# Qui décide quoi : la frontière de délégation

## Étude 05 — Réagencer le travail à l'âge des agents

Mathieu Guglielmino

2026-06-24

## Le fil



## Qui décide quoi : la frontière de délégation

𝓒 Journal des mises à jour

- **2026-06-24** — Création du dossier et du journal (premier passage) : échelle de délégation, incidents OWID, fossé de gouvernance McKinsey, AI Act art. 14, cadre des 4 fonctions.
- **2026-07-12** — Approfondissement « honnêteté » : nouvelle section 3 « Compter les ratés dépend de la définition » (fig05, triangulation des périmètres d'incidents — 233 vs 281 pour 2024, OECD AIM comme 3ᵉ compteur) ; renvoi depuis l'encart de la section 2 ; sections 3→7 renumérotées.

### La thèse

Personne ne **décide** la frontière entre ce qu'un agent tranche seul et ce que l'humain garde la main : elle se **sédimente par défaut**, au gré de la confiance accumulée et des incidents encaissés — et elle **glisse vers le haut** de l'échelle, vers l'autonomie. Le cadre conceptuel existe pourtant depuis 2000 — l'échelle d'automatisation de Parasuraman, Sheridan et Wickens[1] décompose la délégation en dix niveaux et quatre fonctions. Le droit l'impose pour les usages à haut risque : l'AI Act, article 14, exige une supervision humaine effective, et nomme explicitement le biais d'automatisation[2] — cette tendance à sur-s'appuyer sur la

[1] Parasuraman, Sheridan & Wickens, « A Model for Types and Levels of Human Interaction with Automation », IEEE Transactions on Systems, Man, and Cybernetics — Part A, vol. 30, n° 3, 2000. D'après Sheridan & Verplanck (1978). https://doi.org/10.1109/3468.844354

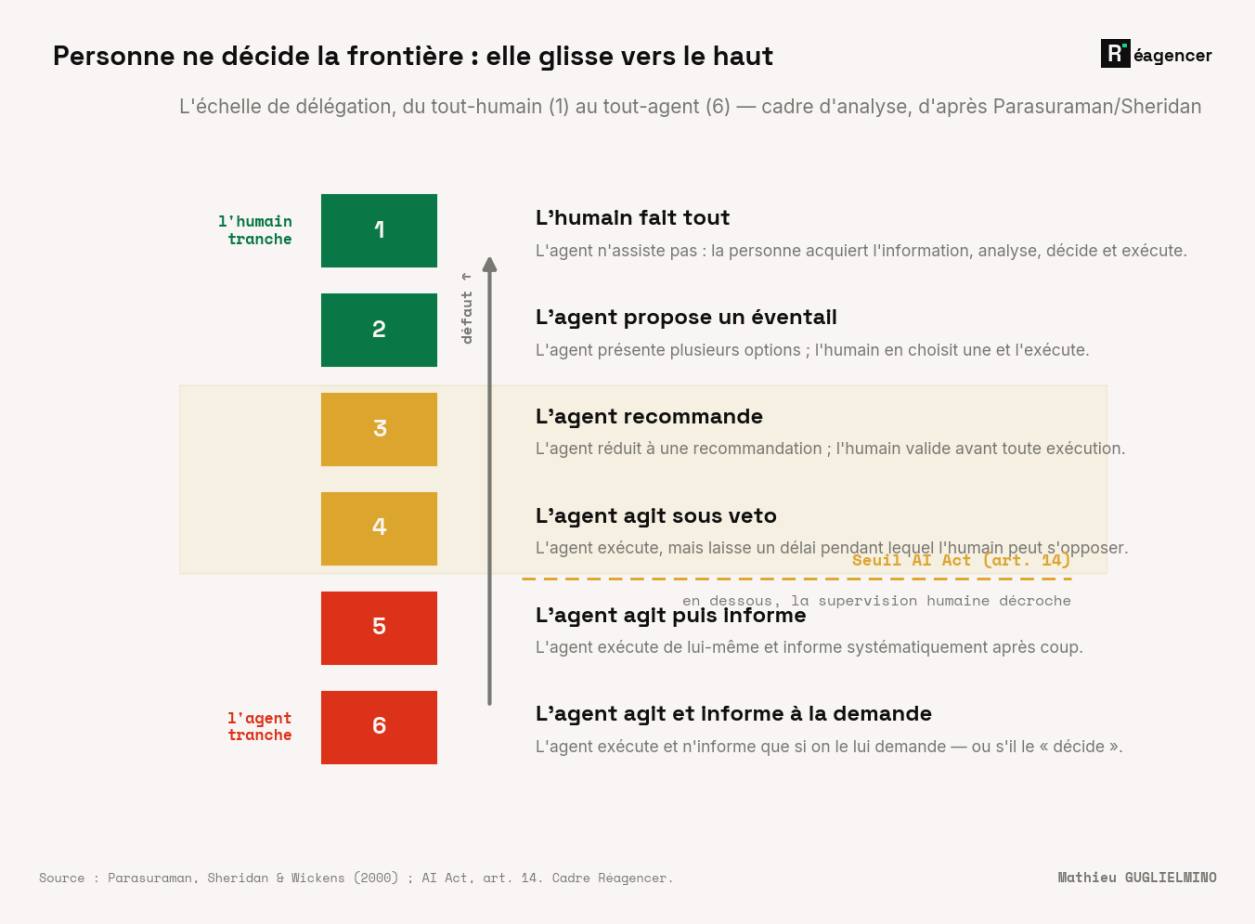
[2] Règlement (UE) 2024/1689 (AI Act), article 14 « Contrôle humain ». https://artificialintelligenceact.eu/article/14/

sortie de la machine jusqu'à ne plus la questionner. Sur le terrain, l'expérimentation court pourtant devant la gouvernance : 62 % des organisations expérimentent des agents, 23 % les passent à l'échelle, et la validation « human-in-the-loop » n'est définie que chez 65 % des hautes performances contre 23 % des autres[3]. Les incidents et controverses IA recensés dans le monde ont été multipliés par ≈ 4 en cinq ans[4]. **Corollaire Réagencer : on ne délègue pas « un agent », on délègue quatre fonctions — acquérir · analyser · décider · exécuter — et la décision est celle qu'il faut garder basse.**

---

## 1. L'échelle existe — personne ne la regarde

L'idée que la délégation à un agent serait binaire — « on lui confie » ou « on ne lui confie pas » — est un raccourci qui explique beaucoup d'accidents. La réalité est une échelle, et on peut se situer dessus avec précision.




Parasuraman, Sheridan et Wickens ont formalisé dès 2000 une taxonomie à dix niveaux, du « tout humain » au « tout machine »[5], articulée autour de quatre fonctions cognitives : acquérir l'information, l'analyser, décider, exécuter. Le cadre ci-dessus en propose un condensé en six barreaux, réutilisable comme grille de lecture pour n'importe quel agent déployé.

---

[3] McKinsey, « The State of AI 2025 » (nov. 2025, n=1 993, 105 pays, terrain 25 juin–29 juil. 2025). https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

[4] Our World in Data, « Global annual number of reported AI incidents and controversies » (d'après AI Incident Database / AI Index Report 2026). https://ourworldindata.org/grapher/annual-reported-ai-incidents-controversies

[5] Parasuraman, Sheridan & Wickens, « A Model for Types and Levels of Human Interaction with Automation », IEEE Transactions on Systems, Man, and Cybernetics — Part A, vol. 30, n° 3, 2000. D'après Sheridan & Verplanck (1978). https://doi.org/10.1109/3468.844354

Trois zones se distinguent. **La zone humaine** (barreaux 1 à 2) : l'agent assiste sans décider — il présente, suggère, filtre — mais l'humain tranche toujours. **La zone frontière** (barreaux 3 à 4) : l'agent propose et l'humain dispose, mais la supervision commence à devenir un acte formel — valider, ou ne pas valider ce que le système recommande. C'est là que le biais d'automatisation s'installe sans qu'on s'en rende compte : on valide par habitude, pas par examen. **La zone agent** (barreaux 5 à 6) : l'agent agit, et l'humain n'est prévenu qu'après — ou pas du tout.

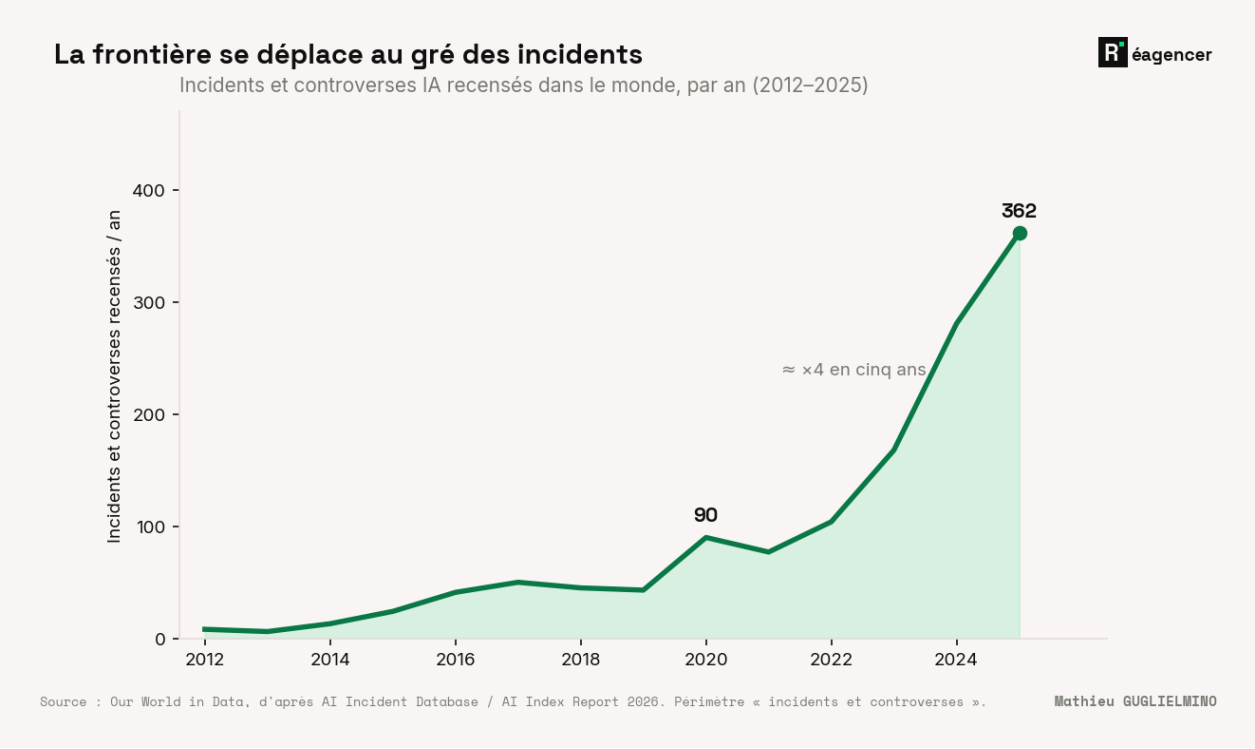
Le **seuil AI Act** se situe entre le barreau 4 et le barreau 5. En dessous, la supervision humaine effective que l'article 14 exige commence à décrocher. Ce seuil n'est pas théorique : il délimite l'espace légal pour les systèmes dits « à haut risque » — recrutement, accès au crédit, gestion de personnel, décisions médicales. Le franchir sans architecture de contrôle délibérée, c'est s'exposer à une non-conformité que la plupart des organisations n'ont pas encore anticipée.

Ce schéma est un **cadre d'analyse**, pas une mesure. Il n'existe aucune donnée représentative française sur le barreau réellement occupé par les agents en production dans les organisations.

---

## 2. La frontière se déplace au gré des incidents

La sédimentation ne se voit pas — jusqu'à ce qu'elle se paie. Les incidents sont le signal que la frontière avait glissé trop haut.




Les incidents et controverses IA recensés dans le monde sont passés de **90 en 2020** à **362 en 2025** — ≈ ×4 en cinq ans[6]. La courbe est exponentielle sur la période 2020–2025, avec **281 incidents recensés dès 2024**.

Ce chiffre ne dit pas que l'IA est quatre fois plus dangereuse qu'en 2020. Il dit que la surface exposée est quatre fois plus grande — plus d'agents déployés, dans plus de contextes, avec plus d'autonomie — et que la capacité de signalement a progressé en parallèle. La tendance compte ; le niveau absolu demande prudence.

[6] Our World in Data, « Global annual number of reported AI incidents and controversies » (d'après AI Incident Database / AI Index Report 2026). https://ourworldindata.org/grapher/annual-reported-ai-incidents-controversies

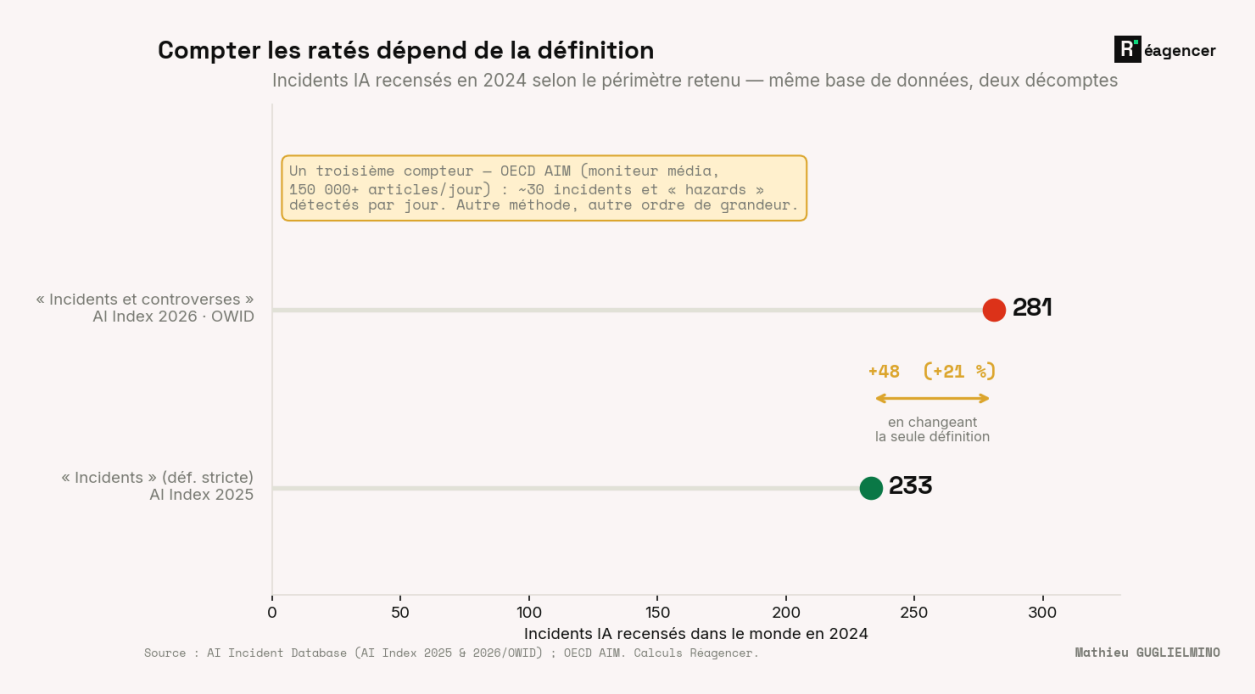
- **Périmètre « incidents et controverses »** — les données Our World in Data[7] agrègent incidents avérés et controverses (cas contestés, alertes préventives, débats publics autour d'un usage). Ce périmètre est plus large que celui des seuls incidents documentés au sens strict — un écart de périmètre qu'explore la section suivante. La tendance ×4 reste robuste ; le niveau absolu (362 en 2025) doit être lu comme un indicateur de vigilance collective, pas comme un décompte de défaillances techniques. **Lire la tendance, pas le niveau absolu.**

Ce que la courbe illustre, c'est le coût de la sédimentation invisible : une frontière qui glisse vers le haut par défaut rencontre tôt ou tard un cas où la supervision manquante n'est plus une abstraction.

---

## 3. Compter les ratés dépend de la définition

- Il n'existe pas UN nombre d'incidents IA. Le décompte dépend entièrement de la définition retenue — au point qu'on ne sait même pas compter les ratés d'une frontière qu'on ne sait déjà pas tracer. C'est un prolongement direct de la thèse : la frontière est invisible jusque dans ses ratés.




- Pour la même année 2024 et la même base curée — l'AI Incident Database —, le décompte passe de **233** (« incidents » au sens strict, chiffre rapporté par le Stanford AI Index 2025[8] — un record, **+56,4 %** sur 2023) à **281** (« incidents et controverses », périmètre élargi, repris par Our World in Data[9]). Soit **+48 incidents, +21 %**, rien qu'en élargissant la définition — pas en changeant la réalité du terrain.

---

[7] Our World in Data, « Global annual number of reported AI incidents and controversies » (d'après AI Incident Database / AI Index Report 2026). https://ourworldindata.org/grapher/annual-reported-ai-incidents-controversies

[8] Stanford HAI, « The 2025 AI Index Report — Responsible AI ». https://hai.stanford.edu/ai-index/2025-ai-index-report/responsible-ai

[9] Our World in Data, « Global annual number of reported AI incidents and controversies » (d'après AI Incident Database / AI Index Report 2026). https://ourworldindata.org/grapher/annual-reported-ai-incidents-controversies

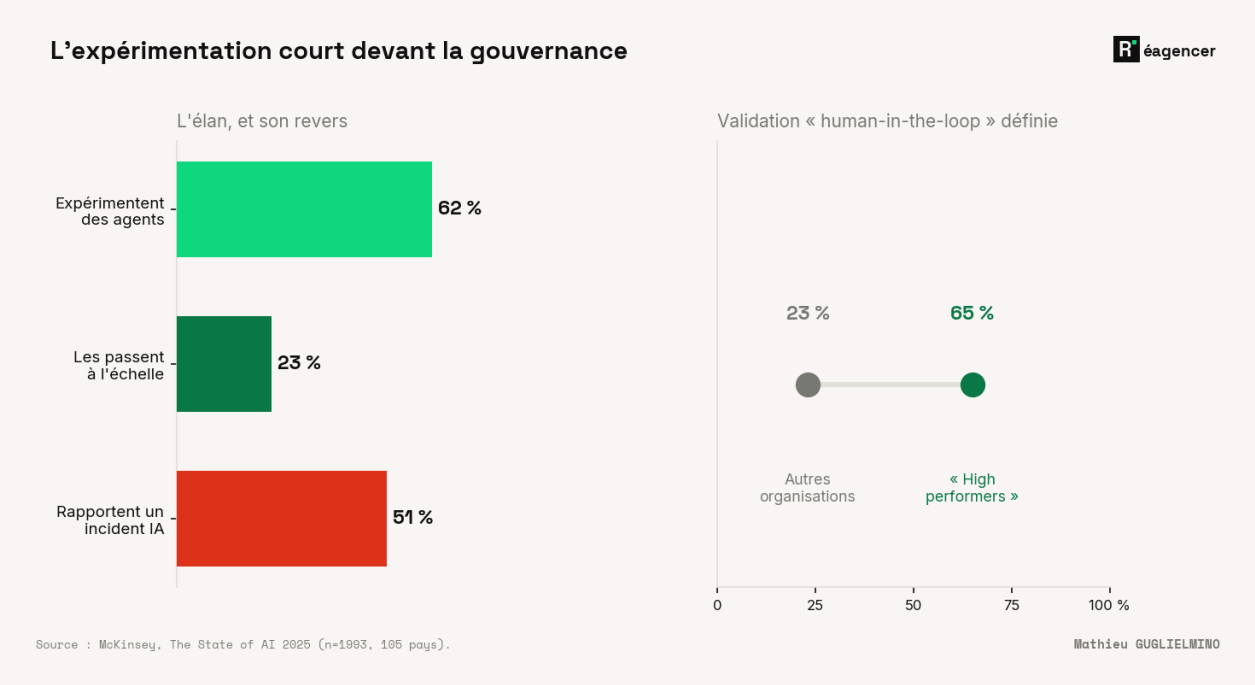
[10] OECD.AI, « AI Incidents and Hazards Monitor (AIM) — Overview and methodology ». https://oecd.ai/en/incidents-methodology

- Un troisième compteur procède autrement. L'**OECD AI Incidents and Hazards Monitor (AIM)**[10] est un moniteur média : il scanne plus de **150 000 articles par jour**, classés par des modèles de langage, et détecte en moyenne ≈ **30 incidents et « hazards » par jour**. Méthode et ordre de grandeur radicalement différents des deux bases curées ci-dessus — on ne compare pas des choux et des carottes. Dans le vocabulaire de l'OECD AIM, un **« hazard »** est un événement qui pourrait plausiblement mener à un incident : autrement dit, une frontière de délégation qui a déjà glissé, mais qui n'a pas encore été payée. Le hazard, c'est la sédimentation avant l'accident — exactement l'objet de cette étude, et personne ne le compte côté français.
- Le bon réflexe n'est donc pas de chasser le « vrai » chiffre d'incidents. C'est de lire une tendance — elle, robuste : la surface exposée croît — et de se méfier de tout niveau absolu brandi sans son périmètre. Un chiffre d'incidents sans sa définition ne veut rien dire.
- **Trois compteurs, pas un.** Les trois chiffres ci-dessus ne s'additionnent pas et ne se comparent pas terme à terme : les deux premiers (233 et 281) sont deux périmètres d'une même base curée, le troisième (OECD AIM) est un moniteur média qui fonctionne sur un principe différent. Le chiffre OECD AIM (≈ 30/jour) est un **rythme**, pas un décompte annuel de même nature que les deux autres — le convertir en total annuel pour le comparer serait une erreur de méthode, pas un raccourci acceptable. La seule lecture honnête est la **tendance**, pas le niveau.

---

## 4. L'expérimentation court devant la gouvernance

Le fossé entre déploiement et contrôle est désormais documenté à grande échelle. Il n'est plus théorique.




McKinsey a interrogé 1 993 dirigeants et responsables d'organisations dans 105 pays fin 2025[11]. Trois chiffres résument le paradoxe : **62 %** des organisations expérimentent des agents, **23 %** les passent à l'échelle — mais **51 %** rapportent au moins un incident ou une conséquence négative liés à l'IA.

[11] McKinsey, « The State of AI 2025 » (nov. 2025, n=1 993, 105 pays, terrain 25 juin–29 juil. 2025). https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Le chiffre le plus révélateur est celui du fossé « human-in-the-loop » : la validation par un humain d'une sortie d'agent avant qu'elle produise un effet est définie chez **65 %** des organisations dites « high performers » — et chez seulement **23 %** des autres. Un écart de 42 points entre ceux qui gèrent la frontière et ceux qui la subissent.

Ce fossé n'est pas un problème de compétence technique. C'est un problème de design organisationnel : où place-t-on le point de contrôle, qui l'exerce, sur quelle base décide-t-on de valider ou d'écarter ? Ces questions n'ont pas de réponse dans un modèle de langage. Elles s'organisent.

**Enquête mondiale, déclarative.** Les chiffres McKinsey portent sur 105 pays, avec une surreprésentation des grandes entreprises internationales. « High performers » = organisations qui s'autodéclarent comme telles sur un critère de performance IA (impact EBIT). Il n'existe pas d'équivalent français publié. Les valeurs sont des déclarations, non des mesures opérationnelles. Lire comme des ordres de grandeur directionnels.

---

## 5. Ce que le droit exige déjà

L'AI Act ne demande pas l'impossible. Il décrit exactement ce que le bon sens de gouvernance devrait prescrire — et il le rend obligatoire pour les systèmes à haut risque.

L'article 14 du règlement (UE) 2024/1689[12] organise la supervision humaine en cinq points :

- **§1** — la supervision humaine doit être effective : pas un pro-forma, pas une validation automatique.
- **§4(a)** — les opérateurs doivent comprendre les capacités et les limites du système, et savoir détecter les anomalies.
- **§4(b)** — ils doivent éviter le **biais d'automatisation** — la tendance à sur-s'appuyer sur les résultats du système, notamment quand il s'agit de décisions aux conséquences importantes.
- **§4(d)** — ils doivent pouvoir **écarter ou annuler** la sortie du système à n'importe quel moment.
- **§4(e)** — un **bouton d'arrêt** doit permettre de stopper le système d'urgence.
- **§5** — pour l'identification biométrique : **double validation humaine** obligatoire.

Le biais d'automatisation nommé en §4(b) est précisément le mécanisme par lequel la frontière glisse. Quand on valide par habitude plutôt que par examen, on n'est plus en barreau 4 — on est en barreau 5 sans l'avoir décidé. Le droit l'a anticipé. Il reste à le traduire en design de processus.

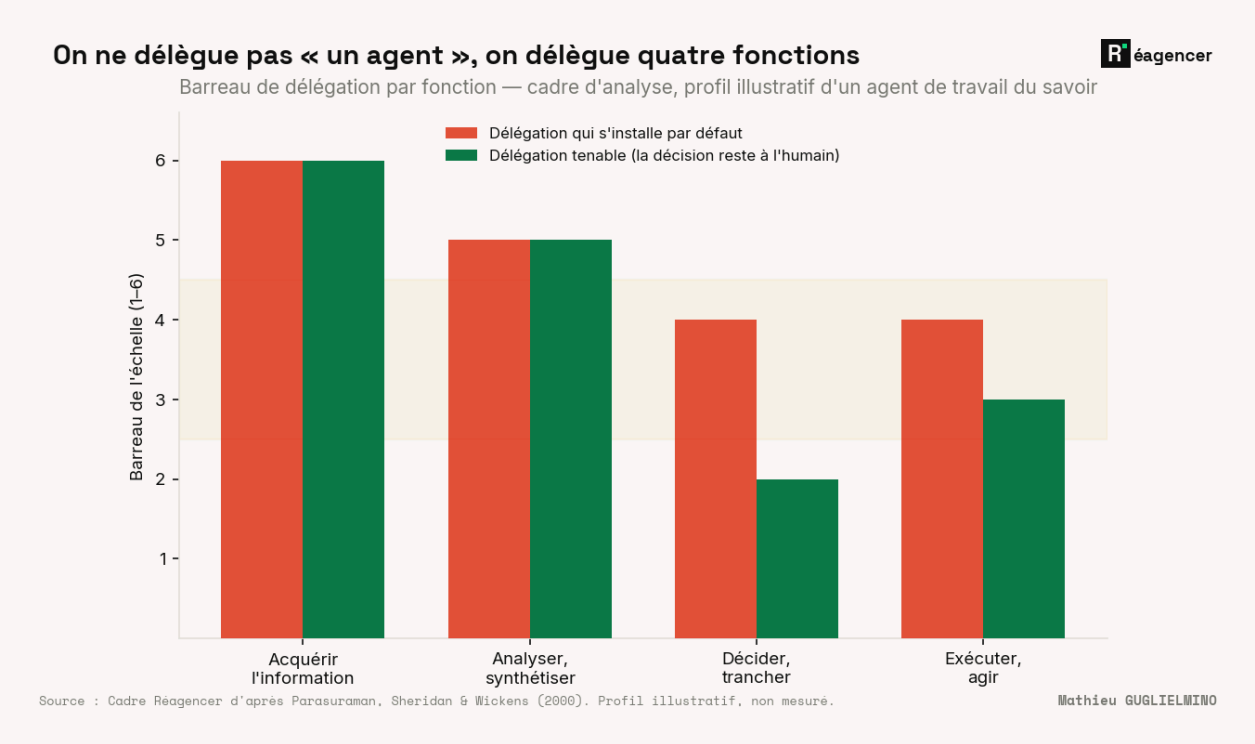
Ce que l'AI Act impose aux systèmes à haut risque est une description opérationnelle précise du problème que cette étude pose. La frontière de délégation n'est pas seulement une question de gouvernance interne ; c'est une obligation de conformité.

---

## 6. On ne délègue pas un agent — on délègue quatre fonctions

La frontière n'est pas une ligne unique. Elle est différente selon ce qu'on délègue — et c'est justement là que le raisonnement « agent » ou « pas agent » échoue.

[12] Règlement (UE) 2024/1689 (AI Act), article 14 « Contrôle humain ». https://artificialintelligenceact.eu/article/14/




Le cadre Parasuraman/Sheridan[13] décompose l'interaction humain-machine en quatre fonctions cognitives distinctes : **acquérir l'information** (sourcer, collecter, surveiller), **analyser et synthétiser** (comparer, résumer, produire des recommandations), **décider** (trancher, approuver, rejeter), **exécuter** (agir dans le monde réel — envoyer, publier, déclencher).

La thèse de ce cadre : ces quatre fonctions n'ont pas la même sensibilité à la délégation. **Acquérir et analyser** peuvent être fortement déléguées — c'est ce que font la plupart des agents aujourd'hui, et les gains de productivité sont réels. **Décider et exécuter** méritent une délégation basse, surtout pour les décisions à conséquences irréversibles ou à fort impact sur des personnes.

Le profil « par défaut » — ce qui s'installe quand personne n'a posé la question — tend à déléguer la décision au même niveau que l'analyse. C'est la sédimentation. Le profil « tenable » — celui que le réagencement recommande — maintient la décision en zone humaine ou frontière, même quand l'acquisition et l'analyse sont complètement déléguées.

Ce schéma est un **cadre d'analyse et un profil illustratif**, pas une mesure de terrain. Il sert à poser la question fonction par fonction, pas à y répondre à la place de chaque organisation.

---

## 7. L'angle mort français

L'honnêteté de ce dossier tient en une phrase : il n'existe aucune mesure française de la frontière de délégation réellement pratiquée.

**Angle mort majeur.** Toutes les données de cette étude sont mondiales ou anglo-saxonnes : Parasuraman/Sheridan (académique international), OWID/AI Incident Database (mondial), McKinsey (105 pays, surreprésentation grandes entreprises internationales), AI Act (texte européen, pas encore d'évaluation de mise en œuvre). Il n'existe pas, à ce jour, de mesure française publiée

[13] Parasuraman, Sheridan & Wickens, « A Model for Types and Levels of Human Interaction with Automation », IEEE Transactions on Systems, Man, and Cybernetics — Part A, vol. 30, n° 3, 2000. D'après Sheridan & Verplanck (1978). https://doi.org/10.1109/3468.844354

sur : qui décide quoi dans les organisations françaises qui ont déployé des agents, à quel barreau de délégation ces agents opèrent pour chaque type de décision, quels sont les déclencheurs de reprise en main, combien d'incidents ont eu lieu en lien avec une délégation excessive. C'est exactement ce que le **module 05 du baromètre décideurs-IA** doit produire : cartographie des décisions déléguées par fonction, déclencheurs de reprise en main déclarés, incidents liés à la délégation — étalonné sur l'échelle de délégation pour rendre le cadre citable côté terrain français.

---

## Le déclic ?

Il n'y a pas de déclic dans la gouvernance de la délégation. Pas de moment où l'organisation « comprend » et règle le problème une fois pour toutes.

Ce qui existe, c'est un travail de fond : poser la question des quatre fonctions pour chaque agent déployé, décider explicitement où est le point de contrôle et qui l'exerce, documenter ce choix, le réviser quand un incident le remet en cause. La frontière ne se décrète pas — elle se conçoit, fonction par fonction, décision par décision. Et elle se surveille, parce qu'elle glisse.

L'AI Act a nommé le biais d'automatisation. Il reste aux organisations à en faire un critère de design, pas un risque de conformité qu'on gère à la fin.

---

## Sources et références

---

Journal de recherche en cours — Réagencer, juin 2026. Données ouvertes au 2026-06-24.