L'agent qui décide n'a pas de signataire

Pourquoi la deuxième vague de shadow IT n'est plus un sujet d'IT, mais une crise d'imputabilité décisionnelle

Une décision a été prise hier, au nom de l'entreprise, par un agent. Un client la conteste aujourd'hui. Sa reconstitution exige quatre conditions cumulatives : la version exacte du modèle exécuté à l'heure de l'appel, l'état de la mémoire de l'agent au moment de la prise de décision, la composition exacte de la chaîne d'agents et de connecteurs ayant convergé vers cette décision, et l'intégrité vérifiable de chacun de ces éléments à la date de l'incident. Aucune de ces conditions n'est réunie. La version du modèle n'est plus disponible chez le fournisseur. La mémoire n'a jamais été journalisée par la plateforme citoyenne qui l'a hébergée. La chaîne d'appels n'est exposée nulle part dans une représentation versionnée. Les connecteurs n'étaient pas verrouillés sur une signature. La décision a eu lieu. Elle a engagé l'entreprise. Elle n'est pas reconstructible.

Cette situation n'est pas un cas limite. Elle est un mode de fonctionnement structurellement encouragé par l'agentique citoyenne telle qu'elle se diffuse aujourd'hui dans la plupart des grandes organisations. Elle est aussi, et c'est l'objet de cet article, *l'apparition d'une catégorie de risque dont le droit positif a déjà nommé les contours sans que la pratique d'entreprise ne les ait encore intégrés*. Le shadow IT classique reproduisait. Le shadow IT agentique décide. Une macro hérite des droits de l'utilisateur ; un agent hérite de son jugement. Et lorsque ce jugement engage l'entreprise sans qu'aucune chaîne reconstructible ne permette d'en désigner le porteur, on n'a pas affaire à de la dette technique. On a affaire à une *dette décisionnelle non assignable* : un acte qui produit des effets au nom de l'organisation sans signataire identifiable, c'est-à-dire sans personne, fonction ou organe capable d'assumer, reconstruire et défendre la décision au nom de l'organisation, au sens où le contrôle interne, la responsabilité civile, le RGPD et l'AI Act entendent ce terme.

La thèse défendue ici est qu'aucune organisation soumise à responsabilité ne peut, en l'état du droit applicable depuis 2024, continuer à industrialiser cette pratique sans décision explicite, seuil documenté, et capacité de reconstruction proportionnée à l'impact.

1. De quoi parle-t-on

Quatre objets se présentent aujourd'hui sous la même étiquette « automation » et il faut les séparer avant tout autre travail.

La *macro VBA*, telle qu'elle a prospéré entre 1997 et 2015, est un automate déterministe local. Elle s'exécute sur le poste de l'utilisateur, dans un classeur dont l'état est intégralement observable, avec des instructions reproductibles. La défaillance est localisable à une ligne de code, le diagnostic est possible au prix d'un effort raisonnable, et l'objet ne sort pas du périmètre maîtrisé par l'organisation.

Le *RPA classique* (UiPath, Blue Prism, Automation Anywhere) prolonge cette logique à l'échelle de plusieurs systèmes mais conserve le déterminisme : un robot RPA fait, à chaque exécution, exactement la même séquence de clics sur les mêmes champs des mêmes interfaces. Sa défaillance la plus typique (un sélecteur DOM modifié par une mise à jour applicative) reste localisable, et son audit s'appuie sur des artefacts stables. Le RPA est, à cet égard, le dernier représentant pleinement gouvernable de la lignée des automates citoyens.

L'*agent générique citoyen*, troisième catégorie, est ce qui se construit aujourd'hui sur les plateformes no-code et low-code grand public, qu'elles soient indépendantes ou intégrées à un éditeur dominant. L'utilisateur métier compose visuellement un workflow, branche des connecteurs vers son CRM, sa messagerie, son ERP, son outil de ticketing, et appelle un modèle de langage généraliste pour les étapes de raisonnement. La promesse est l'autonomie locale ; le mode de production est l'opacité par construction. Ce n'est pas un défaut accessoire de l'outil : c'est sa proposition de valeur.

L'*agent architecturé*, quatrième catégorie, est ce qu'une organisation construit lorsqu'elle traite l'agent comme un système d'information à part entière : modèles versionnés, contrats d'interface entre composants, identité distincte de l'utilisateur appelant, journalisation reconstructible, propriétaire de chaîne identifié, validation formelle préalable à toute mise en production. La distinction entre agentique citoyenne et agentique architecturée n'est pas une distinction de plateforme ; c'est une distinction de régime de gouvernance. La même brique technique peut, selon la discipline organisationnelle qui l'entoure, basculer d'un côté ou de l'autre de la frontière. Et la frontière n'est pas étanche : *un système composite est gouvernable au niveau de son maillon le plus faible*. Une chaîne contenant un seul appel non gouverné est une chaîne non gouvernée. Il n'existe donc pas, dans la réalité opérationnelle, de zone propre isolée d'une zone citoyenne. Il existe une gouvernance continue, ou une gouvernance perdue.

La critique qui suit vise la troisième catégorie, et sa contamination prévisible de la quatrième.

2. La non-localisabilité causale comme propriété du système

Le point décisif n'est pas que l'agent puisse se tromper. Tous les systèmes se trompent. Le point décisif est que l'agent puisse se tromper *sans qu'il soit possible, ex post, de reconstituer la cause de l'erreur dans un espace stable*.

Cette propriété, que l'on peut nommer *non-localisabilité causale*, distingue qualitativement l'agent générique citoyen des automates qui l'ont précédé, et constitue le terme doctrinal central de cet article. Elle se compose de trois sous-propriétés solidaires.

La rupture du déterminisme. Une macro VBA produit un bug localisable à une ligne de code identifiée, reproductible en re-jouant la macro avec les mêmes entrées. Un agent composite produit une décision dont le rejouement, *même en principe*, ne garantit pas le même résultat : le modèle sous-jacent est probabiliste, la fenêtre de contexte modifie l'état informationnel disponible à chaque exécution, l'état interne de l'agent évolue. Cette propriété n'est pas un retard de pratique que la maturité comblera ; elle est intrinsèque au substrat. Les outils de gouvernance hérités du logiciel déterministe (revue de code, tests unitaires, traçabilité ligne à ligne) sont, sur ce plan, structurellement insuffisants. La gouvernance applicable doit être *probabiliste, statistique, et contractuelle*, pas seulement procédurale.

L'externalisation par défaut. La macro était locale, exécutée dans le périmètre maîtrisé de l'utilisateur. L'agent court chez un fournisseur dont le modèle est mis à jour, déprécié, ou retiré sans contrat opposable sur la version exécutée hier. Les conditions générales standard des éditeurs de modèles de fondation prévoient le droit de modifier ou de retirer une version, parfois avec un préavis court, sans engagement de fournir, à un tiers en contentieux, l'instance exacte qui aura produit la décision contestée. La conséquence est qu'au moment où une organisation aurait besoin de prouver ce qui a été décidé et comment, elle se trouve dans la situation d'un constructeur qui ne pourrait pas accéder aux plans du véhicule mis en circulation parce que le sous-traitant en a effacé la version exacte.

L'opacité de la chaîne. Un agent composite mobilise, au moment de sa décision, une suite d'appels imbriqués (agent appelant agent, outil appelant outil) dont la topologie effective n'est pas exposée par les plateformes citoyennes dans une représentation auditable. La chaîne existe dans l'instant de l'exécution ; elle n'est pas conservée dans une forme reconstructible. Les connecteurs ont pu changer silencieusement parce qu'une API distante a été dépréciée, remplacée, ou simplement ré-paramétrée par son éditeur. Le point d'audit observable a posteriori n'est pas la chaîne ; c'est, au mieux, sa trace résiduelle.

Ces trois propriétés ne sont pas des risques accessoires. Elles sont *constitutives* de l'agentique citoyenne dans son régime de production actuel. Elles définissent un système dont l'erreur n'est pas localisable, et dont la décision n'est, en conséquence, pas reconstructible.

3. Qualification juridique : RGPD, AI Act, contrôle interne

La non-localisabilité causale est une propriété du système. Sa qualification juridique est un travail distinct, et il faut le mener avec un soin particulier sur le périmètre, parce que la critique perd toute force si elle survend la portée des textes mobilisés.

Le RGPD encadre, depuis 2018, la décision exclusivement automatisée produisant des effets juridiques ou significatifs sur la personne concernée (Article 22). Il pose une *interdiction générale*, assortie d'exceptions strictes, et impose dans tous les cas le droit pour la personne concernée d'obtenir une intervention humaine, d'exprimer son point de vue, et de contester la décision (Considérant 71). La Cour de justice de l'Union européenne, dans son arrêt SCHUFA du 7 décembre 2023 (C-634/21), a précisé qu'un score automatisé jouant un rôle déterminant dans une décision prise par un tiers, même formellement humain, relève bien de l'Article 22. Cet arrêt ferme la porte de sortie habituelle (« nous ne sommes pas exclusivement automatisés puisqu'un humain valide ») dès lors que la validation humaine est, dans les faits, dictée par la sortie de l'agent.

Le périmètre exact mérite attention : *le RGPD ne capture pas toute agentique citoyenne ; il capture précisément les cas où l'organisation ne peut plus prétendre que l'agent n'était qu'un outil*. Un agent qui rédige un brouillon que son auteur réécrit substantiellement n'est pas concerné. Un agent qui produit la sortie qui sera, en pratique, transmise sans amendement à la personne concernée l'est. Cette distinction se joue sur la nature effective de la chaîne décisionnelle, pas sur sa description formelle.

L'AI Act, entré en vigueur le 1er août 2024, impose pour les *systèmes à haut risque* (au sens de son Annexe III, qui couvre notamment certains usages en RH, scoring crédit, accès aux services essentiels, justice, infrastructures critiques) une journalisation automatique des événements pertinents (Article 12), une supervision humaine effective (Article 14), une conservation des journaux par le déployeur (Article 19), et une responsabilité explicite tout au long de la chaîne de valeur (Article 25). La portée du règlement n'est pas universelle : les obligations contraignantes visent en priorité les systèmes haut risque et les systèmes à usage général posant un risque systémique. Mais lorsque l'usage franchit ces catégories, ou produit des effets comparables sur des personnes, *l'AI Act fournit déjà la grammaire de gouvernance que les plateformes citoyennes ne satisfont pas par défaut*.

Au-delà de ces deux régimes, la responsabilité du responsable de traitement (RGPD Article 24), la responsabilité civile au titre de l'Article 82, et les obligations de contrôle interne (LSF, SOX selon les cas, dispositifs sectoriels MIFID, Solvabilité II, Bâle pour les acteurs financiers, dispositifs ASN, ANSM, ARS pour les acteurs de la santé) imposent, partout, une *capacité de démonstration* : la possibilité de retracer, à la demande, comment une décision a été prise, par qui, sur la base de quelles données, avec quel processus de validation. Le périmètre exact varie ; le principe ne varie pas.

La rupture est la suivante. Tous ces régimes supposent qu'une décision prise au nom d'une organisation puisse être *retracée jusqu'à un signataire identifiable* et *reconstruite dans ses motifs*. La promesse opérationnelle de la plateforme citoyenne, à l'inverse, est une exécution rapide qui se passe précisément de cette traçabilité. La collision n'est pas une zone grise à explorer ; elle est l'affirmation simultanée de deux thèses inconciliables. Tant qu'elle n'est pas explicitée au niveau du COMEX, le droit applicable rend déjà difficilement défendable la pratique courante de l'agentique citoyenne sans cadre.

4. Le multiplicatif comme mode par défaut du design produit

On objectera que tout cela ne vaut que pour des compositions complexes, et que la plupart des agents citoyens sont des automatismes ponctuels sans prétention. L'objection ne tient pas au regard du mode opératoire des plateformes elles-mêmes.

Le rayon d'effet, d'abord, n'est pas comparable à celui des automates antérieurs. Une macro modifie un classeur. Un agent agit dans des systèmes de production : il met à jour un CRM, envoie un message au nom du salarié, modifie un ticket, alloue une ressource, signe une réponse contractuelle, déclenche un paiement, ajuste un prix. La défaillance d'une macro coûte une après-midi de retraitement ; la défaillance d'un agent peut coûter une fuite de données personnelles relevant de l'Article 33 du RGPD, une promesse contractuelle non honorée, un biais discriminatoire dans un pré-tri RH, une décision financière documentée par une donnée erronée. Le rayon d'effet n'est pas additif par rapport au shadow IT précédent ; il est dans une autre catégorie d'engagement.

La composition, ensuite, n'est pas un usage avancé que l'on découvre après six mois de pratique. Elle est *l'argument commercial principal* des plateformes agentiques génériques. Marketplaces de connecteurs comportant plusieurs milliers de templates, agents qui appellent d'autres agents par chaînage graphique, bibliothèques de workflows pré-assemblés, intégrations multi-systèmes proposées par défaut dès l'écran d'accueil : tout, dans le design produit, encourage la composition. La complexité effective d'un workflow agentique en production dépasse, en quelques semaines d'usage, la

compréhension du créateur citoyen qui l'a assemblé. Selon Gartner, un Fortune 500 global moyen pourrait dépasser cent cinquante mille agents en exploitation en 2028, contre moins d'une quinzaine en 2025, et seulement 13 % des organisations estiment aujourd'hui disposer de la gouvernance adéquate face à cette prolifération.

La conséquence opérationnelle de la composition non déclarée échappe au métier qui la produit. Lorsqu'un agent A appelle un agent B, qui appelle un connecteur C vers un modèle externe D, dont la sortie alimente un agent E qui prend la décision finale, l'incident sur l'un quelconque des cinq maillons produit un effet sur la décision sans qu'aucun signal explicite ne le rattache à sa cause. La dette ne s'additionne pas, *elle se compose au sens algébrique*. Un système à n maillons non gouvernés n'a pas n défauts potentiels : il a un nombre de chemins de défaillance qui croît avec les chemins d'interaction entre maillons, et non avec leur seul nombre. Une fraction seulement de ces chemins est observable depuis n'importe quel point d'audit donné. Le multiplicatif n'est pas un dérapage. C'est le régime nominal.

Auto-immunisation : ce qui est critiqué ici n'est pas la composition agentique en tant que telle. Les architectures composées sous discipline d'ingénieur sont, par ailleurs, ce que défendent les travaux les plus sérieux sur les jumeaux numériques régulés et sur l'IA en environnement clinique. La composition est nécessaire ; c'est la composition *non déclarée*, sans contrat d'interface, sans propriétaire de chaîne, sans capacité de reconstruction décisionnelle, qui est intenable. La distinction est analogue à celle qui sépare, dans le génie civil, un assemblage par platelage et un assemblage par soudure normée. Les deux relient des éléments. Un seul soutient une responsabilité.

5. Audit trail n'est pas gouvernance opposable

Les plateformes agentiques citoyennes sont vendues, sans exception notable, comme remède au shadow IT. Le vocabulaire est constant : *audit trails natifs*, *observability* intégrée, *governance ready*, *enterprise-grade compliance*, *fully traceable*. Cette promesse est, dans son intention, sincère ; elle n'est pas, dans sa réalisation, opposable.

Le point décisif n'est pas marketing. Il est juridique, et il se formule ainsi : *un audit trail qui ne permet pas de reconstruire la décision dans un espace stable est juridiquement et opérationnellement inutile*. Un log technique liste des événements horodatés. Une gouvernance opposable reconstruit une décision : elle restitue, à la demande, l'instance du modèle, l'état de la mémoire, la version des connecteurs, la chaîne d'appels effective, la donnée d'entrée non altérée, le résultat tel qu'il a été produit, et le chemin par lequel ce résultat est devenu un acte engageant. La distance entre les deux objets est l'écart entre une fiche d'incidents et un acte notarié.

L'AI Act, dans ses Articles 12, 14 et 19, ne se contente pas d'imposer la conservation d'événements ; il vise une journalisation *conçue pour permettre la reconstruction*. Le RGPD, dans son Article 24, charge le responsable de traitement de mettre en œuvre les *mesures techniques et organisationnelles appropriées pour démontrer que le traitement est conforme*. L'Article 82 prévoit la responsabilité civile pour tout dommage résultant d'une violation, et l'Article 83 fixe les sanctions administratives à des niveaux qui rendent l'exposition matérielle. La confusion, aujourd'hui dominante au niveau exécutif, entre la disponibilité d'un journal et la capacité de démonstration, est l'erreur catégoriale qui rend la première signature possible. Tant qu'elle n'est pas dissipée, le rapport d'audit interne décrira un dispositif conforme là où le contentieux constatera une décision orpheline.

6. Le précédent VBA, et ce que nous désapprenons

L'argument historique est court, et n'est pas mobilisé pour effrayer. Il est mobilisé pour disqualifier l'objection selon laquelle l'organisation apprendra en marchant.

Entre 1997 et 2015, les macros VBA ont produit un patrimoine ingérable dans la quasi-totalité des grandes organisations. Audits successifs, projets de cartographie, migrations de plateformes, dispositifs ALM citizen, cellules de gouvernance dédiées : il a fallu, dans la plupart des secteurs, entre quinze et vingt ans pour ramener cet héritage dans un périmètre de contrôle, au prix d'incidents répétés (modèles financiers défaillants, calculs réglementaires erronés, erreurs de production) et d'investissements considérables. Cet apprentissage est suffisamment documenté pour que l'argument de l'ignorance ne soit plus recevable.

L'observation à faire est la suivante : *la même organisation qui a investi pour assainir la dette VBA reproduit, sur un substrat technologique présenté comme remède, exactement le motif qu'elle prétendait avoir corrigé*. Mêmes commanditaires métiers, même contournement de la DSI, même opacité par construction, même absence de propriétaire identifié au moment où le créateur change de fonction. Cette régularité n'est pas une coïncidence d'inattention ; elle suggère que le mécanisme institutionnel produisant le shadow IT n'a jamais été traité, et qu'il reproduira la pathologie sur tout substrat offrant aux métiers une voie de contournement de la gouvernance. Nous avons déjà appris. Nous sommes en train de désapprendre. Et la fenêtre n'est pas de vingt ans cette fois.

7. Pourquoi la fenêtre est de douze à vingt-quatre mois

Quatre paramètres rendent la diffusion qualitativement différente de celle des macros, et concentrent dans un horizon court ce qui s'était étalé sur deux décennies.

La *vitesse de diffusion* est supérieure de plusieurs ordres de grandeur. L'interface en langage naturel abaisse la barrière d'entrée à un niveau qu'aucune génération précédente d'outil n'avait atteint. Le marketing, descendant depuis le PDG qui a vu une démonstration, met l'organisation en posture d'adoption avant toute délibération de gouvernance. Selon Gartner, environ 40 % des applications d'entreprise intégreront un agent à des tâches spécifiques d'ici fin 2026, contre moins de 5 % en 2025. Selon le Cyber Pulse Report 2026 publié par Microsoft, plus de 80 % des Fortune 500 exploitent désormais activement des agents construits avec des outils low-code ou no-code, et 29 % des salariés déclarent recourir à des agents non sanctionnés par leur organisation.

La *dépendance externe* est immédiate, à la différence de la macro qui restait locale. Modèle, connecteurs, infrastructure : trois sources de dérive incontrôlable installées dès le premier déploiement. Le *coût marginal d'expérimentation* est quasi nul, ce qui signifie que la prolifération n'est pas freinée par un arbitrage budgétaire local. Et la *surface d'action* est, dès le premier jour, en systèmes de production, parce que l'intérêt même de l'agent est d'agir, et non de calculer.

Conséquence : la dette agentique ne met pas vingt ans à apparaître. Elle devient critique au moment où un premier incident exige une reconstruction décisionnelle qui n'existe pas. Les indicateurs avancés disponibles à fin 2025 et début 2026 (selon l'IBM *Cost of a Data Breach Report* 2025, 37 % seulement des organisations disposent d'une politique formelle de gouvernance IA ; selon Netskope, l'entreprise moyenne enregistre de l'ordre de deux cent vingt violations mensuelles de politique de données liées à l'usage de la GenAI) suggèrent une fenêtre de l'ordre de douze à vingt-quatre mois entre l'industrialisation de l'usage et l'incident de gouvernance opposable. Ces chiffres ne décrivent pas un risque marginal ; ils décrivent une trajectoire plausible de matérialisation rapide, qui ne laisse pas le temps des apprentissages séquentiels caractéristiques de l'ère VBA.

8. Quatre décisions, pas quatre principes d'ingénierie

La proposition n'est pas une bonne pratique IT. Elle est une délibération exécutive à inscrire formellement au plus haut niveau du dispositif de contrôle. Quatre décisions, prises explicitement, séparent l'organisation qui signe encore de l'organisation qui a cessé de signer ce qu'elle ne peut pas reconstruire.

Première décision : seuil de criticité explicite. L'organisation doit définir, par délibération formelle, le seuil au-delà duquel un agent est traité comme système critique au sens du dispositif de contrôle interne. Le seuil mobilise les axes habituels de criticité : impact financier (volume des engagements pris), impact contractuel (engagements vis-à-vis de tiers), impact RGPD (traitement de données personnelles, décision automatisée au sens de l'Article 22), impact AI Act (qualification haut risque au sens de l'Annexe III, en particulier RH, scoring crédit, accès aux services essentiels, infrastructures critiques, justice). En deçà du seuil, sandbox déclarée avec périmètre borné. Au-delà, obligations identiques à celles d'un système d'information critique, sans exception liée à la plateforme de construction. L'absence de seuil délibéré n'est pas la neutralité : c'est le seuil zéro, c'est-à-dire l'engagement implicite à signer toute décision agentique quel qu'en soit l'impact.

Deuxième décision : interdiction explicite de patterns. Quatre patterns doivent être bannis par décision écrite, et leur détection doit déclencher une alerte au niveau de la conformité, pas au niveau de l'IT : composition d'agents non déclarée à la cartographie ; agent en production sans propriétaire identifié, nominativement, avec un suppléant désigné et un cycle de revue ; agent ayant accès en écriture à un système de production sans validation formelle de la chaîne complète d'appels ; agent dépourvu de capacité de reconstruction décisionnelle (instance du modèle, mémoire, connecteurs, chaîne, entrée, sortie, chemin de décision). Ces interdictions ne suppriment pas l'agentique citoyenne ; elles la cantonnent à son domaine de validité.

Troisième décision : kill switch décisionnel, distinct du kill switch technique. La capacité technique de suspendre l'exécution d'un agent existe sur la plupart des plateformes. La capacité organisationnelle de suspendre la *valeur engageante* de ses décisions, le temps qu'une chaîne humaine reprenne la main, n'existe presque jamais. Cette seconde capacité doit être formalisée : circuit de remontée, pouvoir explicite d'invalider rétroactivement les actes émis dans une fenêtre définie, dispositif de notification aux parties prenantes contractuelles. Le kill switch décisionnel ne supprime pas les décisions déjà émises ; il cantonne leur portée le temps de la reconstruction. Il est, sur ce plan, l'équivalent agentique du rappel produit dans l'industrie manufacturière.

Quatrième décision : circuit de validation indexé sur l'impact métier, pas sur la technologie. Le critère de qualification d'un agent ne doit pas être la plateforme de construction, mais l'impact métier de ses actes. Un agent rédigeant un compte-rendu interne ne relève pas du même circuit qu'un agent clôturant un ticket client avec engagement contractuel implicite. Cette indexation par l'impact, et non par la technologie, est ce qui empêche l'arbitrage par la commodité technique de produire un arbitrage de fait sur la responsabilité.

Ces quatre décisions ne sont pas des principes. Elles sont des *signatures à apposer*, par les organes de gouvernance compétents, sous une forme reproductible par un auditeur. Tant qu'elles n'ont pas été apposées, l'organisation continue à signer en aveugle.

9. Articulation au corpus

La distinction posée ici (*agentique citoyenne* opposée à *agentique architecturée*, et plus profondément *décision avec signataire* opposée à *décision sans signataire*) est cohérente avec le geste épistémique mobilisé ailleurs dans ce corpus. Elle est analogue, dans l'ordre de l'ingénierie, à la distinction entre paradigme LLM-centré et architecture composite régulée. Elle est analogue, dans l'ordre de l'architecture, à la condition d'hexagonalité comme prérequis de gouvernabilité (référence Soleau TI-2026-ART5-ES). Elle est analogue, dans l'ordre opérationnel, à la nécessité d'une architecture événementielle pour produire l'observabilité que les plateformes citoyennes ne livrent pas.

Le motif récurrent est le même : nommer la distinction structurante qui sépare deux régimes de fonctionnement, l'un viable sous un cadre régulé, l'autre non. La présente contribution projette ce motif sur le plan de l'imputabilité. Elle propose deux termes doctrinaux articulés. Le premier, *non-localisabilité causale*, désigne la propriété du système : un agent dont l'erreur ne peut être ni reproduite, ni rattachée à un maillon stable, ni reconstruite à partir des artefacts disponibles. Le second, *dette décisionnelle non assignable*, désigne la conséquence organisationnelle : un acte engageant l'entreprise sans signataire identifiable au sens du droit applicable. Les deux termes sont solidaires. Le premier est conceptuel ; le second est opérationnel. La distinction *agentique citoyenne / agentique architecturée* est, sur ce plan, une distinction de second rang : c'est l'opérateur qui sépare les régimes de production produisant l'une ou l'autre catégorie de dette.

10. Limites

Quatre objections méritent d'être traitées sans complaisance.

Le seuil de criticité est non trivial à fixer. C'est exact. Il n'existe pas de seuil universel ; il existe un seuil propre à chaque organisation, fonction de son secteur, de son exposition réglementaire, et de son appétit de risque. Cette difficulté n'est pas une raison de ne pas délibérer. Elle est précisément l'objet de la délibération.

La frontière agentique citoyenne / agentique architecturée est graduelle, pas binaire. C'est exact, et c'est traité explicitement à la section 1. La gouvernance applicable est

continue, pas catégorielle ; elle se perd au premier maillon non gouverné. Ce qui ne signifie pas qu'il n'y ait pas de seuil décisionnel : la décision est de savoir à partir de quel point l'organisation accepte de s'engager.

Les organisations à faible maturité IT n'ont pas la masse critique pour l'agentique architecturée. C'est exact. La conséquence n'est pas qu'elles peuvent industrialiser l'agentique citoyenne sans cadre, mais que la décision d'industrialisation, dans ce cas, devrait être suspendue. C'est précisément la décision que ces organisations évitent de prendre, et c'est l'évitement de cette décision, plutôt que l'absence de capacité, qui produit le risque.

Le cadre régulé déplace mais n'élimine pas la dette. C'est exact. L'objectif n'a jamais été d'éliminer la dette ; il est de la rendre *assignable*. Une dette assignable est traitable ; une dette non assignable ne l'est pas, parce qu'aucune contrepartie ne peut en être responsabilisée. La proposition vise ce déplacement, pas une utopie de risque zéro.

11. Conclusion

L'organisation dispose, à la date de ce texte, de vingt-cinq ans d'expérience documentée sur l'échec de gouvernance des macros, de huit ans d'application du RGPD sur la décision automatisée, et de l'entrée en vigueur effective de l'AI Act depuis l'été 2024. Que ces trois corpus convergent pour rendre intenable la pratique courante de l'agentique citoyenne sans cadre, et que l'organisation continue néanmoins d'industrialiser cette pratique sans les contremesures nécessaires, est une régularité qui doit être nommée pour être interrompue.

Le problème n'est pas qu'un agent puisse se tromper. Le problème est qu'une entreprise puisse être engagée par une décision qu'elle ne sait *ni reconstruire, ni expliquer, ni assigner*.

Tant qu'un comité exécutif classe ce sujet à la direction des systèmes d'information, il signe sans le savoir des décisions sans signataire. La question n'est pas de savoir s'il faut construire de l'agentique. Elle est de savoir à partir de quel point l'organisation cesse de signer ce qu'elle ne peut pas reconstruire.

Références

Cadres réglementaires

Règlement (UE) 2016/679 (RGPD), articles 22 (décision individuelle automatisée), 24 (responsabilité du responsable de traitement), 33 (notification de violation), 82 (droit à indemnisation), 83 (sanctions administratives) ; Considérant 71.

Règlement (UE) 2024/1689 (AI Act), articles 12 (record-keeping pour systèmes haut risque), 14 (supervision humaine), 19 (logs automatiques générés et conservés par le déployeur), 25 (responsabilité tout au long de la chaîne de valeur), 26 (obligations des déployeurs), Annexe III (catégorisation des systèmes haut risque). Entrée en vigueur 1er août 2024.

Cour de justice de l'Union européenne, affaire C-634/21, OQ contre Land Hessen (SCHUFA), arrêt du 7 décembre 2023.

Données de cadrage

Gartner, *Press Release : 40 % of Enterprise Apps Will Feature Task-Specific AI Agents by End of 2026*, 26 août 2025.

Gartner, *Press Release : Six Steps to Manage AI Agent Sprawl* (référence aux projections de cent cinquante mille agents par Fortune 500 en 2028 et au taux de 13 % d'organisations estimant disposer de la gouvernance adéquate), 28 avril 2026.

Gartner, *Top Predictions for Data and Analytics in 2026*, 11 mars 2026.

Microsoft, *Cyber Pulse Report 2026* (taux de 80 % de Fortune 500 exploitant activement des agents low-code / no-code, et 29 % de salariés recourant à des agents non sanctionnés).

Netskope, *Cloud and Threat Report 2026* (de l'ordre de 220 violations mensuelles moyennes de politique de données liées à l'usage de la GenAI).

IBM, *Cost of a Data Breach Report 2025* (37 % d'organisations dotées d'une politique formelle de gouvernance IA).

Forrester, *AEGIS Framework : Agentic AI Enterprise Guardrails for Information Security* ; *Predictions 2026*.

World Economic Forum, *Global Cybersecurity Outlook 2026*.