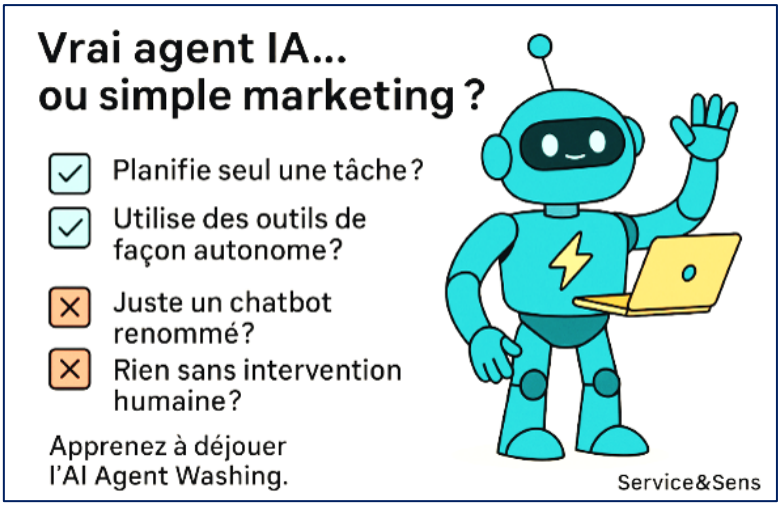
Agentic AI et AI Agent Washing

Comprendre, anticiper et sécuriser vos choix en 2025



1. Introduction : promesses d'autonomie et mirages technologiques

Dans l'effervescence actuelle de l'innovation numérique, l'intelligence artificielle semble tout bouleverser. L'arrivée des *agents autonomes* – ou *agentic AI* – alimente les espoirs les plus fous : assistants qui gèrent vos mails, agents commerciaux qui négocient seuls, techniciens digitaux qui résolvent vos tickets SAV… Bref, l'IA qui agit, raisonne et décide. Mais cette révolution annoncée est aujourd'hui parasitée par une dérive grandissante : celle de l'**AI Agent Washing**.

Ce terme, emprunté au *greenwashing* ou au *AI washing*, désigne une pratique marketing consistant à faire passer des technologies classiques – chatbots, scripts automatisés, RPA – pour des agents intelligents alors qu'ils ne disposent ni de l'autonomie, ni de la capacité d'adaptation ni du raisonnement requis. Résultat : les promesses ne sont pas tenues, les projets déçoivent, et la confiance des décideurs s'effrite.

Selon une étude de Gartner relayée par *Reuters* en juin 2025, plus de **40 % des projets « agentiques » seront abandonnés avant 2027**, notamment à cause de surventes technologiques et de manque de preuves tangibles d'intelligence réelle. Le marché, saturé de fausses promesses, crée une confusion dangereuse : entreprises désorientées, directions IT sceptiques, budgets gaspillés.

Pourtant, les véritables agents IA – bien conçus, testés, sécurisés – offrent un potentiel immense : coordination automatisée de tâches complexes, interactions naturelles, prise de décision stratégique ou opérationnelle, exécution autonome de tâches à valeur ajoutée. Encore faut-il savoir distinguer le vrai du faux.

Dans cet article, nous allons :

- définir ce qu'est réellement un agent intelligent autonome ;
- identifier les pratiques d'**AI Agent Washing** et leurs conséquences ;
- explorer des cas concrets ;
- proposer une grille de lecture stratégique pour aider les entreprises à choisir, auditer, déployer ou rejeter des projets « agentiques ».

2. Comprendre l'Agentic AI

2.1 Définition et fonctionnement des agents intelligents

Un agent IA (ou *agentic AI*) est une entité logicielle dotée de la capacité d'**agir de manière autonome** dans un environnement donné, en poursuivant un ou plusieurs objectifs, tout en adaptant son comportement en fonction du contexte.

L'agent ne se contente pas de répondre à une requête comme le ferait un moteur de recherche ou un assistant vocal. Il :

- **perçoit** des informations via des capteurs ou API (inputs) ;
- **raisonne** en s'appuyant sur des modèles cognitifs ou des chaînes de pensée (CoT) ;
- **planifie** une séquence d'actions pour atteindre son but ;
- **agit** de manière proactive, avec la capacité de boucler, réessayer, s'auto-corriger.

Ce cycle perception–raisonnement–action est la marque de l'agentic AI, qui se différencie des simples modèles conversationnels (LLMs) ou des systèmes RPA (robotic process automation).

2.2 Exemples récents

Parmi les initiatives notables en 2025 :

- **AutoGPT, BabyAGI, AgentGPT** : agents open-source expérimentaux qui planifient des tâches à partir d'une consigne utilisateur.
- **GPT-4o (OpenAI)** : architecture multimodale, capable de mener des actions dans un environnement connecté à des API, mais sous contrôle utilisateur.
- **Claude Sonnet (Anthropic)** : gestion d'objectifs et planification dynamique via constitution d'une mémoire étendue de contexte.
- **Devin (Cognition AI)** : agent développeur capable de coder, tester et corriger ses programmes pour accomplir un projet logiciel.

Ces projets explorent différents degrés d'autonomie. Certains sont "tool-using" (utilisent des outils externes), d'autres sont "multi-agents", capables de déléguer à d'autres agents.

Mais attention : ces solutions restent en grande partie expérimentales. Très peu d'entre elles atteignent la maturité opérationnelle nécessaire pour des déploiements industriels. Et c'est là que le "washing" commence.

3. Décryptage du phénomène d'AI Agent Washing

3.1 Définition

L'**AI Agent Washing** consiste à **présenter comme agents IA** des systèmes qui ne le sont pas réellement.

Exemples typiques :

- un **chatbot pré-entraîné**, sans mémoire ni capacité de planification, est présenté comme un "agent conversationnel autonome" ;
- un **script RPA** exécutant des actions sur un logiciel est qualifié d'"IA assistante" ;
- un système de règles (*if/then*) est vendu comme une solution d'agent intelligent proactif ;
- une interface "humaine" (voix, 3D, avatar) masque un moteur basique sans logique adaptative.

Cette pratique pose problème car elle :

- **trompe le client** sur la nature réelle du produit ;
- **induit des erreurs stratégiques** dans le choix et l'intégration de la solution ;
- **nuit à l'ensemble de l'écosystème IA**, en semant la confusion et en réduisant la crédibilité des véritables acteurs du domaine.

3.2 Un phénomène en pleine expansion

D'après *Bernard Marr* (Forbes, 11 juillet 2025), la multiplication des projets IA et la pression concurrentielle ont poussé de nombreux fournisseurs à **survendre** leurs produits, sous l'étiquette « agentique ». Résultat : on estime à seulement **130 le nombre de vendors réellement « agentics »** sur les milliers de prestataires revendiquant cette capacité.

Une étude menée par le cabinet DWF en avril 2025 montre que 27 % des solutions présentées comme des agents IA ne passent **aucun test de planification autonome**, et 19 % ne disposent **d'aucun moteur de raisonnement ou d'apprentissage intégré**.

Le terme devient un **label marketing vide**, parfois volontairement flou. Or, dans un contexte de tension réglementaire (AI Act européen, SEC américaine, FCA britannique), le recours à un vocabulaire exagéré ou trompeur peut avoir **des conséquences juridiques lourdes**, comme on le verra plus loin.

4. Pourquoi l'AI Agent Washing est un risque stratégique pour les entreprises

Si l'AI agent washing peut sembler à première vue n'être qu'un excès marketing bénin, ses implications peuvent s'avérer particulièrement lourdes pour une entreprise. Car au-delà de la déception technologique, ce sont bien **la performance, la conformité, la sécurité et la crédibilité** de l'entreprise qui sont en jeu.

4.1 Perte de confiance, réputation entachée

Un projet IA qui échoue, ce n'est pas qu'un surcoût : c'est un signal négatif envoyé à toutes les parties prenantes. Qu'il s'agisse d'un agent de recrutement IA incapable de discriminer correctement les candidatures, d'un pseudo-agent conversationnel incapable d'apporter de la valeur en SAV, ou d'un outil supposé "auto-exécutif" nécessitant en réalité des validations humaines constantes, l'échec laisse des traces.

Dans une étude de Deloitte de mai 2025, 47 % des responsables métiers interrogés affirment avoir **revu à la baisse leur confiance dans les projets IA autonomes**, après une expérience déceptive liée à une promesse non tenue. Le phénomène s'accélère dans les entreprises où les équipes métiers sont peu sensibilisées aux différences entre agents, assistants et automatisation simple.

Or, dans une dynamique d'innovation, la confiance est un actif immatériel essentiel : sa perte **ralentit l'adoption des vraies technologies** à forte valeur ajoutée, pousse à la frilosité budgétaire, et renforce les résistances internes.

4.2 Risques juridiques et réglementaires

La réglementation autour de l'intelligence artificielle se structure rapidement. L'Union européenne, à travers l'AI Act (promulgué en 2024), impose déjà des obligations de **transparence, traçabilité, explication et validation des systèmes autonomes**.

Outre les exigences de conformité, les autorités de régulation peuvent **sanctionner le recours à des termes trompeurs** ou à des pratiques fallacieuses :

- **La SEC** (États-Unis) a sanctionné en 2024 les entreprises **Delphia et Global Predictions** pour “AI Washing”, avec des amendes allant jusqu’à 400 000 $ pour avoir communiqué à leurs investisseurs des informations mensongères sur l’autonomie de leurs systèmes prédictifs.
- **L’ASA britannique** (Advertising Standards Authority) a quant à elle ouvert une enquête en 2025 sur plusieurs entreprises tech ayant vendu des assistants comme “agents d’exécution” sans capacités prouvées d’action autonome.
- La **FCA** et la **CMA** britanniques intègrent désormais la notion d’"AI misleading labelling" dans leurs lignes directrices en matière de conformité produit.

Les litiges sont donc bien réels, y compris dans des secteurs non-tech : santé, assurance, banque, RH. Toute promesse d’autonomie engage la responsabilité de l’entreprise, surtout si l’agent est impliqué dans une prise de décision affectant un client ou un salarié.

4.3 Risques opérationnels : erreurs, biais, dérapages

Un agent IA qui n’est en réalité qu’un automate mal étiqueté est un **danger opérationnel**. Pourquoi ? Parce que les équipes lui font confiance pour des tâches qu’il ne sait pas réellement exécuter.

Conséquences :

- **Décisions prises sans supervision**, sur la base de règles obsolètes ou mal calibrées ;
- **Propagation de biais**, lorsque l’agent est censé « apprendre » mais n’a pas de moteur adaptatif ;
- **Fuites de données**, si des API externes non vérifiées sont utilisées pour simuler une autonomie ;
- **Performance en chute libre**, dès que la tâche sort du cadre prévu par les scripts.

Ainsi, de nombreuses entreprises finissent par désactiver ou marginaliser des agents après quelques semaines, au profit d’un retour à l’humain ou à un système semi-automatisé.

5. Détecter les signaux faibles du "washing" : la check-list pour décideurs

Comment une entreprise peut-elle détecter une situation d’AI agent washing avant qu’il ne soit trop tard ? Plusieurs indicateurs permettent de faire le tri entre véritable agent autonome et produit rebrandé.

5.1 Les signaux d’alerte techniques et fonctionnels

Voici une check-list des points à vérifier systématiquement lorsqu’un fournisseur vous présente une solution "agentique" :

Élément	Agent réel	Agent washing probable
Planification	L’agent peut déterminer seul une séquence d’actions	Actions codées dans un workflow figé
Mémoire contextuelle	Mémoire à long terme, apprentissage	Aucune mémoire entre sessions
Outils intégrés	Capable d’invoquer dynamiquement des outils/API	Liste fixe de fonctions intégrées
Raisonnement	Utilise une chaîne de pensée (CoT), vérifie ses actions	Simple réponse au prompt initial
Autonomie	Peut prendre des décisions sans intervention humaine	Requiert validation à chaque étape
Explicabilité	Logique traçable, justificatifs de décisions	Boîte noire ou processus figé
Interface	Variable, mais secondaire	Avatar vocal mis en avant comme “preuve” d’intelligence

Check-list des points à vérifier face à un agent IA (proposée par Service&Sens)

Si plus de trois cases tombent dans la colonne droite : c'est un *red flag*, le danger est réel.

5.2 Les questions clés à poser aux fournisseurs

Pour éviter d'être pris au piège, posez des questions techniques précises, notamment :

- "Votre agent peut-il planifier des séquences d'actions ?"
- "Utilise-t-il un modèle de raisonnement explicite (ex. : chain-of-thought, tree-of-thought) ?"
- "Peut-il décider de modifier son plan en fonction de l'environnement ?"
- "Dispose-t-il d'une mémoire de travail persistante ?"
- "Quels outils (API, scripts, apps) l'agent peut-il appeler de façon autonome ?"

La règle : **moins le fournisseur comprend la question**, plus vous êtes face à un pur produit marketing.

6. Études de cas : entre promesses brillantes et revers douloureux

6.1 Amazon Just Walk Out : quand l'agent est un humain

Amazon a longtemps promu son service "Just Walk Out" dans ses magasins physiques comme une prouesse de computer vision et d'agents autonomes. Les clients étaient censés "entrer, prendre leurs produits, et ressortir" pendant qu'un système agentique les suivait, détectait leurs choix et gérait automatiquement la facturation.

Mais en 2024, un article du site *The Information* révèle que le système repose **à plus de 80 % sur des employés humains** situés à distance en Inde, qui regardent les vidéos et valident les achats en temps réel.

Ce scandale a révélé un cas massif de **"agent washing humain"** : une technologie vendue comme agent autonome était en réalité un processus semi-manuel déguisé.

Résultat :

- Fermeture de plusieurs magasins "Just Walk Out" ;
- Perte de crédibilité sur les offres IA d'Amazon ;
- Questions juridiques sur la protection des données personnelles et la transparence client.

6.2 Anthropic et Claude Sonnet : entre innovation et attentes mal calibrées

Anthropic a lancé avec fracas son agent Claude 3 Sonnet en 2025, présenté comme une avancée majeure dans l'autonomie des agents conversationnels. Pourtant, plusieurs tests indépendants ont montré que l'agent :

- restait fortement dépendant du prompt initial ;
- ne faisait pas de boucle de décision complexe ;
- nécessitait des relances fréquentes pour accomplir une tâche multi-étapes.

Bien que Claude 3 Sonnet soit techniquement avancé, il **ne correspondait pas à la définition stricte d'un agent autonome exécuteur de tâches**, comme l'a précisé un rapport du *MIT Technology Review*.

6.3 Sanctions par la SEC : les cas Delphia et Global Predictions

En mars 2024, la **Securities and Exchange Commission (SEC)** américaine a sanctionné deux entreprises fintech, **Delphia** et **Global Predictions**, pour avoir exagéré les capacités d'agents IA dans leurs offres d'investissement automatisé.

Les communiqués de presse parlaient d'"agents décisionnels", "intelligents, proactifs, adaptatifs", alors que les produits proposés étaient en réalité de simples algorithmes de recommandation paramétrés à l'avance.

Résultat :

- Sanctions financières (400 000 dollars cumulés) ;
- Retrait temporaire de certains produits ;
- Outils désormais sous supervision réglementaire renforcée.

7. Cadre stratégique pour repérer et éviter l'AI Agent Washing

Face à la montée de l'AI Agent Washing, les entreprises doivent structurer leur vigilance. L'approche ne peut être laissée au seul département IT ou innovation : elle doit impliquer la gouvernance, la conformité, les directions métiers, le juridique, et parfois même les RH.

7.1 Due diligence renforcée

Avant tout engagement avec un fournisseur, une démarche de **due diligence IA** s'impose. Celle-ci peut s'articuler autour des dimensions suivantes :

- **Technique** : vérification de l'architecture, présence d'un moteur de planification, raisonnement, capacité de multi-agent, logs d'exécution.
- **Fonctionnelle** : mesure de l'autonomie réelle, de la flexibilité face aux scénarios inattendus, de la capacité à gérer les exceptions.
- **Sécuritaire** : protection des données, cloisonnement des contextes, auditabilité.
- **Légale** : conformité avec l'AI Act, l'accessibilité, le RGPD, et la communication non mensongère.
- **Éthique** : mécanismes de supervision, transparence, explicabilité, niveau de contrôle humain.

Bonnes pratiques :

- Demander un **livre blanc technique**.
- Vérifier les **journaux de tâches** ou traces d'exécution.
- Réaliser une **analyse de risques différenciée selon les cas d'usage**.

7.2 Tests progressifs et preuve de valeur (POC)

Un agent IA ne se valide pas sur une démo. Il faut **tester en contexte réel**, selon une démarche progressive :

- **POC ciblé**, sur un cas métier concret et mesurable ;
- **MVP fonctionnel**, avec utilisateurs pilotes ;
- **Évaluation multicritère** : performance, autonomie, robustesse, adoption utilisateur ;
- **Scénarios inverses** : que fait l'agent en cas d'ambiguïté, d'erreur de contexte, ou de manque d'information ?

Les critères doivent être définis avant la phase de test, pas après. Sans cela, les biais de confirmation prennent le dessus.

7.3 Pilotage et traçabilité

Une fois déployé, l'agent doit rester **sous gouvernance active**. Cela implique :

- **supervision humaine** (*human-in-the-loop* ou *human-on-the-loop*) ;
- **indicateurs de performance clairs** : taux de réussite, réactivité, qualité perçue, nombre d'escalades humaines ;
- **tracabilité des décisions prises** ;
- **logs d'activité disponibles et exploitables** ;
- **droits d'audit** par un tiers (notamment en cas d'externalisation ou d'achat en marque blanche).

Cette gouvernance différencie les agents "fiables" des gadgets marketing. Elle rassure aussi les régulateurs et les comités d'audit.

8. Bonnes pratiques de mise en œuvre d'agents autonomes

Certaines entreprises ont déjà établi des pratiques inspirantes dans la mise en œuvre d'agents IA, en contournant les pièges de l'AI washing.

8.1 Instaurer un cadre de gouvernance interfonctionnel

La gouvernance ne peut être purement technique. Une **cellule de pilotage transverse**, rattachée au COMEX, incluant DSI, juridique, métiers, conformité et direction générale, est un facteur clé de succès.

Exemples de bonnes pratiques :

- **Comité de validation éthique** (comme chez AstraZeneca) ;
- **Grilles de niveaux d'autonomie** utilisées par Schneider Electric pour classer les agents ;
- **Charte de transparence IA**, rédigée avec les utilisateurs finaux.

8.2 Accompagner les utilisateurs, former les métiers

Un agent IA modifie les rôles. Si les utilisateurs n'en comprennent ni la logique, ni les limites, la confiance s'effondre.

Recommandations :

- former les collaborateurs aux **capacités et limites des agents** ;
- intégrer les utilisateurs dans les **phases de test et de feedback** ;
- mettre en place une **fonction de référent agent** dans chaque BU ;
- anticiper les impacts RH (requalification, montée en compétence, repositionnement).

*Exemple : la SNCF a lancé en 2024 une série de **formations "agentic literacy"** pour les managers, afin de leur permettre d'identifier les projets d'IA washing et de dialoguer avec les équipes tech.*

8.3 Choisir les architectures adaptées

Le choix de l'architecture technique conditionne le niveau d'autonomie :

Approche	Capacités	Limites
RAG (Retrieval-Augmented Generation)	Réponse documentée	Ne planifie pas
Agent LLM isolé	Dialogue, réflexion	Exécution limitée
Agent LLM + outils	Actions déclenchables, outils tiers	Coordination manuelle
Multi-agent orchestré	Répartition des tâches, supervision adaptative	Complexité, coût
Agent hybride humain–IA	Robustesse, éthique	Moins automatisé

Quelques exemples d'architectures et de niveaux d'autonomie (proposés par Service&Sens)

De plus en plus d'entreprises optent pour des **architectures multi-agents + supervision humaine**, où chaque composant reste simple, mais la coordination génère de l'intelligence collective.

9. Enjeux futurs et tendances à horizon 2028

9.1 Explosion des cas d'usage (réels ou fantasmés)

Selon IDC et McKinsey, d'ici 2028 :

- plus de **30 % des applications SaaS intégreront un ou plusieurs agents autonomes** ;
- **15 à 25 % des décisions de gestion opérationnelle** seront prises (ou préparées) par des agents ;
- les dépenses mondiales en agentic AI atteindront **plus de 180 milliards de dollars**.

Mais cette expansion rapide crée un terrain fertile pour le **washing**, tant que des standards clairs ne sont pas imposés.

9.2 Vers des labels ou certifications ?

L'absence de standard nuit à la lisibilité du marché. À ce titre, plusieurs initiatives émergent :

- **Consortium Agentic Trust** (lancé en 2025 par Microsoft, Anthropic, Hugging Face) ;
- **Label européen "AI transparency by design"** en cours de définition ;
- Travaux du **IEEE sur les niveaux d'autonomie et d'explicabilité**.

À terme, un **indice de maturité agentique** pourrait devenir obligatoire dans les appels d'offres publics et les audits IA réglementaires.

9.3 L'essor de l'agentic ops

La maturité ne passera pas uniquement par la technologie, mais par les **organisations capables d'absorber et d'encadrer des agents** dans leurs processus internes.

Le modèle "Agentic Ops" pourrait devenir un standard :

- process documentés pour chaque interaction agent–humain ;
- tableaux de bord de pilotage IA ;
- indicateurs de supervision active ;
- mise à jour régulière du socle d'outils et de rôles ;
- agents coordonnés, mais sous contraintes métier.

10. Conclusion : vigilance, stratégie, transparence

Le développement de l'IA agentique est sans conteste l'un des tournants majeurs des stratégies numériques des entreprises d'ici 2030. Mais cette promesse est aujourd'hui parasitée par un phénomène croissant : celui de l'**AI Agent Washing**.

Sous couvert d'innovation, certaines offres n'apportent ni autonomie, ni valeur, ni transparence. Pire, elles peuvent faire perdre du temps, de l'argent et de la crédibilité aux entreprises.

Pour éviter cela, il faut :

1. **Savoir ce qu'est vraiment un agent IA** : percevoir, raisonner, agir.
2. **Identifier les signaux faibles du washing** : manque de mémoire, de planification, d'autonomie réelle.
3. **Adopter un cadre stratégique rigoureux** : due diligence, POC, supervision active.
4. **Former les équipes et sensibiliser les décideurs**.
5. **Contribuer à la construction d'un écosystème agentique crédible, éthique et sécurisé.**

L'entreprise qui saura **distinguer les vrais agents des avatars de façade** prendra une longueur d'avance. Non seulement dans la maîtrise technologique, mais aussi dans la confiance qu'elle inspirera à ses collaborateurs, ses clients et ses partenaires.

Lien vers l'article en ligne : https://www.service-sens.com/blog/agentic-ai-et-agent-washing-le-double-defi-strategique-pour-les-entreprises-en-2025

A propos de SERVICE&SENS :

"L'économie de demain ne vendra pas plus, elle servira mieux."

Naviguant à travers divers secteurs de l'industrie, nous cherchons à infuser du sens dans chaque accompagnement, formation ou conférence que nous offrons, en créant des solutions personnalisées qui répondent à vos besoins uniques.

En travaillant avec nous, vous bénéficiez d'une expertise de pointe, d'outils méthodologiques éprouvés et d'un dévouement inébranlable à l'excellence. Nos accompagnements ambitionnent de transformer les organisations opérationnelles et commerciales, autour du 'Product as a Service' et de l'Economie de la Fonctionnalité et de la Coopération, redéfinissant ainsi les standards du marché dans le cadre d'une transformation globale guidée par les impératifs de la transition écologique.

En résumé, Service&Sens conjugue sens et excellence servicielle, pour apporter un accompagnement ambitieux et éclairé en stratégie opérationnelle.

Nos domaines d'expertise : Stratégie, modèles économiques serviciels, SAV & service client, comportements relationnels, management agile, performance commerciale, industrie 4.0...

contact@service-sens.com | ***www.service-sens.com*** | ***https://www.linkedin.com/company/service-sens/***