sécurité, conformité et impact environnemental.

Selon leur écosystème ou leur stratégie de plateforme cible, les entreprises devront arbitrer entre innovation, contrôle et efficacité, tout en intégrant les contraintes de coûts, de SLA, de latence, de résilience, de sécurité, de protection des données, de conformité et de sobriété environnementale.

Enfin, au-delà des seules décisions technologiques, les retours d'expérience soulignent que la valeur apportée par l'IA dépend autant du choix du déploiement que de l'architecture retenue.

Un principe clé : « Choisir une architecture multi-agents orchestrée pour mieux gérer la complexité, sans oublier les chantiers critiques de la gouvernance et de la supervision » selon l'**Experte Data , Xdecisio, Delphine Chardon.**

6/ Quelques recommandations clés pour vous accompagner dans la conception d'une IA Agentique

- **Adopter une stratégie d'adoption progressive et maîtrisée** L'identification des cas d'usage à faible risque et fort impact opérationnel constitue un très bon point de départ.

- **Cartographier et prioriser vos processus métier à automatiser, (ré)inventer ou réutiliser** dans différents flux de travail. La réalisation d'un état des lieux et d'un audit approfondi de vos processus s'impose avant tout déploiement.

- **Favoriser des architectures multi-agents IA (autonomes), spécialisées et modulaires, spécialisées aux besoins**. Il s'agit ici de pouvoir à la fois maximiser flexibilité, performance et maîtrise des menaces et des vulnérabilités de la pile technologique.

- **Procéder par étapes dans la mise en œuvre technique, en intégrant les innovations tout en investissant dans des fondations stables**. L'identité de agents IA (autonomes), leur observabilité et leur mémoire ne sont pas facultatives ; ce sont les piliers qui différencient les automatismes ad hoc des systèmes d'entreprise, en consolidant la fiabilité, la sécurité, la maitrise du patrimoine informationnel et en permettant une conformité réglementaire.

- **Sélectionner et adopter des standards d'interopérabilité robustes**. L'adoption de protocoles de communication et de standards ouverts dès le départ, même dans le cadre de preuves de concepts ou de projets pilotes, garantit la pérennité et l'évolutivité de votre architecture et permet d'éviter des travaux ultérieurs. Les protocoles MCP (Model Context Protocol), A2A (Agent to Agent), ou les interfaces Open API pour les outils, se distinguent actuellement mais ne sont pas les seuls. Ces fondamentaux universels constitueront la colonne vertébrale technique de votre écosystème agentique pour les différents échanges qui en résulteront.

- **Adopter un Framework d'agents IA adapté à vos besoins**. Le choix d'un Framework est susceptible de conditionner votre capacité à implémenter et vos agents IA et orchestrer efficacement vos flux de travail (workflow). Parmi les solutions open source éprouvées : CrewAI, LangChain & LangGraph, Microsoft Agent Framework (résultant de l'unification d'Autogen et de Semantic Kernel).

- **Identifier, mesurer et atténuer les risques des agents IA.** Les agents IA appellent des outils et exécutent des tâches, transforment le mode de fonctionnement des organisations. Ces capacités induisent de nouveaux types de risques, que la gouvernance logicielle traditionnelle n'a jamais été conçue pour gérer.
  Les logiciels traditionnels sont déterministes et soumis à un examen par les pairs. Les agents, en revanche, sont probabilistes et contextuels, orchestrés par des modèles de langage (SLM vs. LLLM) qui prennent des décisions dynamiques au moment de l'exécution. Cette flexibilité les rend puissants, mais aussi imprévisibles.
  Profiter des avantages des agents IA impose de pouvoir déployer de tels agents avec le même niveau de confiance, de sécurité et de responsabilité que les logiciels traditionnels. Les pratiques d'analyse d'impact (d'IA responsable vs. de confidentialité des données personnelles/sensibles (DPIA au sens du RGPD), de modélisation des menaces de systèmes d'IA, de « red teaming » IA, etc. permettent d'établir une cartographie précise des risques. Selon l'OWASP, les principaux risques sont la fuite de données sensibles, le détournement d'agent (jailbreak, attaques par injections d'invite (prompt), directes ou indirectes), le mauvais alignement des tâches, l'exécution de tâches prohibées avec en ligne de mire le détournement des objectifs poursuivis.
  La capacité à mesurer la prévalence de tels risques et la pertinence des garde-fous sont les conditions nécessaires pour ensuite les atténuer et adopter des pratiques opérationnelles adaptées. Des garde-fous sont non seulement nécessaires pour les entrées et les sorties des modèles, mais également notamment vis-à-vis de l'invocation outils, ou de leur résultat produit.
  Cela implique, une observabilité ainsi qu'une approche hybride (déterministe vs. non-déterministe).

- **Maintenir l'humain au centre des décisions critiques (Human-in-the-Loop)**
  L'autonomie des agents IA ne doit pas signifier l'absence de supervision humaine. Intégrez systématiquement des points de validation à travers une approche hybride, et en apportant le bon équilibre : « Une surveillance excessive va à l'encontre de l'objectif des agents. Une surveillance insuffisante peut entraîner des dommages ou une perte de confiance. Les agents IA les plus intelligents ne sont pas les plus autonomes, mais ceux qui savent quand demander de l'aide. »
  Philippe Beraud, Responsible AI Lead, Microsoft France.

Pour aller au-delà / Source - Bibliographie

L'IA agentique, l'émergence d'une nouvelle vague :

- Etude The state of AI: How organizations are rewiring to capture value de McKinsey.
- 2025: The year the Frontier Firm is born de Microsoft.
- Article Agentic AI Is The Next Competitive Frontier du Forrester.

L'IA agentique, une exigence stratégique :

- Étude Agentic AI : The Emerging Strategic Frontier de Harvard.

L'IA agentique : un défi managérial prioritaire pour passer de la cohabitation à la collaboration entre l'humain et la machine :

- Rapport Rise of agentic AI How trust is the key to human-AI collaboration de Cap Gemini.
- Rapport Annual Work Trend Index on WorkLab de Microsoft (résumé, rapport complet).
- Lettre d'information LinkedIn AI at work de Microsoft, conçue pour aider les chefs d'entreprise à décoder les dernières innovations en matière d'IA et l'avenir du travail.
- Papier The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise.

L'IA agentique : une nécessaire gouvernance :

- Livre-blanc From Risk to : The Business Case for Responsible AI basé sur l'enquête mondiale d'IDC sur l'IA responsable, sponsorisée par Microsoft.
- Publications de Partnership on AI sur la gouvernance des agents IA : Prioritizing Real-Time Failure Detection in AI Agents, Preparing for AI Agent Governance et AI Agents & Global Governance: Analyzing Foundational Legal, Policy, and Accountability Tools.
- Papier Practices for Governing Agentic AI Systems.
- Cadre OCDE 2024, qui fixe des standards internationaux pour une IA digne de confiance.
- Normes internationales du ISO/IEC JTC 1/SC 42 dédiées à la normalisation technique et éthique de l'IA, comme couvertes dans ISO policy : Harnessing international standards for responsible AI development and governance.
- Cadre NIST AI Risk Management Framework (AI RMF) américain, qui cible la gestion pragmatique des risques, et le profil AI RMF Generative AI Intelligence Profile.
- Règlement (UE) 2024/1689 sur l'intelligence artificielle (EU AI Act), pour une conformité réglementaire en Europe.
- Modèle européen AIGA, qui promeut une gouvernance systématique et intégrée.
- Cadre ISACA Digital Trust Ecosystem, axé sur la confiance numérique dans les environnements IA.

L'IA agentique : une nécessaire gouvernance :

- Livre-blanc From Risk to : The Business Case for Responsible AI basé sur l'enquête mondiale d'IDC sur l'IA responsable, sponsorisée par Microsoft.
- Publications de Partnership on AI sur la gouvernance des agents IA : Prioritizing Real-Time Failure Detection in AI Agents, Preparing for AI Agent Governance et AI Agents & Global Governance: Analyzing Foundational Legal, Policy, and Accountability Tools.
- Papier Practices for Governing Agentic AI Systems.
- Cadre OCDE 2024, qui fixe des standards internationaux pour une IA digne de confiance,
- Normes internationales du ISO/IEC JTC 1/SC 42 dédiées à la normalisation technique et éthique de l'IA, comme couvertes dans ISO policy : Harnessing international standards for responsible AI development and governance.
- Cadre NIST AI Risk Management Framework (AI RMF) américain, qui cible la gestion pragmatique des risques, et le profil AI RMF Generative AI Intelligence Profile.
- Règlement (UE) 2024/1689 sur l'intelligence artificielle (EU AI Act), pour une conformité réglementaire en Europe,
- Modèle européen AIGA, qui promeut une gouvernance systématique et intégrée,
- Cadre ISACA Digital Trust Ecosystem, axé sur la confiance numérique dans les environnements IA.

L'IA agentique, des exigences technologiques majeures :

- Article Building effective agents d'Anthropic quant à des modèles de conception (patterns) simples et composables plutôt que des cadres complexes comme approche architecturale avec des LLM.
- Série de billets de blog Agent Factory de Microsoft pour parcourir le chemin de la construction d'agents en entreprise et jeter les bases des premiers cas d'utilisation et modèles de conception (patterns), aux outils et flux de travail des développeurs nécessaires pour passer du prototype à la production, aux pratiques d'observabilité, standards ouverts pour l'interopérabilité et enfin aux principes de gouvernance et de sécurité nécessaires pour déployer des agents de manière responsable.
- Document Taxonomy of Failure Mode in Agentic AI Systems de Microsoft quant aux modes de défaillance de l'IA agentique.
- OWASP Gen AI Security Project (https://genai.owasp.org/initiatives/#agenticinitiative) afin d'aider à rendre la sécurité de l'IA agentique réelle, actionnable et efficace :
    - Agentic AI Threats Navigator
    - Agentic AI - Threats and Mitigations
    - Multi-Agentic system Threat Modeling Guide v1.0
    - Securing Agentic Applications Guide 1.0
    - State of Agentic AI Security and Governance 1.0

Remerciements à Bérengère Arnold (Impact AI), Philippe Beraud (Microsoft), Delphine Chardon (Xdecisio), Anne-Marie Jonquiere, Ludovic Letort (AG2R La Mondiale), Julie Ravillon (Salesforce), Roxana Rugina (Impact AI), Olivier Simon (Orange) et Emilie Sirvent-Hien (Orange) pour leur engagement et leur contribution essentielle à ce Brief de l'IA responsable et à tous les participants aux ateliers qui ont bien voulu partager leurs expériences et cas d'usage.

Retrouvez-nous sur

www.impact-ai.fr

pour rejoindre nos initiatives et nos ateliers !


IMPACT AI