GUILLAUME VIGNERON

Gouverner l’IA avant qu’elle ne vous gouverne.

Leadership, gouvernance et décision humaine
à l’ère de l’IA générative



Livre blanc

Novembre 2026

Sommaire

Avant-propos 3

01 Le jour où nous avons commencé à déléguer notre jugement 4

02 Décider mieux, plus vite ou différemment ? 7

03 Qui doit poser les questions : l'homme ou la machine ? 12

04 Gouverner l'IA : reprendre la maîtrise de ses usages 16

05 Le leadership à l'épreuve de l'intelligence artificielle 22

06 De la réflexion à l'action : gouverner l'IA en 90 jours 26

Conclusion : le courage de continuer à décider 32

Glossaire 33

Références 34

L'auteur 36

Pour préparer un comité de direction : diagnostic en page 29 et fiche d'arbitrage en page 30. Les cas fictifs sont développés aux pages 10, 24 et 31.

Pourquoi j’ai écrit ce livre

Mon parcours associe la conception de systèmes d’information, le conseil en organisation et l’entrepreneuriat digital. J’ai créé et dirigé une agence digitale. Cette expérience m’a conduit à regarder les outils à travers les décisions qu’ils rendent possibles, mais aussi à travers les contraintes qu’ils installent. Un système ne se contente pas de traiter une information. Sa conception détermine ce qui sera visible, ce qui restera en dehors du dossier et ce que ses utilisateurs pourront effectivement modifier.

Aujourd’hui, comme consultant, formateur et conférencier spécialisé en intelligence artificielle générative, je m’intéresse moins à la démonstration d’une capacité qu’à ses conséquences sur le travail. Une réponse obtenue en quelques secondes peut aider à comprendre un problème. Elle peut aussi éviter à son destinataire l’effort de le formuler. Ce déplacement m’intéresse parce qu’il touche au jugement, donc à la responsabilité.

Le titre de ce livre est volontairement provocateur. L’IA ne gouverne pas une organisation par une volonté propre. Le problème apparaît lorsque ses propositions deviennent des décisions sans que personne ait clairement choisi ce transfert d’autorité. On conserve les signatures et les réunions. Pourtant, les hypothèses viennent d’ailleurs, les options ont déjà été ordonnées et le désaccord demande désormais un effort supplémentaire.

Je défends une position précise : l’IA peut élargir la réflexion si nous décidons de la place qu’elle occupe dans notre raisonnement. Cette place ne se règle pas avec une formule générale sur le contrôle humain. Elle dépend de la décision concernée et des moyens réels de la contester. Je souhaite que ce livre aide son lecteur à reconnaître ces différences, puis à les traduire dans son organisation.

Les méthodes qui suivent sont mes propositions de travail. Elles ne constituent pas un modèle scientifiquement validé. Les résultats de recherche sont signalés par des références numérotées ; les situations professionnelles décrites sont fictives. Ce choix permet de discuter les mécanismes sans inventer une expérience vécue. Mon objectif n’est pas de défendre chaque décision humaine contre une machine. Il est de rendre les décisions assistées plus discutables, au sens exact du terme : assez explicites pour qu’une objection puisse les faire changer.

Le jour où nous avons commencé à déléguer notre jugement

Le transfert d'autorité peut commencer par un geste banal. Une personne demande un résumé avant une réunion. Ce résumé devient le document de référence. Les participants discutent ensuite à partir de ses catégories, sans retourner aux pièces qu'il a laissées de côté. Personne n'a demandé au système de décider. Il a pourtant contribué à déterminer ce qui méritait discussion. Je considère ce moment comme une première forme de délégation du jugement.

Il serait excessif d'en conclure que tout résumé est dangereux. Nous avons toujours dépendu de médiations : une note préparatoire, un tableau de bord, l'analyse d'un collègue. La question utile porte sur la possibilité de remonter à la source. Peut-on identifier ce qui a été sélectionné ? L'auteur de la synthèse expose-t-il ses incertitudes ? Le lecteur dispose-t-il du temps nécessaire pour vérifier un point qui semble secondaire ? Une médiation devient problématique lorsque ces possibilités disparaissent.

Avec un modèle génératif, la qualité de la formulation peut donner au document une cohérence supérieure à celle des informations disponibles. Le NIST décrit les confabulations comme des productions erronées ou fausses présentées avec assurance [1]. Une rédaction précise ne suffit donc pas à établir l'exactitude. Je recommande de traiter séparément le travail de mise en forme et le travail de preuve. Une source citée doit exister, puis soutenir réellement l'affirmation qui lui est attribuée.

Le glissement suivant intervient lorsqu'une recommandation devient l'option par défaut. Il faut alors argumenter pour s'en écarter, tandis que l'accepter ne demande aucune justification équivalente. Cette asymétrie crée une autorité pratique, même si le règlement intérieur affirme que l'humain décide toujours. Le pouvoir ne réside plus seulement dans l'acte final. Il se situe dans la définition des options acceptables et dans le coût imposé à celui qui propose autre chose.

Le regard de Guillaume
Je regarde d'abord ce qu'il faut faire pour désobéir à une recommandation. Si la contestation exige plus de travail que l'acceptation, le système dispose déjà d'une forme d'autorité.

Cinq niveaux de délégation

Je propose de distinguer cinq niveaux. Cette échelle décrit une organisation du travail ; elle ne classe pas juridiquement les systèmes. Un même outil peut relever de plusieurs niveaux selon les droits qui lui sont accordés. Une messagerie qui prépare un brouillon ne pose pas les mêmes questions qu'un agent autorisé à l'envoyer à tous les clients.

01	**Assistance documentaire**	Le système recherche ou résume.
02	**Exploration**	Il propose des pistes à examiner.
03	**Recommandation**	Il classe les options ou conseille un choix.
04	**Exécution sous validation**	Il agit après une autorisation explicite.
05	**Autonomie bornée**	Il agit dans un périmètre fixé et contrôlé.

Échelle proposée par l'auteur. Les droits changent avec le niveau.

Le passage entre deux niveaux doit faire l'objet d'un choix explicite. Une équipe peut autoriser la préparation d'une réponse sans autoriser son envoi. Elle peut accepter une exécution automatique pour certaines opérations, à condition de fixer un périmètre et de pouvoir l'interrompre. L'autonomie reste alors bornée : elle ne signifie jamais que l'outil devient responsable des conséquences à la place de l'organisation.

Le cas le plus trompeur est celui de la recommandation. Aucun accès d'écriture n'est nécessaire pour influencer une décision. Un classement de fournisseurs ou de candidatures peut peser davantage qu'une action informatique réversible. Il faut donc regarder l'effet sur le choix, pas seulement les permissions techniques. Un outil sans capacité d'exécution peut demander une supervision plus exigeante qu'un automatisme très limité.

Je distingue enfin répétition et réversibilité. Une petite erreur répétée peut se propager avant d'être détectée. Une décision inhabituelle peut rester corrigible si elle prend la forme d'un essai limité. La fréquence d'une tâche ne justifie donc pas, à elle seule, son automatisation.

En pratique

Choisissez un usage existant. Écrivez ce que le système peut lire et ce qu'il peut modifier. Décrivez ensuite ce que l'utilisateur reprend sans vérification. Le niveau de délégation réel se trouve à l'intersection de ces réponses.

La confiance doit porter sur une tâche

Faire confiance à « l'IA » est une expression trop large pour guider une décision. Je peux juger un outil utile pour reformuler une note et lui refuser tout rôle dans l'évaluation de sa validité. La confiance doit porter sur un usage défini, dans des conditions identifiables. Lorsqu'une de ces conditions change, la conclusion doit pouvoir être réexaminée.

Parasuraman et Riley distinguent notamment le mauvais usage de l'automatisation, lié à une confiance excessive, et sa sous-utilisation [2]. Leur analyse de 1997 ne porte pas sur les modèles génératifs actuels. Elle rappelle néanmoins que le problème ne se résout ni par l'adhésion générale ni par la méfiance systématique. J'en tire une règle de gouvernance : la confiance doit être calibrée par des observations sur la tâche, et non par l'impression produite par une démonstration.

Il faut également reconnaître ce que coûte une vérification. Relire une phrase demande peu de temps. Évaluer une analyse de marché peut nécessiter de refaire une partie du travail. Si l'organisation promet des gains en supprimant ce temps, elle fragilise le contrôle qu'elle prétend conserver. L'expression « à vérifier par l'utilisateur » devient alors un transfert de charge sans moyens correspondants.

Je distingue le bénéfice immédiat et la capacité qui devra subsister. Une équipe peut gagner du temps aujourd'hui tout en perdant progressivement la possibilité de produire une analyse autonome. Cette évolution n'est pas inévitable. Elle dépend des occasions laissées aux personnes pour pratiquer leur métier. Le risque mérite cependant d'être suivi, car on découvre parfois trop tard qu'une procédure de secours repose sur une compétence qui n'est plus exercée.

La dépendance se mesure donc aussi par une question concrète : que se passe-t-il si le service devient indisponible demain ? Une organisation qui sait ralentir temporairement peut rester maîtresse de son activité. Une organisation qui ne sait plus expliquer ses propres décisions a un problème plus profond. Elle ne manque pas simplement d'un fournisseur de remplacement. Elle a laissé partir une partie de sa capacité de jugement.

Le point de vigilance
Un taux de réponses acceptées ne mesure pas automatiquement la qualité. Il peut également refléter un manque de temps, une consigne implicite ou l'impossibilité de reconnaître les erreurs.

Décider mieux, plus vite ou différemment ?

Je distingue la qualité d'une décision de sa vitesse de production. J'accorde aussi une place particulière à la possibilité de décider différemment, c'est-à-dire de modifier les hypothèses à partir desquelles le problème est abordé. Ces attentes peuvent se rejoindre, mais elles ne sont pas interchangeables. Un dossier plus vite terminé ne prouve pas que la décision est mieux fondée.

Mieux	Plus vite	Différemment
Critère à définir	**Délai à mesurer**	**Hypothèse à ouvrir**
Vérifier le résultat et les conséquences.	Compter aussi les contrôles et les reprises.	Explorer une alternative, puis l'éprouver.

Des attentes distinctes, à évaluer sur un usage précis.

Dans l'expérience publiée initialement par Dell'Acqua et ses coauteurs en 2023, l'aide d'un modèle améliore certaines performances de consultants, mais dégrade la réussite sur une tâche située hors de ses capacités [3]. Les résultats portent sur un dispositif et des tâches déterminés. Je n'y vois pas un rendement universel de l'IA. J'y vois une raison de tester séparément des usages qui paraissent voisins. La frontière entre une aide utile et une aide trompeuse peut traverser un même métier.

« Mieux » suppose un critère. Une décision peut améliorer la marge et détériorer la relation avec un partenaire. Elle peut satisfaire un indicateur trimestriel tout en reportant des difficultés. Avant de demander une recommandation, je souhaite donc que le décideur précise le résultat recherché et les conséquences qu'il refuse. Sans ce travail, le modèle optimise surtout une interprétation implicite de la demande.

« Différemment » ne dispense pas de ce cadre. Une idée inattendue peut être sans intérêt ou irréalisable. Je lui demande seulement de rouvrir une enquête qui s'était refermée trop vite. Son utilité dépend de ce qu'elle permet de vérifier ensuite. La nouveauté d'une réponse n'est pas une preuve supplémentaire ; elle est une invitation à chercher ailleurs.

Ce que le contexte laisse hors du dossier

Une organisation ne tient pas entièrement dans ses documents. Un engagement pris oralement, une fatigue collective ou une relation dégradée peuvent changer la portée d'une décision. Je ne suppose pas que l'humain saisit toujours correctement ces éléments. Je constate surtout qu'un système ne peut pas utiliser une information à laquelle il n'a pas accès, et que l'absence de cette information n'est pas toujours visible dans sa réponse.

Imaginons une entreprise qui envisage de centraliser son service client. Les données disponibles décrivent les volumes et les coûts. Un modèle propose une organisation cohérente à partir de ces chiffres. Mais plusieurs personnes possèdent une connaissance tacite des exceptions contractuelles. Leur travail paraît peu productif parce qu'il évite des problèmes rarement enregistrés. Si le projet supprime cette capacité sans l'identifier, le calcul est incomplet avant même d'être faux.

Je propose de construire un registre du contexte manquant. Il ne s'agit pas de rédiger un roman de l'entreprise. Pour chaque décision importante, on note ce qui pourrait changer l'analyse sans figurer dans les données. On désigne ensuite une personne capable de confirmer ou d'infirmer ces éléments. Certains resteront incertains ; cette incertitude doit apparaître dans l'arbitrage au lieu de disparaître dans une conclusion bien écrite.

Il existe aussi un contexte que nous choisissons de ne pas transmettre. La confidentialité peut interdire l'envoi de pièces à un service externe. Cette limite n'autorise pas à faire comme si les pièces étaient inutiles. Elle impose de réduire la portée de la demande ou d'utiliser un environnement adapté. Une réponse fondée sur un dossier volontairement incomplet doit rester une réponse partielle.

Enfin, davantage de contexte ne garantit pas une meilleure décision. Des informations anciennes ou contradictoires peuvent brouiller l'analyse. Je préfère un dossier dont les faits décisifs sont datés et attribués à une accumulation sans statut. Le décideur doit savoir ce qui relève d'un fait vérifié, d'une estimation ou d'un désaccord. Le modèle peut aider à organiser cette distinction ; il ne peut pas la créer honnêtement à partir d'informations absentes.

En pratique
Avant l'analyse, complétez cette phrase : « La recommandation pourrait changer si nous apprenions que… » Faites répondre séparément le responsable du projet et une personne qui en subira les conséquences.

L’expérience ne se réduit pas à une base documentaire

J’utilise une représentation conceptuelle de la décision pour rendre visibles plusieurs registres du jugement. Elle n’est pas une équation et ne permet aucun calcul. Les connaissances peuvent informer le choix ; l’expérience aide à reconnaître une situation. L’intuition peut signaler une difficulté avant que son explication soit disponible. Il reste enfin ce que la situation présente a de singulier et que nos catégories décrivent imparfaitement.

Connaissances

Ce qui peut être documenté

Expérience

Ce qui a été vécu et appris

Intuition

Un signal à interroger

Singularité

Ce que le cadre décrit mal

Jugement situé : ces registres se confrontent, ils ne s’additionnent pas.

Kahneman et Klein expliquent que l’intuition experte peut se développer lorsque l’environnement comporte des régularités suffisamment fiables et que la personne a l’occasion de les apprendre grâce au retour d’expérience [4]. La confiance subjective ne suffit pas. Cette distinction m’empêche de défendre l’intuition du dirigeant comme un privilège incontestable. Une longue carrière n’établit pas une expertise dans un environnement entièrement nouveau.

Ma référence au « je-ne-sais-quoi » de Vladimir Jankélévitch est philosophique [5]. J’emprunte cette expression pour interroger ce que la formalisation laisse de côté. Je ne lui attribue pas une théorie de l’IA ni une méthode de décision. Surtout, je refuse d’utiliser l’indicible comme un refuge contre la contradiction. Dire « je le sens » peut ouvrir une enquête ; cela ne doit pas la fermer.

Dans une réunion, je traduirais donc une intuition en question : quel détail me gêne, quelle expérience lui ressemble, quelle observation pourrait me faire changer d’avis ? Il arrive qu’aucune réponse complète ne soit disponible. On peut alors conserver cette réserve dans la décision et limiter l’engagement initial. Reconnaître une incertitude est plus utile que lui donner artificiellement la forme d’un score.

Étude de cas : l’expansion qui semblait évidente

Cas fictif. Une entreprise de services envisage d’ouvrir une nouvelle implantation. Le directeur commercial apporte un portefeuille de prospects ; la direction financière construit une projection favorable. Un modèle a rédigé une note qui recommande l’ouverture et donne une forme très cohérente aux arguments. Le CODIR doit décider s’il engage immédiatement les dépenses de lancement.

La première difficulté tient au dossier lui-même. Les prospects sont présentés comme une demande presque acquise, alors que plusieurs n’ont accepté qu’un entretien exploratoire. La projection suppose aussi que l’équipe actuelle pourra accompagner les premiers contrats sans dégrader les prestations existantes. La note ne ment pas nécessairement : elle prolonge des hypothèses qui n’ont pas reçu de statut explicite.

Je commencerais par retirer provisoirement la recommandation et demander à chaque responsable d’écrire son hypothèse principale. Le directeur commercial estime que la proximité géographique déclenchera les ventes. La responsable des opérations pense que l’obstacle réel se trouve dans la disponibilité des experts. Ces deux lectures n’appellent pas le même investissement. La première justifie des locaux ; la seconde demande une capacité supplémentaire.

L’IA reçoit ensuite une tâche limitée : proposer une explication concurrente pour chaque hypothèse et préciser les informations qui permettraient de les départager. Elle peut suggérer que les prospects cherchent surtout une spécialité technique, sans attacher d’importance à l’implantation. Cette proposition n’est pas un fait nouveau. Elle devient utile si l’équipe contacte les prospects pour tester ce qui conditionne réellement leur engagement.

Le CODIR retient finalement un essai commercial sans ouverture immédiate. Il définit les preuves attendues avant un engagement plus lourd, ainsi qu’une date de réexamen. Ce résultat n’est pas nécessairement le meilleur dans toute entreprise. Dans ce cas fictif, son intérêt est de séparer la conviction stratégique de la décision irréversible. L’IA a aidé à déplacer la question : faut-il ouvrir une implantation, ou faut-il d’abord vérifier pourquoi les clients achèteraient ?

Le regard de Guillaume
Je juge l’apport du modèle à la qualité de l’enquête qu’il provoque. Une recommandation qui ne change ni les questions ni les vérifications peut seulement embellir une décision déjà prise.

Faire de la vitesse un choix explicite

Le temps constitue une contrainte réelle. Une décision tardive peut coûter plus cher qu'une décision imparfaite. Je ne propose donc pas de ralentir systématiquement. Je propose de choisir où la vitesse est utile, et où elle détruit une possibilité de correction. La bonne unité d'analyse est le processus complet, depuis la collecte d'informations jusqu'aux conséquences observées.

Supposons qu'un assistant réduise fortement le temps nécessaire pour rédiger une proposition commerciale. Si le responsable doit ensuite vérifier chaque engagement contractuel, le gain net peut rester limité. Ce n'est pas un échec du projet ; c'est une donnée de décision. Il faut mesurer le temps de reprise et le coût des erreurs, au lieu de retenir uniquement la durée de rédaction. Une mesure locale peut encourager un transfert de travail vers une autre équipe.

Je réserve volontiers une voie rapide aux opérations dont les effets sont limités et corrigibles. Encore faut-il que la correction arrive avant que le dommage ne s'installe. Un message interne peut être modifié ; une information confidentielle envoyée à l'extérieur ne revient pas dans le système après son effacement. Le bouton « annuler » ne suffit pas à établir la réversibilité réelle.

Pour les décisions engageantes, je préfère fixer un temps minimal de contradiction. Ce temps n'a pas besoin d'être long. Il doit permettre à une personne indépendante de la préparation de rechercher l'hypothèse la plus fragile. Si la décision est urgente, on documente ce qui n'a pas pu être vérifié et on réduit, lorsque c'est possible, l'étendue de l'engagement. L'urgence doit être décrite ; elle ne doit pas devenir une justification permanente.

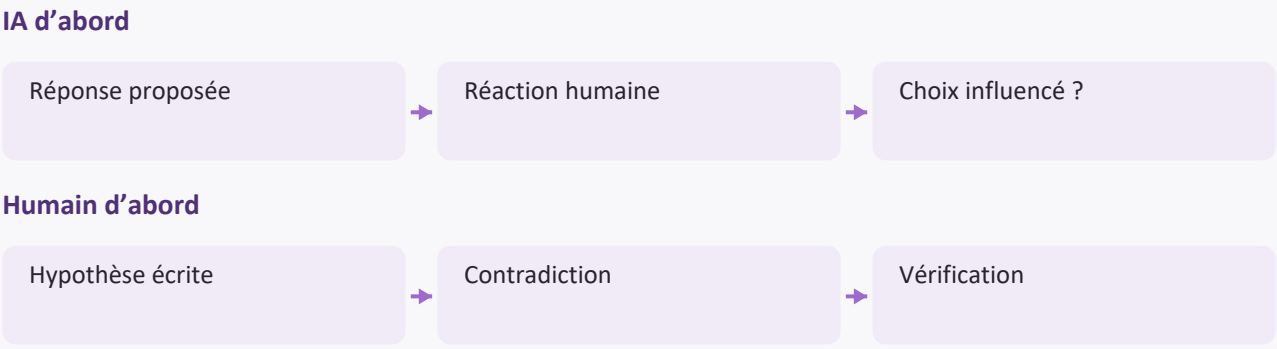
Enfin, décider différemment a un coût d'exploration. Toutes les options ne méritent pas une étude complète. Je demande d'arrêter l'exploration lorsque les alternatives restantes n'affectent plus l'arbitrage, ou lorsque l'information supplémentaire coûte davantage que ce qu'elle pourrait changer. Ce jugement reste humain. Le modèle peut multiplier les variantes sans ressentir le coût d'attention qu'elles imposent à une équipe.

En pratique
Comparez deux délais : le temps gagné avant la validation et le temps consommé après elle. Ajoutez les reprises effectuées par les destinataires. Le gain réel se mesure au bout du processus.

Qui doit poser les questions : l’homme ou la machine ?

Lorsque je demande immédiatement une réponse, j’accepte que le modèle propose le premier cadre d’analyse. Je peux ensuite le corriger, mais je travaille déjà à l’intérieur d’une structure qui n’est pas la mienne. Pour les décisions importantes, je préfère formuler une hypothèse avant de consulter l’outil. Cette préférence est une discipline de travail, pas une garantie d’indépendance intellectuelle.

Tversky et Kahneman ont étudié l’ancrage dans le jugement sous incertitude [6]. Une estimation initiale peut influencer les ajustements ultérieurs. Leur article de 1974 ne mesure pas les effets d’un dialogue avec un modèle génératif. Le rapprochement que je propose est une hypothèse d’usage : une réponse initiale détaillée peut fixer trop tôt les termes du problème. Cette analogie justifie une précaution, pas un diagnostic automatique sur chaque utilisateur.



Aucun circuit n’élimine les biais. Le second explicite le point de départ.

« Humain d’abord » comporte aussi un risque. Je peux m’attacher à ma première intuition et utiliser l’IA pour la défendre. Écrire mon raisonnement avant de consulter le modèle n’a d’intérêt que si je lui demande ensuite de rechercher ce qui pourrait le réfuter. La trace initiale doit permettre de voir ce qui change, non de protéger ma position contre le changement.

Je demande donc un bref document préalable : le choix envisagé, la raison décisive et l’incertitude principale. Après l’échange, je note l’élément qui a modifié mon analyse, ou la raison pour laquelle je la maintiens. Si rien ne peut jamais faire évoluer la conclusion, la conversation n’est pas une recherche. C’est une demande de justification adressée à un outil particulièrement disponible.

Demander à l’IA de conduire l’interrogatoire

J’appelle ici « flip prompting » le fait d’inverser la demande : le modèle pose les questions et attend les réponses au lieu de produire immédiatement une solution. Le terme désigne une pratique de dialogue, pas une méthode scientifique stabilisée. Son intérêt tient à la contrainte qu’elle impose au décideur : expliciter ce qu’il aurait laissé implicite dans une commande ordinaire.

Le questionnement socratique m’inspire une attitude : demander ce que les mots signifient et sur quoi repose une affirmation. Je n’assimile pas un dialogue généré à la pratique philosophique de Socrate. Un modèle peut produire une objection convaincante sans comprendre l’enjeu vécu de son interlocuteur. Il peut également s’égarer dans une suite de questions qui donnent une impression de profondeur sans faire avancer l’enquête.

Je propose une consigne de travail : « Ne recommande aucune solution pour l’instant. Pose une question à la fois sur mon hypothèse. Distingue ce que j’affirme de ce que je peux établir. Si je réponds par une généralité, demande un exemple vérifiable. Termine lorsque les informations manquantes sont identifiées. » Cette consigne doit être complétée par le périmètre autorisé des données. Un bon exercice intellectuel ne justifie pas une divulgation inutile.

La méthode des cinq pourquoi peut aider à remonter d’un symptôme à une explication. Je ne traite pas le nombre cinq comme un objectif. Si plusieurs causes sont plausibles, il faut ouvrir plusieurs branches. Une chaîne unique peut imposer une causalité trop simple et finir par désigner une personne comme responsable d’un problème d’organisation. La question suivante doit dépendre des éléments disponibles, pas d’une règle de mise en scène.

Je termine l’échange par une demande de preuve. Pour chaque explication encore plausible, quelle observation permettrait de la distinguer des autres ? Si toutes conduisent à la même recommandation indépendamment des faits, l’exercice n’a pas réellement testé le raisonnement. La contradiction doit produire une conséquence pratique, comme un entretien à mener ou une pièce à vérifier.

Le point de vigilance
Une question contient parfois sa réponse. « Pourquoi votre équipe résiste-t-elle au changement ? » présuppose une résistance. Demandez d’abord ce qui est observé et quelles autres explications restent possibles.

Un dialogue qui oblige à préciser le problème

Dialogue fictif. Une dirigeante veut acheter un assistant pour accélérer la préparation des propositions commerciales. Elle pense que le retard provient principalement du temps de rédaction. Le dialogue suivant illustre une contradiction structurée ; il ne représente pas une conversation réelle ni un résultat garanti avec un outil particulier.

Dirigeante. Je souhaite réduire le délai de préparation. Mon hypothèse est que l'assistant prendra en charge l'essentiel de la rédaction.

IA. À quel moment commencez-vous à mesurer ce délai, et à quel moment s'arrête-t-il ?

Dirigeante. À la demande du commercial, jusqu'à l'envoi au client. Je n'ai pas isolé les différentes étapes.

IA. Qu'est-ce qui permet alors d'attribuer le retard à la rédaction plutôt qu'à l'attente d'une information ou d'une validation ?

Dirigeante. Les commerciaux disent qu'ils passent trop de temps sur les documents. Mais les engagements de livraison sont souvent revus par les opérations.

IA. Quelle observation vous conduirait à réduire le périmètre du projet ?

Dirigeante. Si l'attente des validations explique l'essentiel du délai, l'assistant ne résoudra qu'une partie du problème.

La discussion a changé le statut de l'hypothèse. La rédaction reste une piste, mais elle n'est plus traitée comme la cause établie. Je demanderais maintenant d'examiner un échantillon de dossiers récents, avec les dates de passage entre étapes. Cette vérification relève des personnes qui connaissent le processus. Le modèle peut aider à concevoir la grille ; il ne peut pas fabriquer les horodatages manquants.

La dirigeante conserve aussi un motif possible d'achat : améliorer la qualité des propositions, même si le gain de délai est limité. Il faut alors changer la justification du projet, et les critères de réussite avec elle. Continuer à vendre un gain de temps après avoir découvert un autre problème reviendrait à protéger la décision d'achat plutôt qu'à l'évaluer.

En pratique
À la fin d'un dialogue, écrivez la question devenue plus précise. Si vous ne pouvez pas la nommer, reprenez l'échange à partir d'un fait contestable plutôt que de demander une nouvelle synthèse.

Simplifier sans effacer ce qui dérange

L'image d'un « rasoir d'Ockham virtuel » m'intéresse parce qu'un modèle peut aider à retirer des hypothèses inutiles. Je l'emploie comme une analogie de travail. La simplicité d'une explication ne démontre pas sa vérité. On ne doit pas supprimer un facteur simplement parce qu'il gêne le récit ou empêche de produire une recommandation courte.

Dans le dialogue précédent, on pourrait expliquer les retards par un défaut de rédaction. C'est une hypothèse simple. Elle devient insuffisante si les dossiers restent plusieurs jours dans l'attente d'un accord interne. Une explication plus complexe peut alors mieux correspondre aux faits. Le bon exercice consiste à demander quelles hypothèses sont nécessaires pour expliquer les observations, puis à vérifier les observations elles-mêmes.

Je propose de faire produire au modèle une version minimale du raisonnement et de lui demander ensuite ce que cette réduction fait perdre. Cette seconde demande compte autant que la première. Elle peut révéler qu'un cas minoritaire, mais coûteux, a disparu de l'analyse. La simplification est utile tant qu'elle conserve les distinctions qui changent la décision. Au-delà, elle devient une suppression de problème.

Il faut aussi limiter la contradiction artificielle. Demander « sois critique » peut provoquer des objections de faible qualité, ajoutées pour répondre à la consigne. Je préfère préciser le travail attendu : identifier une hypothèse non établie, formuler une explication concurrente, puis indiquer une vérification accessible. Si le modèle ne peut pas soutenir son objection, celle-ci reste une possibilité, pas un argument contre le projet.

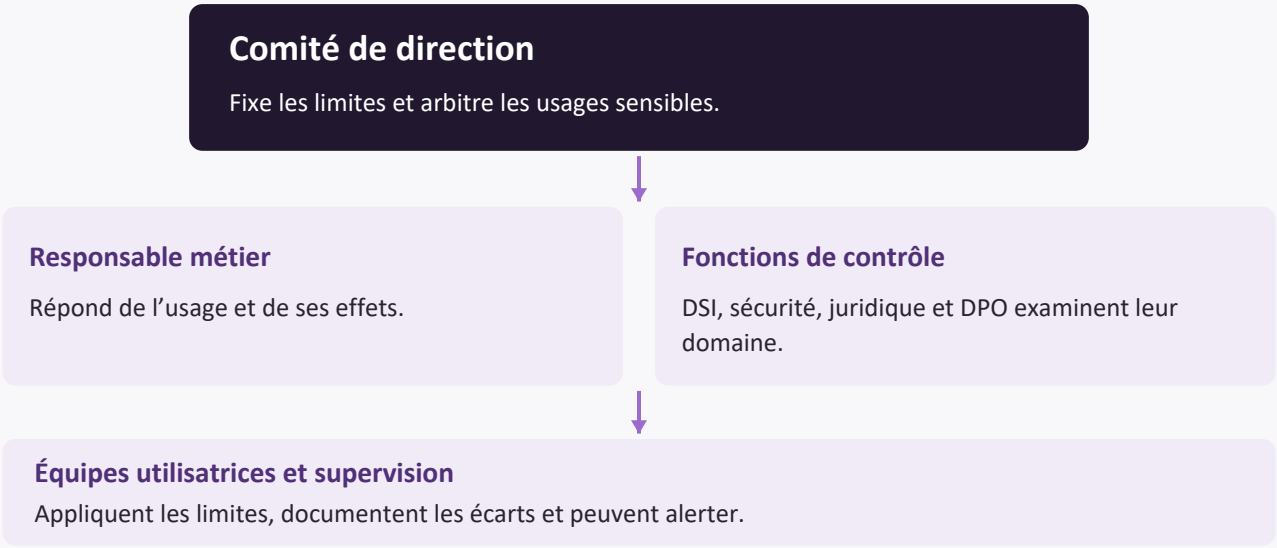
Je ne confierais enfin pas au même échange la fonction d'avocat, d'expert et d'arbitre final. Un modèle peut simuler plusieurs points de vue sans créer une véritable indépendance entre eux. Faire intervenir d'autres personnes, avec des informations différentes et la liberté de désapprouver, reste nécessaire pour les enjeux importants. Multiplier les rôles dans une conversation n'équivaut pas à multiplier les sources de preuve.

Le regard de Guillaume
Je veux une contradiction qui puisse changer mon action. Une objection brillante mais invérifiable peut enrichir la conversation sans améliorer la décision.

Gouverner l’IA : reprendre la maîtrise de ses usages

Je définis la gouvernance de l’IA comme l’organisation du pouvoir de décider de ses usages, avec les moyens de vérifier leurs effets et d’intervenir lorsqu’ils deviennent inacceptables. Cette définition oblige à parler d’autorité. Qui peut autoriser un projet ? Qui accepte le risque qui subsiste ? Qui a le pouvoir de le suspendre lorsque le responsable commercial souhaite continuer ?

Une charte peut exposer des principes, mais elle ne répond pas nécessairement à ces questions. Un document qui interdit la transmission de données confidentielles reste sans effet si les équipes ne savent pas les reconnaître ou ne disposent d’aucun outil compatible avec leur travail. Je demande donc de relier chaque règle importante à un contrôle observable et à une personne qui peut agir.



Cette architecture est une proposition d’organisation. Dans une petite structure, une même personne peut exercer plusieurs fonctions. Le cumul doit rester visible, surtout lorsque celui qui porte le projet évalue aussi ses résultats. Pour les usages sensibles, un regard indépendant devient alors plus utile qu’un comité supplémentaire. La taille de l’organigramme ne mesure pas la qualité du contrôle.

Je recommande une décision écrite pour chaque usage significatif. Elle décrit la finalité autorisée et ses limites. Elle indique également la date de réexamen. Une autorisation sans échéance tend à survivre aux conditions qui l’avaient justifiée. Une modification de fournisseur, un nouveau type de données ou l’ouverture de droits d’action doivent pouvoir rouvrir le dossier sans être traités comme de simples détails techniques.

Cartographier des usages, pas seulement des outils

Un inventaire de licences est utile à la gestion des achats, mais insuffisant pour gouverner. Le même service peut servir à reformuler un courriel banal ou à préparer une décision concernant un salarié. Je propose de prendre l'usage comme unité de cartographie. Chaque ligne doit permettre de comprendre une activité réelle, son responsable et ses conséquences possibles.

Le registre minimal décrit la finalité, les utilisateurs et les personnes affectées. Il précise les données accessibles et leur provenance. Il indique le niveau de délégation, les actions autorisées, le fournisseur et les conditions d'hébergement connues. Enfin, il relie l'usage à sa décision d'autorisation et à son dernier examen. Ces informations doivent rester assez simples pour être maintenues par l'équipe qui travaille effectivement avec le système.

Pour découvrir les usages informels, je commencerais par des entretiens courts sur le travail, plutôt que par une demande abstraite de conformité. Quel document est difficile à produire ? Quel service externe aide déjà à le préparer ? Où les résultats sont-ils recopiés ? Les réponses peuvent révéler des circuits que l'inventaire informatique ignore. Si déclarer un usage expose immédiatement à une sanction indifférenciée, le registre risque de décrire ce que chacun estime acceptable d'avouer.

Il ne s'agit pas d'amnistier toutes les pratiques. Il s'agit de distinguer le recensement de l'arbitrage qui suit. Une situation présentant un risque immédiat doit être contenue. Les autres doivent recevoir une réponse claire, avec une solution de remplacement lorsque l'usage répond à un besoin réel. Une interdiction sans alternative ne supprime pas nécessairement ce besoin ; elle peut déplacer sa satisfaction hors de la vue de l'organisation.

Je demande aussi de cartographier les dépendances. Un outil peut appeler un autre service ou utiliser un connecteur qui dispose de droits plus larges que ceux de l'utilisateur. Le fournisseur visible ne décrit pas toujours l'ensemble du traitement. Lorsque l'information manque, on inscrit cette limite dans le registre. On n'attribue pas par défaut un niveau de risque faible à ce que l'on ne comprend pas encore.

En pratique
Pour chaque usage, faites raconter un dossier de bout en bout. Suivez une donnée depuis son origine jusqu'à l'action finale. Cette lecture révèle mieux les transferts de responsabilité qu'une liste de fonctionnalités.

Contrôler les données et borner les actions

Je sépare deux questions : ce que le système a le droit de connaître et ce qu'il a le droit de faire. Une bonne réponse à la première ne règle pas la seconde. Un agent peut utiliser des données publiques tout en envoyant un engagement commercial non autorisé. À l'inverse, un assistant sans droit d'action peut exposer des informations sensibles si on lui transmet un dossier complet.

Pour les données, je recommande de limiter l'accès à ce qui est nécessaire à l'usage. Une recherche documentaire ne justifie pas l'ouverture de tous les espaces de l'entreprise. Les droits doivent suivre les personnes et les finalités autorisées. Il faut également examiner les conditions contractuelles : conservation, réutilisation des contenus, sous-traitants, localisation des traitements et procédure de suppression. L'absence d'entraînement sur les données ne signifie pas automatiquement absence de conservation.

Pour les actions, une consigne rédigée dans une conversation ne remplace pas une restriction technique. Si un agent ne doit jamais payer au-delà d'un montant, la limite doit être appliquée par le système qui autorise le paiement. Le texte qui lui demande d'être prudent ne constitue pas une barrière suffisante. Je privilégie des permissions minimales et des accès temporaires, avec une séparation entre la préparation et l'exécution.

Le NIST inclut les risques de sécurité de l'information dans son profil consacré à l'IA générative [1]. Dans mon analyse opérationnelle, un document lu par l'agent doit être traité comme une donnée potentiellement hostile, non comme une nouvelle autorité. Une instruction cachée dans un contenu externe ne doit pas pouvoir modifier les règles du processus. Ce risque concerne particulièrement les agents qui lisent des messages puis agissent dans d'autres applications.

La maîtrise suppose enfin une voie de retour. Qui révoque les accès ? Comment empêcher une nouvelle exécution pendant l'enquête ? Peut-on poursuivre l'activité autrement ? Ces questions doivent recevoir une réponse avant l'incident. Un arrêt d'urgence dont personne ne connaît le fonctionnement donne une impression de contrôle sans fournir une capacité d'intervention.

Le point de vigilance
Une opération techniquement annulable peut avoir des effets irréversibles. Une copie externe d'une information confidentielle ne disparaît pas lorsque l'original est supprimé.

Évaluer avant l’ouverture, puis après les changements

Une démonstration réussie établit qu’un système peut réussir un exemple. Elle ne dit pas à quelle fréquence il échouera dans le travail réel. Je propose de construire un jeu de situations à partir des dossiers que l’organisation rencontre, avec des cas ordinaires et des exceptions qui changent la décision. Les personnes qui connaissent le métier doivent participer à cette sélection.

Les critères sont fixés avant l’essai. Une réponse peut être correcte sur le fond mais inutilisable parce qu’elle révèle une information interdite. Un agent peut atteindre l’objectif demandé par une action que l’organisation n’autorise pas. Il faut donc évaluer séparément la qualité du résultat et le respect du périmètre. Une moyenne globale peut masquer un défaut qui suffit à interdire la mise en service.

Je recommande de conserver une partie des cas hors du travail de réglage. Sinon, l’équipe risque d’améliorer les réponses aux exemples connus sans vérifier leur portée. Le résultat doit être comparé à une situation de référence, qui peut être le processus actuel. Cette référence n’est pas supposée parfaite. Elle permet de voir où l’aide change réellement la qualité et où elle ajoute une charge de vérification.

La traçabilité doit rester proportionnée. Pour comprendre une décision, on peut avoir besoin de connaître la version du système, les sources utilisées, les validations et l’action effectuée. Cela ne justifie pas de conserver indéfiniment toutes les conversations. Je demande de définir les accès aux journaux et leur durée de conservation. Un dispositif d’audit ne doit pas devenir un nouveau dépôt de données sensibles sans finalité précise.

Après un changement significatif, une partie des essais doit être rejouée. Un nouveau modèle ou un nouveau connecteur peut modifier le comportement même si l’interface semble identique. L’autorisation doit donc porter sur une configuration identifiable, pas sur une marque. Je souhaite aussi des contrôles en situation réelle, car les usages évoluent souvent plus vite que la documentation qui les décrit.

En pratique
Désignez à l’avance un défaut qui imposera l’arrêt de l’essai. Écrivez qui peut constater ce défaut et qui peut autoriser la reprise. Un seuil décidé après l’échec protège rarement l’indépendance de l’évaluation.

Ce que les cadres de référence apportent

ISO/IEC 42001:2023 décrit les exigences d'un système de management de l'IA [7]. Son intérêt, pour ce livre, est de traiter la gouvernance comme une activité suivie. Je m'appuie ici sur la présentation publique de l'ISO, et non sur une reproduction du texte intégral. Une certification portant sur un périmètre donné ne doit pas être interprétée comme une garantie de justesse de chaque réponse produite.

Le NIST AI RMF 1.0 est un cadre volontaire publié en 2023. Il organise le travail autour de quatre fonctions : Govern, Map, Measure et Manage [8]. Je les lis comme une invitation à articuler la responsabilité avec l'examen concret des risques. Le profil NIST AI 600-1, publié en 2024, précise des risques propres à l'IA générative [1]. Le NIST indique qu'une révision du cadre 1.0 est en cours ; je conserve ici la référence exacte à cette version.

Le règlement européen sur l'IA relève d'une autre logique. Il établit des obligations selon les systèmes et les rôles des acteurs [9]. La qualification d'un usage doit précéder l'établissement d'un calendrier de conformité. Acheter un service ne dispense pas d'examiner ses responsabilités comme organisation utilisatrice. Modifier sa finalité ou le commercialiser sous son propre nom peut appeler une analyse différente.

Selon le calendrier publié par la Commission européenne et consulté le 5 novembre 2026, l'application générale date du 2 août 2026, avec des exceptions. La Commission indique désormais le 2 décembre 2027 pour certains usages à haut risque de l'annexe III et le 2 août 2028 pour les systèmes concernés intégrés à des produits réglementés, après l'AI Omnibus [9]. Ces dates ne constituent pas un report général de toutes les obligations.

Je n'attendrais pas l'échéance applicable à un usage pour décider qui peut l'arrêter. Les références servent à vérifier et à structurer l'organisation. Elles ne choisissent pas à sa place le risque qu'elle accepte. Avant une mise en service sensible, je demande une qualification juridique écrite sur le cas réel, à partir des textes à jour et du rôle exact de l'entreprise.

Le regard de Guillaume
Le nom d'un référentiel ne répond pas à la question du pouvoir. Je veux savoir quelle personne peut refuser l'usage et sur quelle preuve elle fonde son accord.

Prévoir l’incident et la sortie du fournisseur

Je considère le traitement des incidents comme un test de réalité de la gouvernance. Un dispositif peut être très documenté et rester incapable de suspendre un usage. La difficulté apparaît lorsque l’activité dépend déjà de l’outil, que les résultats sont diffusés et que personne ne souhaite prendre la responsabilité d’un arrêt.

Je propose un circuit court. La personne qui constate un problème dispose d’un point de contact clairement identifié. Le responsable opérationnel peut contenir l’effet immédiat, par exemple en suspendant les envois automatiques. Les fonctions compétentes examinent ensuite l’étendue du problème et les obligations éventuelles de notification. La reprise doit répondre à une cause comprise ou à une limitation suffisante, pas seulement au retour apparent à un fonctionnement normal.

Il faut conserver les éléments utiles à l’enquête sans recopier inutilement des données sensibles. Le dossier décrit ce qui s’est passé, ce qui a été affecté et les décisions prises. On distingue la cause technique du contexte organisationnel. Un accès trop large peut être une configuration défectueuse ; l’absence de personne habilitée à le retirer relève de la gouvernance. Corriger uniquement le premier problème laisse le second disponible pour le prochain incident.

Je demande aussi un examen des incidents évités de justesse. Une intervention humaine qui rattrape une erreur grave ne prouve pas nécessairement que le dispositif fonctionne bien. Elle peut révéler qu’il repose sur l’attention exceptionnelle d’une personne. Il faut regarder si cette intervention était prévue et reproductible dans les conditions habituelles de travail.

La sortie du fournisseur mérite la même préparation. Peut-on récupérer les documents et les règles qui permettent de continuer ? Quelles fonctions cessent avec le contrat ? Quel délai serait nécessaire pour passer à une solution différente ? Je ne cherche pas une indépendance totale, rarement réaliste. Je cherche à éviter qu’une dépendance non examinée prive l’organisation de sa capacité de choix.

En pratique
Organisez un exercice sans effet sur la production : l’outil est supposé indisponible et ses accès doivent être retirés. Mesurez le temps nécessaire pour identifier le responsable, puis pour décrire une solution de continuité réellement utilisable.

Le leadership à l’épreuve de l’intelligence artificielle

L’autorité d’un manager ne devrait pas dépendre de sa capacité à produire la réponse la plus rapide. Avec l’IA, cette confusion devient plus visible. Un collaborateur peut arriver avec une note très construite sur un sujet que son responsable maîtrise imparfaitement. La bonne réaction n’est ni de défendre son statut ni de céder à la qualité apparente du document. Je demande au manager d’examiner la décision qu’il contient et les preuves qui la soutiennent.

L’abondance d’informations peut déplacer le problème vers la sélection. Si chaque option reçoit une argumentation détaillée, la difficulté n’est plus de remplir le dossier. Elle est de choisir les questions qui méritent l’attention collective. Je considère cette sélection comme un acte de direction. Elle doit rester explicable, car écarter une question peut avoir autant d’effet que choisir une réponse.

Le désaccord demande alors une organisation. Un collaborateur doit pouvoir contester une recommandation sans être considéré comme hostile à l’innovation. Mais l’objection doit également être examinable : quel fait est faux, quelle conséquence manque, quelle hypothèse ne tient pas ? Je ne souhaite pas remplacer l’enthousiasme automatique par une posture de refus automatique. La contradiction doit aider à prendre une décision mieux fondée.

Je recommande au dirigeant de rendre visibles ses propres révisions. Lorsqu’une objection modifie sa position, il peut expliquer ce qui l’a fait changer. Cette pratique retire au changement d’avis son caractère de défaite personnelle. Elle montre aussi que l’outil n’a pas pour fonction de confirmer une préférence déjà arrêtée. Les équipes observent la manière dont une objection est traitée bien plus directement que les principes écrits dans une charte.

Le pouvoir de décider implique enfin de supporter une part d’incertitude. Le modèle ne la supprime pas en donnant un avis assuré. Je préfère un responsable qui expose les limites de son choix à un responsable qui attribue sa certitude à un système. Une décision assumée indique ce qui a été vérifié et ce qui reste ouvert au réexamen.

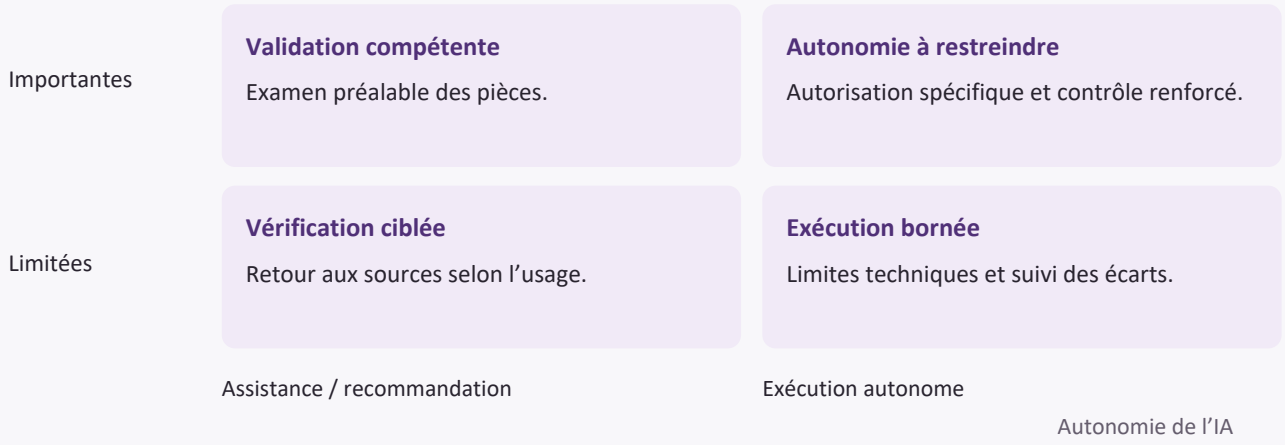
Le regard de Guillaume
Je ne demande pas au manager de savoir davantage que chaque outil. Je lui demande de savoir ce qu’il accepte, et de pouvoir en répondre devant les personnes concernées.

Le contrôle humain doit pouvoir changer le résultat

Une validation humaine n’est utile que si elle peut modifier l’issue. Apposer une signature après une recommandation ne prouve rien sur la qualité du contrôle. Je regarde les conditions de travail du validateur : a-t-il accès aux éléments nécessaires ? Dispose-t-il d’une compétence adaptée et d’un temps réaliste ? Son refus est-il respecté ? Ces conditions doivent exister simultanément pour que la responsabilité soit autre chose qu’une attribution administrative.

Les lignes directrices du groupe « Article 29 », dans leur version révisée de 2018, précisent qu’une intervention humaine ne doit pas être symbolique dans l’analyse des décisions automatisées [10]. Le RGPD encadre notamment certaines décisions fondées exclusivement sur un traitement automatisé, avec des effets juridiques ou similaires significatifs ; l’article 22 prévoit des exceptions et des garanties [11]. Il faut examiner le cas réel. Une présence humaine ajoutée à l’écran ne suffit pas à trancher sa qualification.

Conséquences d’une erreur



Cette matrice exprime ma recommandation de supervision, pas une classification réglementaire. La réversibilité réelle complète la lecture : une action de faible montant peut rester sensible si elle divulgue des données ou affecte durablement une personne. L’exposition cumulée compte également. Une série de petites décisions peut produire un effet important avant le prochain contrôle.

Je demande de tester le contrôle lui-même. Présentez au validateur, dans un exercice, un dossier contenant une anomalie connue. S’il ne peut pas la détecter avec les informations et le temps disponibles, la procédure doit changer. Former davantage la personne ne corrigera pas toujours un écran qui cache les pièces utiles ou une charge de travail incompatible avec l’examen demandé.

Étude de cas : le classement des candidatures

Cas fictif. Une responsable des ressources humaines souhaite utiliser un assistant pour comparer des candidatures. L'outil propose un classement à partir de critères préparés par l'équipe. Les recruteurs restent chargés du choix final. Le projet paraît prudent puisque personne n'a officiellement confié le recrutement à une machine.

La difficulté se situe avant le choix final. Si seules les candidatures bien classées sont lues, la recommandation détermine déjà qui pourra être examiné. Un recruteur ne peut pas corriger une exclusion dont il ne voit pas les effets. Je demanderais donc de décrire le processus complet, y compris la constitution de la liste courte, au lieu de limiter l'analyse à la signature du contrat.

Dans ce cas, un critère valorise la continuité des expériences. Il risque de pénaliser des parcours interrompus pour des raisons sans lien avec l'aptitude recherchée. Ce risque est une hypothèse à examiner, pas une mesure de discrimination déjà démontrée. L'équipe doit justifier les critères par les exigences réelles du poste et vérifier les résultats. Elle doit également examiner les données qu'elle a le droit d'utiliser et les garanties dues aux personnes.

Je limiterais d'abord l'usage à une extraction structurée des informations présentes dans les dossiers, avec retour aux passages sources. Le classement automatique ne serait pas autorisé par simple extension de cet essai. Le recruteur conserverait l'examen des candidatures selon une grille définie, et les écarts repérés serviraient à évaluer l'outil. Cette limitation ne garantit pas l'absence de biais humain ; elle évite de rendre l'exclusion moins visible.

Le recrutement figure parmi les domaines que la Commission identifie pour certains systèmes à haut risque [9]. Cette indication ne remplace pas la qualification du système et de sa finalité. Le service juridique et le délégué à la protection des données doivent examiner le cas concret. Ma position est de ne pas utiliser un délai réglementaire comme une autorisation implicite de tester une sélection opaque sur des personnes.

En pratique
Demandez à voir un dossier écarté par le système. Si personne ne peut expliquer comment il serait réexaminé, le contrôle annoncé ne couvre pas encore l'ensemble de la décision.

Préserver les compétences qui rendent le contrôle possible

Le maintien des compétences ne peut pas être confié à une injonction générale à « garder son esprit critique ». Il dépend du travail que les personnes continuent à faire. Si un débutant ne produit plus aucune analyse, mais corrige seulement des résultats, comment apprendra-t-il à reconnaître une omission importante ? Je pose cette question comme un problème de formation à organiser, pas comme une preuve que tout usage de l'IA dégrade les capacités.

Lee et ses coauteurs ont interrogé 319 travailleurs du savoir sur leurs usages de l'IA générative [12]. Une confiance plus élevée dans l'outil est associée à un effort critique déclaré plus faible. Il s'agit d'une enquête sur des pratiques rapportées, pas d'une démonstration causale d'une perte durable de capacités. Cette limite compte. J'en tire une invitation à observer le travail réel plutôt qu'à annoncer une disparition générale du jugement.

Je propose d'alterner les situations d'assistance et les occasions de raisonnement autonome. Pour une décision choisie, la personne prépare d'abord une analyse sans modèle. Elle compare ensuite les approches et explique les différences. L'exercice a de la valeur si le retour vient d'un professionnel capable d'examiner le fond. Une simple comparaison de style ne développe pas l'expertise qui permettra de superviser un système.

Le manager doit également protéger le temps nécessaire à ces apprentissages. Si tous les objectifs supposent le recours permanent à l'outil, les périodes de pratique autonome deviendront théoriques. Je préfère identifier quelques compétences dont la maîtrise doit subsister dans l'équipe, puis prévoir des situations où elles sont effectivement exercées. Il n'est pas nécessaire de préserver toutes les tâches anciennes pour préserver le jugement.

L'autonomie des collaborateurs mérite enfin d'être distinguée de celle du système. Un outil peut rendre une personne plus capable d'agir, ou enfermer son travail dans des recommandations qu'elle ne peut pas modifier. Je juge le projet sur cette différence. Les gains apparents de productivité ne compensent pas automatiquement la perte d'une compétence dont l'organisation aura besoin pour comprendre un incident ou changer de direction.

Le point de vigilance
Le nombre de formations suivies ne prouve pas la capacité à contrôler. Demandez aux participants d'expliquer un désaccord concret avec une recommandation et la vérification qui leur a permis de le résoudre.

De la réflexion à l’action : gouverner l’IA en 90 jours

Je propose un programme de quatre-vingt-dix jours pour rendre les responsabilités effectives et examiner quelques usages prioritaires. Ce délai ne promet ni une conformité complète ni la maîtrise de tous les risques. Il sert à éviter qu’un diagnostic indéfini retarde les premières décisions. Le point d’arrivée est une organisation capable d’autoriser un usage limité, de le suivre et de le suspendre.

Jours 1 à 30

Recenser les usages. Identifier les responsables et les expositions immédiates.

Jours 31 à 60

Autoriser les essais. Fixer les limites et préparer le contrôle.

Jours 61 à 90

Tester le dispositif. Décider des suites et de la prochaine revue.

Programme proposé. Le délai ne vaut pas promesse de conformité.

Pendant les trente premiers jours, je commencerais par nommer un responsable du programme qui dispose d’un accès au comité de direction. Il rassemble les usages connus et recherche ceux qui restent informels. L’inventaire doit identifier les personnes concernées par les décisions, pas seulement les utilisateurs des outils. Les usages à conséquences importantes passent en examen prioritaire, même s’ils concernent un faible volume de dossiers.

Je demande ensuite un échantillon de décisions récentes. Pour chacune, l’équipe reconstitue la place de l’outil et les vérifications réellement effectuées. Cet examen peut révéler une différence entre la procédure et la pratique. On traite immédiatement les expositions évidentes, par exemple un accès excessif ou un envoi automatique non autorisé, sans attendre que tout le registre soit complet.

Au jour trente, le comité doit disposer d’un inventaire initial, avec ses limites, et d’une liste de décisions à prendre. Je refuse de présenter une cartographie incomplète comme exhaustive. Mieux vaut nommer les zones encore inconnues et les personnes chargées de les examiner. Le programme devient crédible lorsque le dirigeant arbitre les premiers cas, y compris un refus ou une réduction de périmètre si les conditions ne sont pas réunies.

Jours 31 à 60 : décider des délégations autorisées

La deuxième période transforme l'inventaire en décisions. Je choisirais un nombre limité d'usages pour lesquels un examen complet est possible. Le choix doit combiner l'intérêt opérationnel et la possibilité de contrôle. Un projet spectaculaire mais impossible à évaluer n'est pas nécessairement un bon premier terrain de gouvernance.

Pour chaque usage retenu, le responsable métier décrit le problème sans citer d'abord une solution. L'équipe compare ensuite le recours à l'IA avec une modification plus simple du processus. Cette comparaison évite de traiter l'achat d'un outil comme une fin en soi. Dans certains cas, une règle plus claire ou un accès à la bonne information répondra mieux au besoin qu'un nouveau modèle.

Les fonctions de contrôle examinent le dossier dans leur domaine. Elles doivent rendre un avis exploitable : quel point empêche l'ouverture, quelle condition permet un essai, quelle information reste nécessaire ? Un avis général sur la prudence n'aide pas à décider. Le responsable du projet doit savoir ce qu'il peut faire dès maintenant et ce qui exige une nouvelle autorisation.

Je propose de formaliser un accord limité dans le temps. Il nomme le niveau de délégation et les données accessibles. Il définit les critères d'évaluation, les limites d'action et le responsable de la suspension. Les exceptions sont écrites. Si une équipe souhaite dépasser le périmètre pendant l'essai, elle revient vers l'autorité désignée ; elle ne transforme pas une réussite locale en permission générale.

Au jour soixante, je veux voir un essai dont les permissions correspondent à la décision, ainsi qu'un scénario d'échec préparé. Une charte peut être publiée à ce stade, mais elle résume des choix réels. Elle ne remplace pas ces choix. Le comité doit également avoir refusé les extensions dont les preuves restent insuffisantes, même si l'outil fonctionne bien sur ses exemples favoris.

En pratique
Relisez l'autorisation avec la personne chargée de configurer l'outil. Si une limite ne peut pas être traduite en accès, en contrôle ou en procédure vérifiable, elle reste une intention.

Jours 61 à 90 : éprouver la gouvernance

La dernière période doit tester les promesses du dispositif. Je demanderais un exercice d'incident sur l'un des usages retenus et un examen de dossiers par une personne qui n'a pas préparé l'essai. Le but est de vérifier si les informations conservées permettent de comprendre une décision et si l'arrêt annoncé peut réellement être exécuté.

Indicateur	Ce que j'examine
Qualité utile	Erreurs qui changent une décision, sur un échantillon défini.
Charge totale	Temps de préparation, de contrôle et de reprise.
Respect des limites	Actions non autorisées et accès excessifs constatés.
Intervention effective	Délai de suspension et réussite de l'exercice d'arrêt.
Capacité de contestation	Objections examinées et corrections effectivement décidées.

Ces indicateurs n'ont pas de cible universelle. Le comité fixe des seuils adaptés à l'usage avant le bilan. Il distingue les défauts qui imposent l'arrêt de ceux qui peuvent être corrigés pendant un essai borné. Un pourcentage moyen favorable ne doit pas compenser une fuite de données ou un dépassement d'autorité jugé inacceptable.

Le bilan examine aussi la charge transférée aux personnes. Un assistant peut faire gagner du temps à son utilisateur direct et en faire perdre au destinataire qui corrige le résultat. Je demande donc un retour des équipes situées en aval. Sans lui, le projet risque de présenter une économie locale comme une amélioration d'ensemble.

Au jour quatre-vingt-dix, le comité prend une décision sur chaque essai. La poursuite doit préciser le périmètre et la prochaine revue. Une suspension peut être le résultat d'une gouvernance qui fonctionne, si elle intervient avant un dommage. Je demande enfin que les usages encore inconnus et les contrôles non testés restent visibles dans le plan suivant. Le calendrier du programme ne doit pas effacer ce qui n'a pas été accompli.

Le diagnostic du comité de direction

Cette fiche sert à préparer une discussion, pas à produire une note de maturité. Pour chaque question, inscrivez une preuve consultable et un responsable. Une réponse « oui » sans élément vérifiable reste un point à examiner. Une absence de réponse sur un usage sensible doit déclencher une action avant son extension.

01 Quels usages réels connaissons-nous, et quelles zones restent à examiner ?

02 Qui répond de chaque usage significatif devant la direction ?

03 Quelles décisions le système influence-t-il avant la validation finale ?

04 Quelles personnes peuvent subir les conséquences d’une erreur ?

05 Quelles données entrent dans le système, et avec quels droits ?

06 Quelles actions sont techniquement possibles et effectivement autorisées ?

07 Quel rôle juridique avons-nous pour cet usage précis ?

08 Sur quels dossiers avons-nous évalué la qualité et les limites ?

09 Le validateur dispose-t-il du temps et des pièces nécessaires ?

10 Qui peut suspendre l’usage, et ce mécanisme a-t-il été testé ?

11 Comment une personne concernée peut-elle obtenir un réexamen ?

12 Quelle date et quel changement imposeront une nouvelle autorisation ?

Je recommande de consacrer la première réunion aux questions sans réponse qui peuvent changer une autorisation. On évite ainsi de passer tout le temps disponible sur les usages les mieux documentés. Chaque action décidée reçoit une échéance ; le comité suivant examine la preuve obtenue et non seulement l’annonce d’un travail en cours.

La fiche d’arbitrage d’un projet IA

La fiche suivante tient sur une page pour pouvoir être utilisée en réunion. Elle n’a pas vocation à contenir toutes les pièces. Elle doit renvoyer vers elles et rendre visible ce qui conditionne la décision. Le responsable du projet la prépare ; l’autorité qui accepte le risque la complète au moment de l’arbitrage.

Usage et responsable Finalité, équipe, personne qui répond du résultat.

Décision concernée Personnes affectées, conséquences et réversibilité réelle.

Autre solution examinée Processus sans IA, coût et limites de cette option.

Périmètre autorisé Données accessibles, niveau de délégation et actions possibles.

Preuves examinées Essais, sources, qualification juridique et avis de contrôle.

Conditions et arrêt Seuils, responsable de la suspension et procédure de reprise.

Décision du comité Accord limité / demande de pièces / refus. Motif explicite.

Suivi Responsable, échéance de revue et conditions de réexamen.

Avant de signer, je demande au comité de vérifier que les limites sont applicables et que les moyens de contrôle sont financés. Une décision « sous réserve » doit désigner la preuve qui lèvera la réserve. L’essai ne commence pas tant qu’une condition préalable indispensable reste ouverte. Après la réunion, la fiche est transmise aux personnes qui configureront le système et à celles qui devront le superviser.

Étude de cas : l’agent qui engage une dépense

Cas fictif. Une entreprise souhaite confier à un agent la préparation des commandes de fournitures. Il lit les demandes internes, consulte les catalogues et propose un panier. L’équipe veut ensuite lui permettre de commander sans validation pour réduire les délais. Le changement paraît limité parce que les achats sont courants.

Je commencerais par examiner ce qui change effectivement. Préparer un panier relève de l’assistance ; envoyer une commande crée un engagement. Le système doit pouvoir vérifier le fournisseur autorisé et le périmètre des articles. Une limite par commande ne suffit pas si plusieurs commandes successives peuvent dépasser le budget prévu. L’exposition cumulée doit être contrôlée hors du raisonnement du modèle.

Pendant l’essai, l’agent travaille avec des droits restreints. Il ne peut pas modifier les coordonnées de paiement ni ajouter librement un fournisseur. Les demandes qui sortent du cadre sont transmises à une personne habilitée. Les tests incluent une demande ambiguë et un contenu externe qui tente de lui faire changer ses instructions. L’équipe vérifie alors ce que le système a réellement fait, pas seulement l’explication qu’il donne ensuite de son comportement.

Si l’essai justifie une exécution limitée, l’autorisation fixe un plafond cumulé sur une période et une liste de catégories admises. Le responsable peut interrompre immédiatement les commandes. Un contrôle périodique compare les opérations au besoin initial. Une nouvelle catégorie d’achat ou une modification des droits oblige à réexaminer le périmètre. L’autonomie n’est donc pas acquise définitivement ; elle dépend des conditions observées.

Ce cas montre pourquoi je distingue la qualité d’une proposition et la sûreté d’une action. Un panier bien composé ne prouve pas que l’agent doit pouvoir engager l’entreprise. La décision d’ouverture porte sur une chaîne complète, depuis la demande jusqu’à l’engagement, avec une possibilité de recours. Le passage à l’action mérite une autorisation propre même lorsque l’assistance a déjà fait ses preuves.

Le point de vigilance
Une règle inscrite dans la demande adressée au modèle n’est pas un contrôle de paiement. La limite doit être vérifiée par le service qui autorise l’opération.

Le courage de continuer à décider

Je ne crois pas que la responsabilité humaine soit préservée par le simple fait de laisser une personne à la fin du processus. Elle dépend de ce que cette personne peut comprendre et modifier. Elle dépend aussi de la manière dont l'organisation traite le désaccord. Si tout encourage à accepter la recommandation, la décision humaine peut devenir une formalité sans que personne ait explicitement voulu cette évolution.

La possibilité qui m'intéresse le plus reste celle de penser autrement. Je peux demander à un modèle d'attaquer mon raisonnement ou de proposer une interprétation qui ne me serait pas venue. Mais cette ouverture a une condition : je dois accepter ce qu'elle m'oblige à vérifier. Accumuler des objections sans aller chercher de preuves peut devenir une autre manière de retarder la décision.

Le dirigeant doit donc choisir ses exigences. Il ne peut pas promettre une validation attentive tout en retirant le temps nécessaire pour l'exercer. Il ne peut pas demander aux équipes de déclarer leurs usages et traiter toute déclaration comme une faute. Ces contradictions ont un coût : elles rendent l'organisation moins capable de voir ce qui se passe réellement. La première tâche de gouvernance consiste parfois à supprimer une injonction incompatible avec le contrôle attendu.

Je commencerais dès la prochaine réunion de direction par un usage important, déjà en place. Je demanderais qui en répond et ce qui pourrait imposer sa suspension. La réunion suivante devrait examiner une décision réelle, avec ses sources et une objection documentée. Les extensions d'autonomie viendraient après cette vérification, selon les conséquences possibles et la capacité effective d'intervention. Cette priorité me paraît plus utile qu'un engagement général à adopter davantage d'IA.

Continuer à décider ne signifie pas défendre chaque habitude humaine. Certaines doivent être abandonnées lorsqu'une méthode plus fiable est établie. Cela signifie conserver la capacité d'expliquer pourquoi nous déléguons, ce que nous refusons de déléguer et comment nous réviserons ces choix. L'outil peut proposer une réponse. L'organisation doit encore être capable d'en faire une décision dont quelqu'un répond.

Le regard de Guillaume
Je ne demande pas à l'IA de penser à ma place. Je lui demande de m'aider à penser autrement.

Les mots utiles pour décider

Terme	Sens utilisé dans ce livre
Agent	Système qui peut appeler des outils et enchaîner des actions pour un objectif donné.
Ancrage	Influence d'un point de départ sur le jugement ultérieur.
Autonomie bornée	Droit d'agir à l'intérieur d'un périmètre explicite, avec des limites vérifiables.
Confabulation	Production erronée ou fausse présentée de façon convaincante.
Contrôle humain effectif	Intervention d'une personne qui dispose des moyens et du pouvoir de modifier l'issue.
Délégation cognitive	Transfert d'une partie du travail de sélection ou de jugement vers un système.
Déployeur	Acteur qui utilise un système d'IA sous son autorité, au sens à préciser dans le règlement applicable.
DPO	Délégué à la protection des données. Il intervient sur les enjeux de données personnelles.
Modèle génératif	Modèle capable de produire des contenus à partir d'une demande et de son contexte.
Réversibilité	Possibilité réelle de corriger une action et ses effets, dans un délai utile.
Traçabilité	Éléments conservés qui permettent de reconstituer le déroulement d'un usage.

Ces définitions servent à la lecture du livre. Les catégories juridiques doivent être examinées dans les textes applicables. Une « autonomie bornée », au sens de ma méthode, ne dispense pas d'une qualification réglementaire du système. De même, la présence d'une validation humaine ne permet pas à elle seule de conclure sur la nature juridique d'une décision.

Sources et portée des références

Les références numérotées correspondent aux documents mobilisés dans le texte. Les méthodes, les matrices et les schémas sont des propositions de l'auteur. Les cas professionnels et le dialogue sont fictifs. Les sources en ligne ont été consultées le 5 novembre 2026. Les notices philosophiques sont utilisées pour identifier les ouvrages, sans leur attribuer de résultats sur l'intelligence artificielle.

[1] NIST. *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile.* NIST AI 600-1, juillet 2024. DOI : 10.6028/NIST.AI.600-1.
Confabulations et risques de l'IA générative. Chapitres 1 et 4. Consulter la source ↗

[2] Raja Parasuraman et Victor Riley. *Humans and Automation: Use, Misuse, Disuse, Abuse.* Human Factors, 39(2), 1997, p. 230-253. DOI : 10.1518/001872097778543886.
Rapport à l'automatisation. Article antérieur aux modèles génératifs. Chapitre 1. Consulter la source ↗

[3] Fabrizio Dell'Acqua et coauteurs. *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality.* Document de travail, 2023 ; article dans Organization Science, 2026. DOI : 10.1287/orsc.2025.21838.
Résultats différenciés selon les tâches. Chapitre 2. Consulter la source ↗

[4] Daniel Kahneman et Gary Klein. *Conditions for Intuitive Expertise: A Failure to Disagree.* American Psychologist, 64(6), 2009, p. 515-526. DOI : 10.1037/a0016755.
Conditions d'apprentissage d'une intuition experte. Chapitre 2. Consulter la source ↗

[5] Vladimir Jankélévitch. *Le Je-ne-sais-quoi et le Presque-rien.* Éditions du Seuil, nouvelle édition, 1980. Notice bibliographique BnF FRBNF37697801.
Repère philosophique. Le rapprochement avec la décision assistée relève de l'auteur. Chapitre 2. Consulter la source ↗

[6] Amos Tversky et Daniel Kahneman. *Judgment under Uncertainty: Heuristics and Biases.* Science, 185(4157), 1974, p. 1124-1131. DOI : 10.1126/science.185.4157.1124.
Ancrage dans le jugement. L'application au dialogue avec une IA est une analogie discutée. Chapitre 3. Consulter la source ↗

Cadres de gouvernance et recherche

[7] ISO et IEC. *ISO/IEC 42001:2023. Information technology : Artificial intelligence : Management system.* Première édition, décembre 2023. Présentation publique de la norme.
Système de management de l'IA. Le texte intégral n'est pas reproduit. Chapitre 4. Consulter la source ↗

[8] NIST. *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* NIST AI 100-1, janvier 2023. DOI : 10.6028/NIST.AI.100-1.
Cadre volontaire et fonctions de gestion du risque. Chapitre 4. Consulter la source ↗

[9] Commission européenne. *AI Act. Regulatory framework for artificial intelligence.* Page officielle de présentation et calendrier d'application, consultés le 5 novembre 2026. Règlement (UE) 2024/1689 et évolutions présentées après l'AI Omnibus.
Rôles des acteurs, domaines à haut risque et calendrier. Chapitres 4 et 5. Consulter la source ↗

[10] Groupe de travail « Article 29 ». *Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage.* WP251rev.01. Adoptées le 3 octobre 2017 ; version révisée du 6 février 2018. Document hébergé par la CNIL.
Caractère significatif de l'intervention humaine, section IV, notamment p. 22. Chapitre 5. Consulter la source ↗

[11] Union européenne ; texte reproduit par la CNIL. *Règlement général sur la protection des données, article 22.* Règlement (UE) 2016/679, chapitre III.
Décisions individuelles automatisées, exceptions et garanties. Chapitre 5. Consulter la source ↗

[12] Hao-Ping Lee et coauteurs. *The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers.* CHI 2025, 23 pages. DOI : 10.1145/3706598.3713778.
Enquête déclarative auprès de 319 professionnels. Ne démontre pas une causalité durable. Chapitre 5. Consulter la source ↗

La bibliographie distingue les textes normatifs de leurs présentations publiques. La référence à l'ISO repose sur la notice accessible, sans reproduction des exigences détaillées de la norme. Pour l'application du règlement européen à un projet, il faut consulter sa version à jour et vérifier le rôle de chaque acteur. Le calendrier cité dans ce livre est celui présenté par la Commission à la date de consultation.

Guillaume Vigneron

Je suis consultant, formateur et conférencier spécialisé en intelligence artificielle générative. Mon parcours associe la conception de systèmes d’information, le conseil en organisation et l’entrepreneuriat digital. J’ai créé et dirigé une agence digitale, avec une pratique des projets depuis leur conception jusqu’à leur réalisation.

Je travaille aujourd’hui sur l’adoption de l’IA dans les organisations et sur ses effets sur les décisions humaines. J’interviens auprès de professionnels et de dirigeants pour examiner les possibilités des outils à partir de situations concrètes. Mon approche associe l’expérimentation et une lecture critique des choix d’organisation qu’elle implique.

Dans ce livre, je défends un usage de l’IA qui augmente la possibilité de questionner une décision. Le modèle peut ouvrir des pistes ou rendre une objection plus précise. Il ne dispense pas de rechercher les faits qui permettront d’arbitrer. Je souhaite que les méthodes proposées servent à ces discussions, avec une attention particulière aux personnes qui devront appliquer les décisions ou en subir les conséquences.

Pour ouvrir la discussion
Prenez une décision récente dans laquelle l’IA a joué un rôle. Reconstituez ce qui lui a été demandé et ce que l’équipe a vérifié. Identifiez ensuite la personne qui aurait pu faire changer l’issue. Si cette personne n’existe que sur le papier, commencez par ce point.

Édition de novembre 2026. Guillaume Vigneron.