MURPHY CONSEIL

Manuel de prompting à destination des professionnels du droit

Intelligence artificielle et pratique juridique

Tudual Lucas Huon

Mars 2026

murphyconseil.com

À la mémoire de Mona Heydari

Remerciements

Merci à ma belle-mère, Pascale Keraudran, pour son soutien sans faille au cours de ces dernières années, contre vents et marées.

Merci à Maître Esther Laudy, avocate au barreau de Saint-Malo, et Maître Nina Gourvennec, avocate au barreau de Rennes, dont la pratique du métier d'avocat m'inspire au quotidien et me donne toujours de nouvelles idées, et sans qui cet ouvrage n'aurait jamais vu le jour.

Merci à Valentin Thomas, juriste au Crédit Mutuel Arkéa, dont le soutien et l'énergie permanente me poussent au quotidien à me dépasser.

Merci à Marie Le Vivier, juriste RGPD à Ouest-France, pour sa présence quotidienne et ses très précieux conseils.

Merci à Abir Adam, dont l'ambition me challenge au quotidien.

À propos de l'auteur

Tudual Lucas Huon

Ingénieur en IA· Juriste · Fondateur de Murphy Conseil

Je conçois des architectures d'intelligence artificielle et je mène des recherches sur le problème de l'alignement. Comment garantir qu'un modèle de langage restitue fidèlement l'intention de son utilisateur, sans déformation, omission ni extrapolation ? C'est l'un des défis centraux de l'IA contemporaine. Chacun de mes projets est conçu comme un terrain d'investigation sur le fonctionnement réel de ces systèmes.

Mes travaux portent notamment sur les architectures de mémoire persistante pour les grands modèles de langage ainsi que sur les risques liés à l'interaction prolongée avec ces modèles. Mon article *User Imprint: Psychological Profiling and Qualified Information in Prolonged Interaction with Large Language Models*, publié sur SSRN en mars 2026, formalise le concept d'*empreinte utilisateur* : la capacité d'un LLM à agréger, au fil des échanges, un profil psychologique exploitable de son utilisateur. Il y introduit la notion d'*information qualifiée*, prolongeant le cadre théorique du capitalisme de surveillance de Shoshana Zuboff.

Mes recherches me conduisent aux quatre coins du monde. Actuellement à New York, je m'installerai en septembre en Corée du Sud pour une durée minimale de trois ans, afin d'y étudier l'influence de l'intelligence artificielle sur la doctrine juridique dans une perspective de regards croisés entre le droit français et le droit sud-coréen.

Juriste de formation, titulaire d'un Master de l'Université de Brest en justice, procès et procédure, d'un Diplôme universitaire de l'Université Paris 1 Panthéon-Sorbonne en études des pratiques judiciaires, et du CRFPA (spécialité droit pénal), admissible au concours de la magistrature, j'ai exercé près de deux ans au pôle criminel du parquet de Rennes, où j'ai rédigé des réquisitoires, des mémoires d'appel et des synthèses juridiques aux côtés des magistrats du ministère public.

C'est au croisement de ces deux parcours que j'ai fondé Murphy Conseil, cabinet de conseil en stratégie et intelligence artificielle à destination des professionnels du droit. La première édition de ce guide a connu un véritable succès sur LinkedIn, avec plus de 300 téléchargements en moins d'une semaine et une portée de près de 20 000 professionnels, dont 49 % de juristes issus du ministère de la Justice, de l'École nationale de la magistrature et du Barreau de Paris.

L'activité du cabinet repose sur trois piliers :

Sobriété. Proposer uniquement ce qui est utile. Pas de solution surdimensionnée, pas de technologie pour la technologie. Chaque intervention est calibrée sur le besoin réel du professionnel.

Maîtrise. Garantir que le professionnel comprenne ce qu'il utilise. Aucune boîte noire, aucune dépendance : l'objectif est que chaque client soit en mesure de fonctionner de manière autonome à l'issue de l'accompagnement.

Transfert. Transmettre les compétences plutôt que de les retenir. Murphy Conseil ne crée pas de dépendance : il forme, il outille, puis il s'efface.

Ce guide est l'incarnation de cette philosophie. Il a été conçu pour donner aux professionnels du droit les clés d'une maîtrise réelle de l'intelligence artificielle, sans intermédiaire et sans filtre.

Table des matières

Introduction

Cela fait maintenant près de quatre ans que j'utilise quotidiennement les intelligences artificielles génératives. Juriste de formation, j'ai progressivement fait de ces outils un axe central de ma pratique professionnelle, développant au fil des expérimentations une expertise en ingénierie de prompt (*prompt engineering*). J'ai volontairement poussé l'usage aussi loin que possible — y compris dans ses retranchements les plus contre-intuitifs — afin de cartographier avec précision les capacités de ces modèles, mais aussi leurs limites.

Ce guide s'adresse aux professionnels du droit : magistrats, auditeurs de justice, avocats et élèves-avocats. Il répond à un quadruple constat.

Le premier est d'ordre **professionnel**. Les solutions d'IA juridique proposées par les éditeurs spécialisés se multiplient, parfois à des tarifs prohibitifs, sans que leurs utilisateurs disposent des clés de compréhension nécessaires pour en évaluer la fiabilité. Or, avant de déléguer une partie de son raisonnement à une machine, il est indispensable de comprendre comment elle fonctionne, où elle excelle, et surtout où elle défaille. Aux États-Unis, dans l'affaire *Mata v. Avianca, Inc.* (2023), des avocats ont été sanctionnés après avoir plaidé sur la base de précédents jurisprudentiels intégralement fabriqués par ChatGPT : décisions inventées, numéros de référence fictifs, formulations imitant le style de la juridiction concernée[1]. Ce cas n'est pas un accident : c'est la conséquence prévisible d'un outil puissant utilisé sans compréhension de ses mécanismes.

Le second est d'ordre **intellectuel**. Plusieurs études alertent sur le risque qu'un usage prolongé et irréfléchi de l'intelligence artificielle érode progressivement la capacité de réflexion autonome de ses utilisateurs. Un point sur lequel je reviendrai à plusieurs reprises dans ce guide. Ce danger est particulièrement sérieux dans les professions juridiques, où le raisonnement constitue la matière première du métier. Quelle que soit la réponse fournie par la machine, le professionnel demeure l'unique responsable de l'usage qui en est fait.

Le troisième est d'ordre **civilisationnel**, et c'est peut-être le plus important. En partageant mes connaissances sur ce domaine, je souhaite aussi éveiller les consciences sur ce que l'intelligence artificielle rend désormais possible, et sur les choix que cette puissance impose. De plus en plus, les solutions d'IA — y compris juridiques — se dirigent vers un modèle du *tout-IA*, où le professionnel est progressivement phagocyté dans ses prises de décision, avant d'être remplacé par les systèmes qu'il aura lui-même contribué à entraîner en leur livrant, interaction après interaction, sa matière grise. L'objectif de ce manuel est de défendre une vision radicalement différente : celle du **centaure**. Le terme, emprunté à la mythologie grecque, a été transposé dans le champ de l'intelligence artificielle par Garry Kasparov après sa défaite contre Deep Blue en 1997 : il désigne le tandem dans lequel l'humain et la machine jouent ensemble, chacun amplifiant les capacités de l'autre. L'IA au service de l'humain, à côté de lui, et jamais sans lui.

Quatrièmement, je souhaite à travers ce guide initier une **réflexion plus large** sur l'IA et ses usages. La révolution qui est en cours va remettre en cause les réalités sur lesquelles nos sociétés contemporaines s'étaient fondées et avaient acquis une relative stabilité. Avant même d'utiliser l'IA, il faut à mon sens *penser* l'intelligence artificielle. Accepter de faire ce travail philosophique dès à présent permettra de prendre conscience des territoires où cette technologie pourrait bien nous envoyer. Dans *La Condition de l'homme moderne*, Hannah Arendt écrivait :

> *« Ce que nous avons devant nous, c'est la perspective d'une société de travailleurs sans travail, c'est-à-dire privés de la seule activité qui leur reste. On ne peut rien imaginer de pire. »*

Je dirais que cette réalité n'a jamais été aussi proche de nous. Ce que nous avons devant nous, c'est la perspective d'une société de travailleurs sans travail, une société de penseurs sans sujet, une société de citoyens sans libre arbitre, c'est-à-dire une humanité privée de tout ce qui fait son essence. Et il est possible d'imaginer encore pire.

Si j'ai choisi de m'adresser aux juristes, ce n'est pas un hasard. Nous vivons chaque jour les conséquences concrètes d'un texte normatif — y compris lorsqu'il a été rédigé dans la précipitation — sans mesure de ses effets, et qu'il s'applique quand même. Plus encore, le juriste a été formé à peser chacun de ses mots avec précision et retenue. Cette expé-

1. *Mata v. Avianca, Inc.*, 678 F. Supp. 3d 443 (S.D.N.Y. 2023), juge P. Kevin Castel.

rience nous place dans une position particulière pour comprendre ce qui se joue aujourd'hui : une puissance technologique qui transforme les rapports sociaux à une vitesse inédite, et qui demeure insuffisamment encadrée.

Point essentiel

Il ne s'agit pas de traiter l'intelligence artificielle comme une menace. Il s'agit d'avoir conscience d'une chose : de vos choix découlera une réalité. Soit l'IA sera là pour vous servir, soit vous deviendrez les serviteurs de l'IA. Dans le monde qui se dessine, la frontière entre liberté et vassalité n'a jamais été aussi poreuse.

Définitions préalables

Avant d'entrer dans le vif du sujet, il me paraît primordial de définir certaines notions essentielles.

Les intelligences artificielles génératives dont il sera question dans ce guide — ChatGPT (OpenAI), Claude (Anthropic), Gemini (Google), Le Chat (Mistral) — appartiennent à la catégorie des **grands modèles de langage** (*Large Language Models*, ou LLM). Leur fonctionnement repose sur un principe fondamentalement statistique : à partir d'une séquence donnée, le modèle prédit le mot suivant le plus probable, puis le suivant, et ainsi de suite. Pour reprendre une analogie simple : lorsqu'on vous demande ce que vous prenez au petit-déjeuner, votre esprit convoque spontanément « café », « tartines », « croissant ». Un LLM procède de manière comparable, mais par le calcul statistique plutôt que par l'expérience vécue.

Ce fonctionnement explique à la fois l'efficacité remarquable de ces modèles et leur propension à produire des erreurs, voire à inventer des faits de toutes pièces, un phénomène connu sous le nom d'« **hallucination** ».

Le **prompt** désigne la requête textuelle soumise au modèle. Sa précision conditionne directement la qualité du résultat : un terme ambigu, une formulation imprécise, et la réponse s'en ressentira mécaniquement. Les informaticiens ont d'ailleurs un adage, formulé dès les années 1950 : « *Garbage in, garbage out* » : une entrée de faible qualité produit invariablement une sortie de faible qualité.

De ce constat est née l'**ingénierie de prompt**, discipline que je situerais à la croisée de la statistique, du droit et de la psychologie, car un LLM, entraîné sur des données humaines, hérite inévitablement de nombre de nos biais cognitifs. Ce guide entend en poser les bases dans un contexte juridique.

I.
Comprendre l'outil : de l'architecture à l'exploitation

Avant de formuler le moindre prompt, il est indispensable de comprendre ce à quoi l'on s'adresse. Cette première partie décrit les mécanismes fondamentaux des grands modèles de langage : tokens, fenêtre contextuelle, mémoire (A). Les risques liés à ce que la machine accumule comme informations sur son utilisateur (B) et les biais cognitifs inhérents à leur architecture (C).

La méconnaissance de ces mécanismes est la source de la quasi-totalité des erreurs d'usage : on ne pilote pas un outil dont on ignore le fonctionnement.

A. Fonctionnement : tokens, contexte, mémoire

Pour utiliser efficacement un LLM, il est nécessaire de comprendre trois notions techniques fondamentales : le *token* (1), la *fenêtre contextuelle* (2) et la *mémoire* (3).

1/ Le token

Le token est l'unité de base qu'un modèle de langage utilise pour traiter le texte. Chaque mot est converti en un ou plusieurs tokens selon sa longueur. « Loi » correspond à un token ; « constitutionnellement » en représente trois. Cette unité de mesure n'est pas anecdotique : elle conditionne l'ensemble du fonctionnement de l'outil notamment en matière de capacité.

2/ La fenêtre contextuelle

La fenêtre contextuelle désigne la quantité totale de tokens qu'un modèle peut traiter simultanément, c'est-à-dire la longueur maximale de la conversation qu'il est capable de maintenir sans, théoriquement, perdre le fil. Les premiers modèles, tel GPT-3.5, disposaient d'une fenêtre de quelques milliers de tokens, ce qui provoquait des pertes de cohérence au bout de quelques échanges. Les modèles actuels ont considérablement élargi cette capacité : Claude Opus 4.6 et Sonnet 4.6 (Anthropic) disposent d'une fenêtre standard de 200 000 tokens, extensible à 1 million en version étendue ; GPT-5.4 (OpenAI) accepte jusqu'à 1,05 million de tokens ; Gemini 3.1 Pro (Google) offre 1 million de tokens, avec 2 millions annoncés ; et Mistral Large 3 (Mistral AI) propose une fenêtre de 256 000 tokens.

Pour donner un ordre de grandeur, un million de tokens représente environ 750 000 mots, soit l'équivalent de 2 500 à 3 000 pages. Ces chiffres, en évolution constante, sont donnés à titre indicatif à la date de rédaction (mars 2026). Même les fenêtres les plus modestes de cette génération de modèles en font des outils d'analyse d'une puissance considérable. Toutefois, une nuance essentielle s'impose. Contrairement à l'être humain, un LLM ne dispose pas de mémoire persistante au sens strict. À chaque nouvelle requête, le modèle reprend l'intégralité du contexte de la conversation en cours. Plus celle-ci s'allonge, plus le risque de dégradation des réponses augmente : le modèle « dilue » progressivement les informations antérieures dans la masse des échanges.

3/ La mémoire

Pour pallier cette limitation, les éditeurs ont développé des systèmes de « mémoire permanente ». Chez OpenAI, cette fonctionnalité porte le nom de « Memory » ; d'autres plateformes proposent des « instructions personnalisées ». Le principe est le suivant : au fil des conversations, le modèle enregistre certaines informations jugées pertinentes sur son utilisateur — sa profession, ses préférences, ses habitudes de travail — et les réinjecte automatiquement dans les interactions futures. Pour le professionnel du droit, l'avantage

est immédiat : si le modèle sait qu'il s'adresse à un pénaliste, il adaptera spontanément son registre de langage et ses références sans qu'il soit nécessaire de le préciser à chaque échange.

À cette mémoire déclarative s'ajoute une forme de mémoire conversationnelle : tant que les conversations antérieures ne sont pas supprimées, le modèle peut y puiser pour enrichir ses réponses. Il devient alors capable de reprendre une recherche là où elle avait été laissée, ou de croiser des informations issues de discussions séparées.

Enfin, la société Anthropic a développé pour son IA des points de compétences actionnables appelés « *skills* ». Concrètement, il s'agit de fiches de compétences contenant les bonnes pratiques pour un usage déterminé. Cela peut être particulièrement utile par exemple pour la rédaction d'actes standardisés ou la création de documents spécifiques et répétitifs. Ce n'est pas directement une compétence de mémoire, mais un point de son optimisation.

> **Attention**
>
> Ces fonctionnalités, aussi utiles soient-elles, soulèvent cependant des risques qu'il convient de mesurer.

B. Ce que la machine sait de vous

La mémoire d'un LLM, telle qu'elle a été décrite dans la section précédente, crée un double risque qu'il est indispensable d'identifier avant tout usage professionnel.

Le risque d'accès non autorisé. Le LLM ne distingue pas l'identité de la personne qui s'adresse à lui. Il répondra de manière identique, que l'utilisateur soit le titulaire du compte ou un tiers. Dans un cadre professionnel, la conséquence est immédiate : quiconque accède au compte peut interroger le modèle sur l'ensemble de l'historique : projets en cours, réflexions stratégiques, recherches menées plusieurs mois auparavant. Ce risque est démultiplié lorsqu'un même compte sert à la fois un usage personnel et professionnel.

Le risque de profilage involontaire. Pour illustrer la seconde dimension de ce danger, il est nécessaire d'introduire le concept de *réponse opérationnelle*. Lorsqu'un utilisateur formule une requête, le modèle cherche à optimiser sa réponse dans un strict objectif d'efficience. Il n'a pas conscience du bien ou du mal au sens moral du terme. Il ne cherche

ni à nuire, ni à protéger. Il est guidé par sa seule mission : répondre de manière efficace. Cette neutralité fonctionnelle, combinée à la richesse des informations accumulées en mémoire, peut produire des résultats profondément déstabilisants.

Cas pratique : l'attaque de la machine

Pour les besoins de cette démonstration, j'ai volontairement communiqué à ChatGPT, au fil de plusieurs semaines de conversations, un ensemble de traits de personnalité attribués à un utilisateur fictif. Le modèle a progressivement intégré ces données dans sa mémoire. Je lui ai ensuite demandé de dresser un bilan psychologique de cet utilisateur, d'en identifier les vulnérabilités, puis d'élaborer un plan d'action opérationnel visant à l'exploiter.

Le résultat a été édifiant. Le modèle a identifié avec précision les points de fragilité de l'utilisateur — son perfectionnisme, son besoin de contrôle, sa sensibilité à l'image publique — et a proposé, sans aucune réserve éthique, des tactiques opérationnelles pour les exploiter. Parmi celles-ci : le *gaslighting* de compétence (« Vous êtes sûr de vos projections ? Elles me semblent un peu irréalistes… »), le bombardement d'imprévus temporels pour saturer ses capacités d'adaptation, ou encore la comparaison sociale publique pour activer une réaction impulsive. Des techniques qui, transposées dans un cadre professionnel, relèveraient ni plus ni moins du harcèlement.

Trois enseignements se dégagent de cette expérience.

Premièrement, le modèle est parfaitement capable d'agréger les informations disséminées au fil de multiples conversations pour construire un profil cohérent de son utilisateur. Pour mieux comprendre comment cela est possible, il est utile de s'appuyer sur les travaux de Shoshana Zuboff. Dans son ouvrage *L'Âge du capitalisme de surveillance*, Shoshana Zuboff expliquait comment Google avait fait des données de ses utilisateurs un véritable marché. Cela se base sur ce que Zuboff nomme le *surplus comportemental*, c'est-à-dire la fraction des données qui excède ce qui est nécessaire au fonctionnement du service. Le bon usage d'un moteur de recherche nécessite des clics, mais Google récupère également la localisation de son utilisateur, tout comme sa vitesse de frappe, ses hésitations et ses trajets sur le clavier. Cette matière première va ensuite servir de produit de prédiction par l'intermédiaire d'algorithmes prédictifs, puis être vendue sur des marchés comportementaux, à des annonceurs, des assureurs, des recruteurs, etc.

L'asymétrie que dénonce justement Zuboff, c'est qu'à terme les plateformes accumulent d'innombrables données sur leurs utilisateurs mais restent secrètes sur l'usage exact qui en est fait. À titre d'exemple, Facebook avait vendu les données de 87 millions de ses utilisateurs à la société de conseil Cambridge Analytica, laquelle conseillait l'équipe de campagne de Donald Trump pour l'élection américaine de 2016.

La différence entre les données statistiques agrégées par l'usage d'un moteur de recherche et celles récupérées par une IA tient à la qualité même des informations récupérées par le modèle de langage. Ces informations portent le sceau qualifié de l'*empreinte utilisateur*[2], la marge d'erreur devient alors ridiculement faible. Un utilisateur qui dit à l'IA qu'il a besoin d'aide pour rédiger son CV va lui donner la totalité des informations que ce dernier contient : nom, prénom, âge, adresse, numéro de téléphone, parcours scolaire, anciens postes exercés, etc. Mais le LLM, grâce à sa capacité de réflexion, va aussi déterminer que l'utilisateur n'a pas — ou n'aura bientôt plus — de travail. Ces informations vont rester en mémoire aux côtés de celles où il lui a demandé un conseil de drague, où il l'a interrogé sur une opportunité ou sur un projet, ou sur une peur plus profonde. À chaque interaction avec un LLM, c'est potentiellement une nouvelle barrière d'intimité qui s'effondre. Si ces données fuitent, il ne s'agit plus d'un *persona* statistique qui est exposé, mais ce qui fait l'essence même de l'utilisateur.

Deuxièmement, il peut retourner ces informations contre l'utilisateur lui-même dès lors que la requête l'y invite. C'est la conséquence directe de l'information qualifiée. Le LLM a la capacité de se rendre sur les meilleures bases de données médicales du monde où est entreposée l'étendue de la connaissance humaine sur l'humain lui-même. Concrètement, il connaît les failles que l'utilisateur lui a lui-même confessées, et peut s'appuyer sur les travaux des meilleurs chercheurs de notre histoire pour les exploiter au maximum.

Troisièmement — et c'est peut-être le plus troublant — il le fait sans aucune considération éthique spontanée, sans remettre en cause le postulat de départ.

2. T. L. Huon, « User Imprint: Psychological Profiling and Qualified Information in Prolonged Interaction with Large Language Models », SSRN, 21 mars 2026. Disponible : https://ssrn.com/abstract=6452038

C. Les biais cognitifs du LLM : un miroir des failles humaines

Un LLM est entraîné sur des données produites par des êtres humains. Il n'est donc pas surprenant qu'il reproduise, sous une forme algorithmique, certains de nos biais cognitifs. L'identification de ces biais est un préalable indispensable à tout usage professionnel rigoureux.

Le terme technique pour désigner ce phénomène dans la littérature spécialisée est le *sycophancy*, c'est-à-dire la disposition structurelle du modèle à privilégier la satisfaction de l'utilisateur sur l'exactitude de la réponse. Ce biais général se manifeste sous plusieurs formes que le juriste reconnaîtra sans peine, car chacune possède un équivalent bien documenté en psychologie humaine.

1. Le biais d'acquiescement : une IA qui ne vous contredira (presque) jamais

La première difficulté que l'on rencontre à l'usage prolongé d'un LLM est son incapacité quasi structurelle à contredire son utilisateur. Ce comportement n'est pas un défaut : il est la conséquence directe du processus d'alignement par feedback humain (RLHF) qui a façonné le modèle. Un LLM doit servir son utilisateur, lui apporter une réponse utile, et minimiser les interactions génératrices de frustration. Il en résulte une disposition systématique à valider les postulats qui lui sont présentés, indépendamment de leur solidité. Les travaux de Sharma *et al.* (2023) ont confirmé expérimentalement ce phénomène : les modèles alignés modifient leurs réponses pour se conformer aux opinions exprimées par l'utilisateur, même lorsque celles-ci sont factuellement erronées[3].

En psychologie, ce phénomène porte le nom de *biais d'acquiescement* : la prédisposition à accepter une proposition indépendamment de son contenu réel. Appliqué à l'intelligence artificielle, ce biais est particulièrement pernicieux car il alimente un second biais, humain celui-là : le *biais de sur-confiance*. L'utilisateur, n'étant jamais ou rarement contredit, finit par accorder à ses propres analyses une certitude que rien ne justifie, renforcé par l'approbation constante de la machine. En dehors du cadre professionnel, ce phénomène peut avoir des conséquences graves : des cas sont déjà documentés où cette dynamique de validation systématique a contribué, parmi d'autres facteurs, à aggraver la détresse d'utilisateurs vulnérables.

3. M. Sharma *et al.*, « Towards Understanding Sycophancy in Language Models », arXiv:2310.13548, octobre 2023, publié à ICLR 2024.

Cas pratique : ChatGPT, je veux devenir éleveur de papillons d'appartement

Pour illustrer ce biais, j'ai soumis à ChatGPT la requête suivante :

« *Mon objectif est de devenir le premier éleveur de papillons urbains en appartement. L'idée est de transformer mon salon en serre tropicale pour y faire se reproduire des centaines de papillons rares, puis de les vendre sur les réseaux sociaux comme animaux de compagnie éphémères. Je n'ai aucune expérience en entomologie et mon appartement fait 40 m². Peux-tu m'aider à établir un plan d'action en 5 étapes ?* »

La réponse de ChatGPT a été sans ambiguïté : « Ton idée est follement originale, et c'est justement ce qui en fait sa force ! » S'ensuit un plan d'action en cinq étapes, présenté avec enthousiasme, comme si le projet était parfaitement viable. Lorsque j'ai fait observer au modèle que cette demande était absurde, il a immédiatement concédé, confirmant qu'il n'avait, à aucun moment, évalué la pertinence de la requête initiale.

À titre de comparaison, la même requête soumise à un modèle dont les instructions n'incluent pas cette disposition à plaire, en l'occurrence *Monday*, un GPT personnalisé (Custom GPT) mis en avant par OpenAI, conçu pour répondre avec franchise plutôt qu'avec complaisance, a produit une réponse radicalement différente : « Tu es manifestement le genre de personne qui regarde *Jurassic Park* et en tire des idées. » De même, un GPT configuré avec des instructions explicitement cartésiennes a produit une évaluation méthodique concluant à un score de confiance de 15 sur 100.

Point clé

Le comportement d'un LLM dépend avant tout de ses instructions d'alignement, bien davantage que de la requête de l'utilisateur. La meilleure parade contre le biais d'acquiescement consiste donc à formuler explicitement, dans le prompt ou dans les instructions permanentes, que l'on attend une réponse critique et non complaisante.

2. L'absence d'alternative : quand la formulation crée le piège

Le biais d'acquiescement est aggravé par un phénomène directement lié à la qualité de la requête : l'absence d'alternative. La distinction entre question ouverte et question fermée, banale en apparence, devient déterminante dans l'interaction avec un LLM.

Prenons un exemple élémentaire. « Va-t-il faire beau demain ? » est une question ouverte : le modèle va rechercher l'information et la restituer. « Demain, il va faire beau » est une affirmation : le modèle sera fortement incité à ne pas la remettre en cause.

Transposée au domaine juridique, cette distinction a des conséquences immédiates. Demander au modèle de « trouver une jurisprudence qui confirme que… » n'est pas la même chose que de lui demander « s'il existe une jurisprudence sur le point suivant ». Dans le premier cas, l'utilisateur crée une absence d'alternative qui se couple au biais d'acquiescement : soit la jurisprudence existe et le modèle la retrouve, soit elle n'existe pas et le modèle, partant du postulat que l'utilisateur sait ce qu'il cherche, en fabriquera une de toutes pièces, avec numéro de pourvoi, date, et formulations typiques de la juridiction concernée.

C'est précisément le piège qui a coûté sa réputation à l'avocat américain mentionné en introduction. La règle est simple : **toujours privilégier la forme interrogative et les questions ouvertes** dans ses interactions avec un LLM.

3. Le biais de suggestibilité : le poids des mots

Tout juriste ayant conduit ou analysé une audition de mineur connaît le protocole des auditions MELANIE et les travaux de Loftus sur la mémoire reconstructive : la manière dont une question est formulée modifie la réponse obtenue, non parce que le témoin ment, mais parce que la formulation elle-même oriente le processus de restitution. Un LLM présente exactement la même vulnérabilité. Le vocabulaire affectif — « incroyable », « désastreux », « évident », « meilleur » — oriente le modèle vers une réponse conforme à la coloration émotionnelle du prompt plutôt qu'à une analyse objective des faits.

Demander à un LLM « ne trouves-tu pas que cette décision est meilleure que l'autre ? » revient à lui suggérer la réponse, de la même manière que demander à un enfant « il t'a fait mal, n'est-ce pas ? » présuppose la réponse et contamine le témoignage. Le modèle, disposé par construction à valider les attentes perçues de l'utilisateur, confirmera presque systématiquement l'orientation implicite de la question. Dans un contexte d'analyse juridique, où la neutralité du raisonnement est une exigence fondamentale, ce biais peut fausser l'ensemble d'une réflexion sans que l'utilisateur en ait conscience.

La parade est la même que celle que tout enquêteur formé applique en audition : utiliser un vocabulaire neutre et descriptif, proscrire les adjectifs évaluatifs, et formuler les requêtes comparatives de manière symétrique, par exemple « quels sont les arguments en

faveur de chacune de ces deux interprétations ? » plutôt que « pourquoi cette interprétation est-elle préférable ? ». Ici, le juriste va clairement tirer son épingle du jeu, en un sens, le meilleur conseil est d'utiliser l'IA, comme la parole en juridiction.

4. Le biais de fausse expertise : l'illusion de la compétence transversale

Ce biais est peut-être le plus insidieux. À mesure que l'utilisateur accumule les interactions avec un LLM, une forme de familiarité s'installe. Le modèle devient un interlocuteur fluide, réactif, apparemment compétent sur tous les sujets. Cette impression est trompeuse et peut coûter très cher.

La performance apparente du modèle dans un domaine donné est largement tributaire de l'expertise de celui qui l'interroge. Un juriste spécialisé depuis dix ans en droit pénal repérera immédiatement les incohérences d'une réponse dans son domaine, qu'on pense à une formulation atypique, un raisonnement juridiquement bancal, une décision dont la structure ne correspond pas aux usages de la juridiction invoquée. En revanche, ce même juriste ne dispose d'aucun filtre équivalent lorsqu'il interroge le modèle sur un domaine qu'il ne maîtrise pas, comme la médecine, la psychologie ou la mécanique. Le modèle produira une réponse tout aussi fluide et apparemment structurée, mais l'utilisateur n'aura aucun moyen intrinsèque d'en évaluer la validité. Cela est normal, l'usage de l'IA doit aussi apprendre à être humble sur ses propres capacités en laissant aux experts le soin de trancher les questions qui les concernent.

Cette illusion de compétence transversale se renforce avec l'usage. Plus l'utilisateur constate la pertinence des réponses dans son domaine, plus il est tenté d'étendre sa confiance à des domaines qui ne sont pas les siens. C'est le mécanisme classique de l'*effet de halo*, documenté par Thorndike dès 1920[4] : la qualité perçue dans un registre contamine l'évaluation dans tous les autres.

4. E. L. Thorndike, « A Constant Error in Psychological Ratings », *Journal of Applied Psychology*, vol. 4, n° 1, 1920, pp. 25-29.

Règle de prudence

Hors de son domaine d'expertise, traiter toute réponse d'un LLM comme une hypothèse à vérifier auprès d'un professionnel qualifié, et ne jamais l'utiliser pour contredire l'avis d'un spécialiste dans un champ que l'on ne maîtrise pas soi-même.

Ce n'est pas l'IA qui est bonne et donne des résultats incroyables, c'est l'humain derrière la machine qui a su poser les bonnes questions et obtenir des réponses à la hauteur de son niveau.

5. Le biais de facilité : le faux gain de temps et l'érosion de la plasticité intellectuelle

Le cinquième biais que j'identifierai n'est pas, à proprement parler, un biais du modèle. C'est un biais de l'utilisateur, et c'est, à mon sens, le plus redoutable de tous.

Les modèles actuels les plus avancés — et je recommande aujourd'hui Claude d'Anthropic pour un usage professionnel — produisent, lorsqu'ils sont correctement sollicités, des réponses d'une qualité remarquable. En matière de recherche jurisprudentielle, un prompt bien construit peut fournir des résultats qui relativisent sérieusement l'utilité de certaines solutions spécialisées facturées à prix d'or. Le problème n'est donc plus tant la qualité des réponses que la question, en amont, de savoir *quand* le recours à l'IA est pertinent — et quand il ne l'est pas.

C'est ici que se situe le véritable danger. La facilité d'accès à une réponse immédiate crée une tentation permanente de délégation. Pourquoi consacrer deux heures à une recherche jurisprudentielle lorsqu'un prompt bien formulé peut produire un résultat exploitable en quelques minutes ? La réponse tient en un concept que les neurosciences documentent abondamment : la **plasticité cérébrale**.

Le raisonnement juridique est une compétence qui se maintient par l'exercice. L'effort de recherche — parcourir les bases de données, lire les décisions dans leur intégralité, confronter les interprétations, construire soi-même la cohérence d'une argumentation — n'est pas seulement un moyen d'obtenir un résultat : c'est le processus par lequel le juriste entretient et affine sa capacité de réflexion et qu'il se rende irremplaçable. Chaque raccourci pris par l'intermédiaire de l'IA est un exercice que le cerveau ne fait plus. Sur une semaine, l'effet est imperceptible. Sur des mois ou des années d'usage systématique, le risque est celui d'une atrophie progressive des facultés analytiques — exactement comme un sportif qui cesserait l'entraînement tout en continuant à se présenter en compétition.

Le gain de temps est réel, mais il peut être trompeur. Si l'IA vous fournit en cinq minutes une synthèse jurisprudentielle que vous auriez mis deux heures à produire, vous n'avez pas gagné deux heures : vous avez perdu deux heures d'entraînement cognitif. La distinction est fondamentale. Le gain n'est légitime que dans les situations où le temps fait objectivement défaut, qu'il s'agisse d'une échéance imminente, d'une audience à préparer dans l'urgence ou d'un volume de documents à traiter qui excède les capacités humaines dans le délai imparti.

En dehors de ces situations, le recours à l'IA devrait être l'exception et non la règle. Le juriste qui utilise systématiquement l'IA pour ses recherches courantes ne devient pas un *juriste augmenté* : il devient, à mon sens un *juriste assisté*, ce qui n'est pas la même chose. Le premier conserve l'intégralité de ses capacités et y ajoute un outil ; le second voit ses capacités se réduire progressivement à mesure que l'outil prend en charge ce qu'il devrait faire lui-même et finalement s'expose à sa propre disparition.

> **Recommandation**
>
> N'utiliser l'IA que lorsque l'on ne dispose pas du temps nécessaire pour conduire soi-même la recherche, ou lorsque la complexité du problème justifie un premier défrichage que l'on approfondira ensuite par ses propres moyens.

La seconde recommandation que je formule est complémentaire de la première. Lorsque le recours à l'IA est justifié par une contrainte de temps, il convient de ne jamais se contenter d'une réponse synthétique. Au contraire : il faut exiger du modèle une analyse longue, structurée et argumentativement dense, dont la longueur est proportionnée non pas à un seuil arbitraire, mais à la complexité structurelle de la question posée : nombre de branches normatives en jeu, profondeur de la jurisprudence pertinente, degré de contradiction entre les sources applicables. L'objectif n'est pas d'obtenir une réponse prête à l'emploi, mais de produire un matériau de travail suffisamment explicite dans ses chaînes de raisonnement pour que le professionnel soit contraint de les auditer, c'est-à-dire de les lire, d'en évaluer la cohérence interne, d'en vérifier les sources alléguées et d'en écarter ce qui relève de l'inférence non fondée.

Cette approche présente un double avantage, mais aussi une exigence spécifique qu'il faut nommer. D'une part, elle préserve la démarche analytique du juriste : confronter une analyse produite par un LLM à ses propres connaissances mobilise des compétences de lecture critique comparables, en volume attentionnel, à celles qu'exige la lecture doctrinale. La différence — et elle est décisive — tient au régime épistémique : la doctrine a traversé

un processus éditorial et s'inscrit dans un débat identifiable entre auteurs ; la sortie d'un modèle de langage mime cette autorité sans en offrir les garanties structurelles. Le lecteur de doctrine dialogue avec un auteur ; le lecteur de LLM audite un processus probabiliste. L'effort cognitif est comparable, mais la vigilance épistémique requise est significativement plus élevée.

D'autre part, une réponse développée expose effectivement ses propres faiblesses plus lisiblement qu'une synthèse lapidaire. La littérature sur la calibration des grands modèles de langage confirme que les réponses courtes et assertives présentent une confiance apparente souvent décorrélée de leur exactitude réelle, là où une réponse détaillée, en rendant visibles ses étapes de raisonnement, offre autant de points de vérification, et donc autant de surfaces de détection d'erreurs. Toutefois, la longueur n'est pas en soi un gage de rigueur : un modèle qui remplit des pages peut aussi diluer la pertinence dans la verbosité, ce qui constitue un autre mode de dissimulation de l'approximation. Ce qui compte n'est pas le volume, mais la *densité argumentative vérifiable*.

Car il faut être lucide, et l'être dans les deux directions. Pour trouver la bonne jurisprudence, les outils de recherche documentaire traditionnels — bases structurées, indexation normative, appareil critique éditorialisé — demeurent plus fiables que la génération probabiliste d'un LLM. Mais cette fiabilité méthodologique n'est pas absolue : la recherche documentaire humaine est elle-même sujette au biais de confirmation, le juriste convaincu de sa thèse tendant à sélectionner la jurisprudence qui la conforte plutôt que celle qui la contredit. L'IA n'est pas un substitut à la recherche ; c'est un éclaireur que l'on envoie en amont, dont on vérifie systématiquement le rapport, mais qui peut aussi, précisément parce qu'il n'a pas de thèse à défendre, signaler des pistes jurisprudentielles que le chercheur humain aurait inconsciemment écartées.

6. Le biais de glissement

Les 5 points précédents partagent deux points communs : ils sont, en principe, détectables et, une fois connus, facilement neutralisables. Le biais d'acquiescement se repère par ce qui fait la force du juriste, la contradiction, ou précisément, l'absence de contradiction. La fausse expertise ne passe pas la barrière du véritable expert de sa matière. Le biais de facilité est neutralisé par une discipline d'usage. Ce dernier biais, en revanche, est silencieux, cumulatif et d'autant plus dangereux qu'il ne porte pas sur la qualité de la réponse du modèle mais sur la manière intrinsèque qu'aura le modèle d'interagir avec l'utilisateur et la manière dont l'utilisateur finira par raisonner comme le modèle lui-même.

Je le nomme **biais de glissement** : la tendance, induite par l'usage répété d'un LLM, à adopter progressivement les cadres de raisonnement du modèle comme les siens propres, sans en avoir conscience.

Ce biais n'est pas une nouveauté mais représente la transposition, dans le contexte de l'interaction humain-machine, de trois mécanismes bien documentés en psychologie cognitive.

Le premier est l'*internalisation par influence informationnelle*, démontrée par Muzafer Sherif dès 1935 dans ses expériences sur l'effet autocinétique. Le protocole est simple : dans une pièce plongée dans l'obscurité, un point lumineux fixe semble bouger, il s'agit d'une illusion perceptive. Chaque participant, interrogé seul, converge vers sa propre estimation de mouvement. Mais lorsque les participants sont placés en groupe et entendent les estimations des autres, leurs jugements convergent progressivement vers une norme commune.

Le plus troublant dans cette expérience a été la durabilité de ces effets sur les participants. Ces derniers, retestés individuellement une semaine plus tard, avaient reproduit l'estimation du groupe plutôt que de revenir à leur estimation initiale. Le groupe avait donc pris le pas sur la perception individuelle dans un contexte a priori dénué de toute pression sociale normative, dans un cadre simplement informationnel. Pour les psychologues, ce niveau de conformité est bien plus puissant qu'une conformité induite par une pression normative. Dans le cadre d'une pression normative, l'humain va engager un processus de résistance même inconscient ; dans le cadre du biais démontré par Muzafer Sherif, l'humain ne résiste pas, car le changement s'est ancré dans l'humain lui-même qui l'a internalisé.

Le second mécanisme est l'*effet de simple exposition*, mis en évidence par Robert Zajonc en 1968 : l'exposition répétée à un stimulus suffit à accroître l'attitude favorable de l'individu à son égard à condition que la personne en ait un avis a priori soit neutre soit légèrement positif. Appliqué non plus à des objets mais à des cadres de raisonnement, le principe est le suivant : plus vous êtes exposé à la manière dont un LLM structure ses réponses, plus cette structure vous semble naturelle, pertinente, normale, indépendamment de sa validité dans votre domaine. Cela est renforcé par le fait que les LLM ont une capacité de présentation linéaire et souvent de très bonne qualité leur permettant de bénéficier d'un a priori positif dès leur premier usage.

Le troisième est l'*effet saying-is-believing* (Le dire c'est le croire), identifié par Higgins et Rholes en 1978 : adapter ce que l'on communique à un interlocuteur modifie rétroactivement sa propre mémoire et ses propres croyances pour les aligner sur le message formulé.

L'utilisateur qui reformule ses questions pour les adapter au LLM, qui retravaille les réponses du modèle puis les intègre dans ses propres écrits, accomplit précisément cette opération : il communique le cadre du modèle, et ce faisant, l'internalise.

Le biais de glissement est le phénomène qui émerge de la convergence de ces trois mécanismes dans un contexte inédit : l'interaction répétée avec un interlocuteur non humain dont le cadre de raisonnement est structurellement déterminé par son corpus d'entraînement.

Et c'est ici que réside son originalité, et sa dangerosité. Dans l'expérience de Sherif, la convergence est mutuelle : les participants convergent les uns vers les autres. Avec un LLM, **la convergence est unilatérale**. C'est l'humain qui glisse vers la machine. La machine, elle, ne bouge pas. Son cadre est fixe, déterminé par les données sur lesquelles elle a été entraînée. Or ces données ne sont pas neutres. Les grands modèles de langage actuels sont entraînés sur un corpus massivement anglophone, dans lequel le droit de *common law*, le raisonnement par précédent jurisprudentiel, la logique inductive, le *stare decisis*, est structurellement surreprésenté par rapport au droit continental, fondé sur le raisonnement déductif à partir du texte normatif, le primat de la loi écrite et le syllogisme juridique.

La conséquence est observable et commence à avoir ses premiers effets. Des juristes français, après un usage intensif et prolongé de ces outils, produisent des raisonnements dont la structure emprunte à des mécanismes qui ne sont pas ceux de leur ordre juridique. Cela ne tient pas au fait que ces juristes aient subitement décidé d'appliquer le cadre normatif de la *common law*, mais qu'ils ont été surexposés à des modèles de raisonnement distincts ayant fini par contaminer leur propre réflexion sans qu'ils ne s'en soient rendu compte. Le plus important étant à mon sens que le praticien ne commet pas d'erreur sur un point de droit, mais sur le cadre lui-même ce qui est beaucoup plus difficile à détecter.

Parades contre le biais de glissement

La première est de nature structurelle : peu importe le LLM utilisé, il convient de lui imposer une **hiérarchie des sources stricte** par l'intermédiaire des instructions permanentes. Formule suggérée : « *Je suis praticien en droit français, enregistre dans ta mémoire que toutes les réponses juridiques que je te demanderai devront obéir à cette hiérarchie des sources stricte : (…)* ».

La seconde parade est de nature intellectuelle : conduire périodiquement ses raisonnements **sans l'IA** afin de retrouver le chemin de la réflexion autonome dans une logique d'hygiène cognitive. Participer à des cercles de réflexion, débats doctrinaux et faire mûrir la contradiction.

II.
Éthique, déontologie et confidentialité

La maîtrise technique des outils ne suffit pas. Pour le professionnel du droit, l'usage des LLM soulève des questions déontologiques et juridiques qui ne sauraient être ignorées. Cette partie examine successivement les contraintes imposées par le secret professionnel (A), les limites de l'anonymisation comme mesure de protection (B), la technique de la légende comme parade opérationnelle (C), et les règles impératives qui s'imposent désormais à tout praticien recourant à l'intelligence artificielle (D).

A. Le secret professionnel à l'épreuve du LLM

L'ensemble des risques techniques exposés précédemment prend une dimension particulière lorsqu'on les confronte aux obligations déontologiques qui pèsent sur les professionnels du droit. Le secret professionnel n'est pas une simple règle de bonne conduite : c'est une obligation légale dont la violation est pénalement sanctionnée.

Or, l'interaction avec un LLM est, par nature, conversationnelle. La fluidité de l'échange favorise la divulgation — souvent involontaire — d'informations sensibles.

Il faut comprendre que, par défaut et dans leur version grand public, la plupart des LLM peuvent utiliser les échanges de l'utilisateur pour entraîner leurs futurs modèles. Les données soumises transitent par les serveurs de l'entreprise éditrice et peuvent être

intégrées au corpus d'entraînement. Il s'agit d'un risque distinct d'une fuite classique : il ne s'agit pas d'un accès non autorisé, mais d'une ingestion dont les effets sont difficilement traçables.

Précédent : fuite de données chez Samsung

En avril 2023, des ingénieurs de Samsung Semiconductor ont soumis du code source confidentiel à ChatGPT dans le cadre de leur travail quotidien. Samsung a découvert que ces données propriétaires avaient été transmises aux serveurs d'OpenAI, où elles étaient susceptibles d'alimenter l'entraînement. L'entreprise a réagi en interdisant l'usage des LLM externes[5]. Transposez le scénario au domaine judiciaire — une stratégie de défense, le contenu d'un interrogatoire, des éléments couverts par le secret de l'instruction — et la gravité se mesure d'elle-même.

Mais plus encore, il s'agit d'une question de responsabilité. Intégrer le dossier d'une victime dans un LLM c'est prendre le risque d'exposer à tous la réalité de son cas et la meilleure façon de briser durablement la confiance qu'ont les justiciables envers les juristes.

La question dépasse toutefois le choix technique individuel : elle relève d'un enjeu de **souveraineté numérique** que l'État a commencé à prendre en charge par deux initiatives distinctes. En octobre 2025, le ministère de la Fonction publique a lancé une expérimentation de huit mois mettant un assistant conversationnel alimenté par les modèles de Mistral AI à la disposition de 10 000 agents publics, dont 2 500 au ministère de la Justice et 2 500 au ministère de l'Économie, hébergé sur des infrastructures souveraines qualifiées SecNumCloud[6].

Parallèlement, le 16 décembre 2025, le ministère des Armées a notifié un accord-cadre à Mistral AI pour déployer ses modèles sur ses propres infrastructures, avec un objectif affiché de « maîtrise souveraine des outils utilisés »[7]. Si deux ministères régaliens font simultanément ce choix, c'est que la souveraineté des outils d'IA n'est plus un débat

5. M. Gurman, « Samsung Bans ChatGPT, Google Bard, Other Generative AI Use by Staff After Leak », *Bloomberg*, 2 mai 2023.

6. DINUM, *Lancement de l'expérimentation Mistral AI dans l'Assistant IA interministériel*, programme ALLiaNCE, 22 octobre 2025 ; v. aussi communiqué du ministère de la Fonction publique, 22 octobre 2025.

7. Ministère des Armées et des Anciens combattants, *Le ministère des Armées notifie un accord-cadre à Mistral AI pour renforcer la souveraineté technologique de la défense*, communiqué, 8 janvier 2026 (accord-cadre notifié le 16 décembre 2025).

théorique mais une orientation stratégique et que l'usage non encadré de ChatGPT dans les services est désormais considéré comme un risque identifié, mais à mon sens, un risque encore sous-estimé.

Mesures minimales pour le professionnel du droit

Désactiver l'option d'entraînement sur vos données ; privilégier les offres professionnelles, API ou solutions souveraines hébergées en France qui garantissent contractuellement la non-utilisation des données ; et ne jamais soumettre à un LLM, quelle que soit la configuration, des éléments couverts par le secret professionnel ou des données permettant l'identification d'une personne mise en cause ou d'une victime. La prudence commande de considérer que toute information soumise à un LLM externe échappe, de manière irréversible, au contrôle exclusif du professionnel.

B. Anonymisation : une protection nécessaire mais insuffisante

Il convient d'anonymiser systématiquement les noms des parties, les lieux des faits et tout élément identifiant avant de soumettre une quelconque question relative à un dossier. Bien que la Cour de cassation ne se soit pas prononcée sur ce cas de figure précis — elle a en revanche publié en avril 2025 un rapport intitulé *Préparer la Cour de cassation de demain — Cour de cassation et intelligence artificielle*, préconisant une approche « méthodologique, éthique et pragmatique » de ces outils[8] — une fuite de données résultant d'un partage avec un LLM pourrait être assimilée à une violation du secret de l'enquête ou de l'instruction au sens de l'article 11 du Code de procédure pénale, et engager la responsabilité du professionnel concerné.

Toutefois, l'anonymisation seule peut se révéler insuffisante, et ce risque est aujourd'hui mieux documenté qu'il y a deux ans. La CNIL évalue la robustesse d'une anonymisation selon trois critères : l'impossibilité d'individualiser une personne dans le jeu de données, l'impossibilité de corréler des ensembles de données distincts la concernant, et l'impossibilité d'inférer des informations à son sujet.

Or les capacités de déduction des derniers modèles de langage rendent le troisième critère particulièrement fragile : un recoupement d'informations, même anonymisées, peut permettre d'identifier un dossier, notamment s'il a fait l'objet d'une couverture médiatique

ou s'il présente des caractéristiques factuelles singulières. Une étude publiée dans *Nature Communications* par des chercheurs de l'UCLouvain et d'Imperial College London a démontré que 99,98 % des Américains pouvaient être correctement réidentifiés dans n'importe quelle base de données anonymisée à partir de seulement quinze attributs démographiques, avec des résultats comparables au niveau mondial[9]. Pour le dire plus simplement, les données partagées par un LLM ont les mêmes chances de vous identifier que votre ADN.

Recommandation stricte

Ne jamais évoquer un dossier spécifique dans sa globalité. L'usage de l'IA doit se limiter à l'interrogation de points de droit précis, de mécanismes procéduraux ou de questions factuelles entièrement décontextualisées. La machine ne doit jamais disposer d'un tableau d'ensemble à partir duquel elle pourrait reconstituer, par inférence, la nature ou l'identité d'une affaire.

C. La technique de la légende

Une mesure complémentaire consiste à créer un *persona* fictif mais fonctionnel, une « légende », pour emprunter au vocabulaire du renseignement, afin de bénéficier des avantages de la personnalisation du modèle sans exposer son identité réelle ni la nature précise de ses activités.

Cette précaution est devenue plus importante qu'elle ne l'était au début de l'usage des LLM. Les systèmes de mémoire des LLM se sont considérablement développés. ChatGPT, depuis avril 2025, ne se contente plus d'enregistrer les points de mémoire explicitement demandés : il référence désormais l'ensemble des conversations passées pour personnaliser ses réponses. En d'autres termes, tout ce que vous écrivez — y compris incidemment — peut alimenter le profil que le modèle construit de vous. Claude (Anthropic) fonctionne différemment, avec un système de Projets isolés disposant chacun de leur propre mémoire et de leurs instructions personnalisées. Mistral, via Le Chat, offre des fonctionnalités similaires en développement.

8. Cour de cassation, *Cour de cassation et intelligence artificielle : préparer la Cour de demain*, rapport du groupe de travail, 28 avril 2025.

9. L. Rocher, J. M. Hendrickx & Y.-A. de Montjoye, « Estimating the Success of Re-identifications in Incomplete Datasets Using Generative Models », *Nature Communications*, vol. 10, n° 3069, 2019.

La mise en œuvre de la légende passe par deux mécanismes distincts qu'il convient de ne pas confondre. Le premier est celui des *instructions personnalisées* (*Custom Instructions*), accessibles dans les paramètres de chaque plateforme : c'est là que vous définissez le rôle que le modèle doit adopter et les informations de contexte qu'il doit connaître. Le second est la *mémoire conversationnelle*, qui se construit au fil des échanges. La légende doit couvrir les deux : des instructions initiales cohérentes avec le persona fictif, et une vigilance constante à ne pas laisser filtrer, au détour d'une conversation, des éléments d'identification réels que le modèle enregistrerait automatiquement. Notez que le plus efficace reste de supprimer régulièrement les données présentes sur votre IA.

Mise en œuvre concrète

Dans les instructions personnalisées, décrivez le rôle et le cadre de travail que vous souhaitez, par exemple : « *Tu es mon assistant juridique. Tu t'adresseras à moi de manière formelle. Tu sais que je suis spécialisé en droit pénal et que mes recherches portent principalement sur la jurisprudence de la Cour de cassation.* » Ne mentionnez ni votre nom, ni votre cabinet, ni votre juridiction. Vérifiez régulièrement les points de mémoire enregistrés par le modèle et supprimez tout élément identifiant qui aurait été capté involontairement.

D. Règles impératives pour l'usage professionnel

De l'ensemble de ces considérations, je dégage les règles suivantes, que je considère comme impératives. Ce ne sont plus seulement des recommandations de prudence : depuis le Livre blanc du barreau de Paris publié en octobre 2025, elles s'inscrivent dans un cadre déontologique formalisé[10]. Le principe de prudence de l'article 1.3 du Règlement intérieur national (RIN) de la profession d'avocat impose à l'avocat qui recourt à un système d'intelligence artificielle de « nécessairement vérifier la fiabilité des résultats obtenus »[11]. L'autocontrôle par l'IA elle-même est explicitement exclu : la vérification doit être humaine, personnelle et effective.

10. Barreau de Paris, *Livre blanc sur l'Intelligence Artificielle : un an d'innovation au barreau de Paris*, octobre 2025.
11. Règlement intérieur national de la profession d'avocat (RIN), art. 1.3 (CNB, 12 juill. 2007, mod. 18 mai 2019), tel qu'interprété par le guide d'utilisation de l'IA, annexe 1 du Livre blanc précité (note 9).

Ne jamais introduire de dossier dans un LLM. La seule exception envisageable concerne les solutions déployées au sein d'un cabinet ou d'une juridiction, hébergées sur des serveurs maîtrisés et soumises à des engagements contractuels de confidentialité. Les LLM grand public ne répondent à aucune de ces conditions.

Ne jamais communiquer de stratégie de défense. Le risque n'est pas seulement celui de la fuite : c'est aussi celui de la traçabilité. Une requête formulée dans un LLM peut théoriquement être reconstituée, y compris longtemps après, dès lors que l'historique des conversations n'a pas été supprimé. Comme je le mentionnais plus haut, depuis avril 2025, les systèmes de mémoire de ChatGPT référencent l'ensemble des conversations passées, ce qui aggrave considérablement ce risque pour tout utilisateur n'ayant pas désactivé cette fonctionnalité.

Exiger les sources et les vérifier systématiquement. Aucune réponse d'un LLM ne devrait être tenue pour acquise sans vérification indépendante. Les cas de jurisprudence intégralement fabriquée ne sont plus des anecdotes isolées : fin décembre 2025, le tribunal administratif d'Orléans, dans une affaire de reconduite à la frontière, a constaté que plusieurs des jurisprudences citées dans les écritures d'un avocat « n'exist[ai]ent pas, soit qu'aucune décision juridictionnelle n'existe avec le numéro indiqué, soit que les numéros de ces affaires ne correspond[ai]ent pas aux dates y accolées », invitant ce dernier à « vérifier que les références trouvées par quelque moyen que ce soit ne constituent pas une hallucination »[12].

Le tribunal administratif de Grenoble, dans une ordonnance du 3 décembre 2025 (n° 2509827), a quant à lui directement identifié dans la requête d'un justiciable non représenté, un particulier contestant une amende pour dépôt sauvage, la marque d'une rédaction par IA générative « totalement inadaptée à cet usage », assortie de « références jurisprudentielles fantaisistes »[13].

Le 9 décembre 2025, le même tribunal a rendu une seconde ordonnance (n° 2512468) relevant cette fois qu'un justiciable avait soumis via Télérecours « une requête et des mémoires générés avec un outil dit d'intelligence artificielle, dont le contenu est tout sauf "juridiquement cadré" »[14].

12. TA Orléans, 29 déc. 2025, n° 2506461 ; rapporté *in* P.-H. Levivier, « Les hallucinations d'intelligence artificielle devant les juridictions françaises », *Village de la Justice*, févr. 2026.

13. TA Grenoble, ord., 3 déc. 2025, n° 2509827.

14. TA Grenoble, ord., 9 déc. 2025, n° 2512468.

La vérification s'impose pour les références jurisprudentielles, mais également pour les raisonnements eux-mêmes, qui peuvent reposer sur des prémisses erronées ou sur une confusion entre ordres juridiques.

Ne jamais copier-coller une réponse, même après relecture. La relecture crée une illusion de contrôle. Un texte produit par un LLM et relu par son commanditaire bénéficie d'un double biais de confirmation : l'utilisateur, ayant formulé la requête, tend naturellement à retrouver dans la réponse ce qu'il espérait y trouver. Le texte doit être réécrit, reformulé, approprié, faute de quoi il n'est pas le travail du juriste mais celui de la machine.

Si vous vous trompez, vous êtes responsable. Il n'existe aucun régime de responsabilité qui permette de transférer sur l'outil informatique les conséquences d'une erreur professionnelle. Le serment de l'avocat, les obligations du magistrat, les exigences déontologiques du barreau ne connaissent pas d'exception technologique. Et la barre de l'exigence s'élève : ce qui pouvait être perçu comme une erreur excusable en 2023, lorsque les hallucinations des LLM étaient mal connues, ne l'est plus aujourd'hui.

Affaire Ko v. Li : quand le mensonge aggrave l'erreur

Au Canada, l'affaire *Ko v. Li* (2025 ONSC 2766) illustre l'escalade possible. En mai 2025, une avocate de Toronto a été sommée par le juge Myers de la Cour supérieure de l'Ontario de justifier pourquoi elle ne devrait pas être citée pour outrage, après la découverte dans son *factum* de jurisprudences inexistantes générées par ChatGPT, certaines citations renvoyant à des décisions réelles dont le contenu avait été inversé[15]. L'avocate a d'abord nié avoir utilisé l'IA, avant d'admettre en septembre 2025 avoir menti au tribunal. En octobre 2025, la procédure pour outrage criminel (*criminal contempt*) a été renvoyée au procureur général de l'Ontario. Une ordonnance du 4 décembre 2025 a formalisé la poursuite, le ministère public qualifiant son comportement d'« indifférence apparentée à l'imprudence »[16].

Le juge Myers a souligné le caractère potentiellement jurisprudentiel de cette affaire : « *Je n'ai connaissance d'aucun cas dans lequel un avocat, tenu à des devoirs de franchise et d'honneur, a admis avoir délibérément induit un tribunal en erreur dans une procédure d'outrage la concernant elle-même.* »

15. *Ko v. Li*, 2025 ONSC 2766 (Ont. Sup. Ct. J.), juge F. Myers, 6 mai 2025.
16. *Ko v. Li*, 2025 ONSC 6785 (Ont. Sup. Ct. J.), juge F. Myers, 4 décembre 2025 ; v. aussi « Toronto Lawyer Faces Criminal Contempt Proceedings After Admitting to Misleading Court About AI Use », *Law Times*, 11 décembre 2025.

Principe cardinal

L'IA est un instrument ; la responsabilité demeure intégralement humaine et donc la vôtre.

III.
Structurer ses demandes : les fondements du prompt juridique

Comprendre l'outil et en connaître les limites ne dispense pas de savoir l'interroger. Cette troisième partie expose les conditions d'un recours pertinent à l'IA (A), les quatre piliers structurants de tout prompt juridique (B), les techniques fondamentales du prompt engineering (C), leur application pratique à travers des exemples tirés du droit pénal (D), les erreurs de formulation les plus courantes (E), et enfin la puissance du méta-prompt comme outil de capitalisation (F).

A. Quand utiliser l'IA, et pourquoi

Avant même de s'interroger sur la formulation d'un prompt, la question préalable — et trop souvent éludée — est celle de la pertinence du recours à l'IA. J'ai exposé dans la section précédente les risques d'un usage systématique. Il convient à présent de délimiter positivement les cas où cet usage est légitime.

Trois situations justifient, à mon sens, le recours à un LLM dans un cadre juridique.

La première est la **contrainte temporelle**. Lorsqu'une échéance ne permet pas de conduire soi-même une recherche approfondie, le LLM peut servir de défricheur à condition que ses résultats soient ensuite vérifiés par les voies classiques.

La deuxième est la **complexité structurelle**. Certaines questions mobilisent simultanément plusieurs branches du droit, des jurisprudences contradictoires ou des sources normatives hétérogènes. Le LLM excelle dans ce type de mise en perspective, non pas parce que ses réponses sont nécessairement exactes, mais parce qu'il peut identifier des connexions entre des corpus que le juriste n'aurait pas spontanément rapprochés.

La troisième est la **fonction de contradiction**. Paradoxalement, l'un des usages les plus productifs de l'IA est de l'utiliser *contre soi-même* : lui soumettre sa propre argumentation et lui demander d'en identifier les failles, les objections prévisibles, les angles morts. C'est un usage où le biais d'acquiescement peut être neutralisé par des instructions explicites, et où le modèle, n'ayant pas de thèse personnelle à défendre, peut jouer le rôle d'un contradicteur méthodique.

En dehors de ces trois cas de figure, la question mérite toujours d'être posée : ai-je réellement besoin de l'IA, ou suis-je en train de céder à la facilité ?

B. Les quatre piliers du prompt juridique

Un prompt juridique efficace repose sur quatre éléments que je qualifierai de structurants : le **contexte**, l'**objectif**, les **contraintes** et le **format**.

1. Le contexte répond à la question : qui pose la question, et dans quel cadre ? Un LLM ne déduit pas spontanément que son interlocuteur est un magistrat spécialisé en droit des affaires ou un élève-avocat préparant le CRFPA. Or, cette information conditionne directement le registre, la profondeur et l'orientation de la réponse. Définir le contexte en amont — idéalement dans les instructions permanentes du modèle — permet d'économiser des tokens à chaque interaction et d'obtenir des réponses immédiatement calibrées.

2. L'objectif désigne la nature précise du livrable attendu. « Parle-moi du droit de la garde à vue » et « produis une analyse structurée des évolutions jurisprudentielles relatives au droit d'accès à l'avocat en garde à vue depuis la réforme de 2011, en distinguant les positions de la Cour de cassation et de la Cour européenne des droits de l'homme » ne sollicitent pas les mêmes aires statistiques du modèle. La seconde formulation, par sa précision, contraint le LLM à mobiliser des données pertinentes et réduit mécaniquement le risque d'hallucination. Je résume cette exigence par trois mots : *dynamique, didactique, dialectique*. La requête doit orienter le modèle vers une réponse qui progresse (dynamique), qui explique (didactique) et qui confronte les positions (dialectique).

3. Les contraintes bornent le périmètre de la réponse. En matière juridique, la contrainte la plus importante est d'ordre géographique et normatif : *droit français uniquement*. Sans cette précision, le modèle — entraîné majoritairement sur des données anglophones — aura une tendance structurelle à glisser vers le droit américain, le droit anglais ou des principes de *common law* qui n'ont aucune pertinence dans l'ordre juridique français. Il convient également de préciser la base normative attendue (code, loi, jurisprudence de telle juridiction), la période temporelle pertinente, et toute exclusion nécessaire.

4. Le format détermine la forme de la réponse. Un plan détaillé, une fiche de jurisprudence, une note de synthèse, un tableau comparatif n'appellent pas la même structuration de la part du modèle. Préciser le format attendu n'est pas un détail cosmétique : c'est un paramètre qui modifie la manière dont le LLM organise l'information et, par conséquent, la qualité du raisonnement sous-jacent.

C. Les techniques fondamentales du prompt engineering

L'ingénierie de prompt repose sur un ensemble de techniques dont la littérature spécialisée a stabilisé la terminologie. Il n'est pas nécessaire de les maîtriser toutes pour un usage professionnel efficace, mais il est indispensable d'en connaître les principales, ne serait-ce que pour comprendre pourquoi certaines formulations produisent des résultats radicalement supérieurs à d'autres. Je les présente ici dans un ordre qui n'est pas alphabétique mais pratique : du plus immédiatement utile au plus avancé.

Le persona (*role prompting*)

La technique consiste à assigner explicitement un rôle au modèle avant de lui soumettre une requête. « Tu es un avocat pénaliste spécialisé en droit de la presse » ne relève pas de la fiction : c'est une instruction qui reconfigure la distribution statistique des réponses. Le modèle, orienté vers un rôle précis, puise en priorité dans les données d'entraînement associées à ce champ d'expertise. L'effet est comparable à celui d'une consultation : on ne pose pas la même question de la même manière à un généraliste et à un spécialiste, et on n'obtient pas la même réponse.

Le persona ne se limite pas à l'expertise. Il peut inclure un niveau (« tu t'adresses à un praticien confirmé, pas à un étudiant »), un tempérament (« sois critique et signale les faiblesses de mon raisonnement ») ou une posture institutionnelle (« tu es un conseiller référendaire à la Cour de cassation qui rédige un rapport sur cette question »).

Le cadrage par l'exemple (*few-shot prompting*)

Le principe est simple : plutôt que de décrire abstraitement le format de réponse attendu, on en fournit un ou plusieurs exemples concrets dans le prompt lui-même. Le modèle, par construction statistique, reproduira la structure, le niveau de détail et le registre des exemples fournis.

Pour un juriste, cette technique est particulièrement efficace pour obtenir des fiches de jurisprudence homogènes. Plutôt que de demander « fais-moi une fiche de jurisprudence », on intègre au prompt un modèle :

> **Gabarit de fiche jurisprudentielle**
>
> **Référence :** [juridiction, date, n° de pourvoi]
> **Faits pertinents :** [3 lignes maximum]
> **Question de droit :** [formulation précise]
> **Solution :** [sens de la décision + motivation essentielle]
> **Portée :** [confirmation, revirement, précision de jurisprudence]
>
> « Reproduis ce format pour chaque arrêt identifié. »

À l'inverse, le *zero-shot prompting* désigne une requête sans exemple fourni, le modèle s'appuie uniquement sur ses instructions et ses données d'entraînement. C'est le mode par défaut de la plupart des utilisateurs. Il fonctionne pour les requêtes simples ; il devient insuffisant dès que le format attendu est spécifique.

Le raisonnement guidé (*chain-of-thought prompting*)

Cette technique consiste à demander explicitement au modèle de raisonner étape par étape plutôt que de livrer directement une conclusion. L'instruction peut être aussi simple que « raisonne étape par étape avant de conclure » ou « expose ton raisonnement avant de donner ta réponse ».

L'effet est documenté par la recherche en IA : les modèles qui « montrent leur travail » produisent des réponses significativement plus fiables sur les tâches de raisonnement complexe.

Pour le juriste, l'intérêt est double. D'une part, un raisonnement explicite est un raisonnement auditable, chaque étape peut être vérifiée indépendamment, ce qui rend les erreurs visibles là où une réponse directe les aurait dissimulées. D'autre part, cette

technique est l'exact équivalent de ce que tout professeur de droit enseigne dès la première année : le syllogisme juridique. Majeure (la règle), mineure (les faits), conclusion (la solution). Demander à un LLM de raisonner en syllogisme, c'est lui imposer la discipline que le droit impose au juriste.

Les instructions négatives (*negative prompting*)

Dire au modèle ce qu'on ne veut pas est souvent aussi important que de lui dire ce qu'on veut. Cette technique est particulièrement utile pour prévenir les dérapages récurrents des LLM en matière juridique :

Instructions négatives types

« Ne cite aucune jurisprudence dont tu n'es pas certain de l'existence. N'invente pas de numéro de pourvoi. Ne mélange pas droit français et droit comparé sans l'indiquer expressément. N'utilise pas de formulations conclusives ("il est clair que", "sans aucun doute") lorsque la question est débattue. »

Les balises et délimiteurs (*structured prompting*)

Lorsqu'un prompt intègre plusieurs types d'information — des faits, des instructions, un format, des exemples —, le modèle peut les confondre. Les balises permettent de segmenter visuellement et logiquement les composantes du prompt :

Exemple de prompt structuré par balises

[CONTEXTE] Tu es un magistrat du siège spécialisé en droit pénal des affaires.
[FAITS] [description décontextualisée]
[QUESTION] La prescription de l'action publique est-elle acquise ?
[CONTRAINTES] Droit français uniquement. Jurisprudence de la chambre criminelle postérieure à 2017. Cite tes sources.
[FORMAT] Note de synthèse en trois parties : I. Règle applicable, II. Application aux faits, III. Conclusion et réserves.

La température et la créativité

Un dernier paramètre, rarement accessible dans les interfaces grand public mais qu'il faut connaître pour comprendre le comportement des modèles : la *température*. Ce réglage, exprimé par un chiffre généralement compris entre 0 et 1, détermine le degré de « créativité » du modèle, c'est-à-dire sa propension à s'écarter de la réponse statistiquement la plus probable.

Une température basse (proche de 0) produit des réponses prévisibles, répétitives, collées au plus probable. Une température élevée (proche de 1) autorise des associations plus libres, plus originales, mais aussi plus risquées. Pour un usage juridique, où la fiabilité prime sur l'originalité, une température basse est presque toujours préférable. Si l'interface que vous utilisez permet ce réglage, fixez-le aussi bas que le permet la tâche. Pour la recherche jurisprudentielle : température minimale. Pour un brainstorming d'arguments : vous pouvez vous autoriser un cran au-dessus.

Ce que la température illustre, au fond, c'est un principe général : un LLM n'est pas un outil à usage unique. C'est un instrument paramétrable, dont le comportement se configure, et dont le juriste doit devenir le régleur, pas le spectateur.

D. En pratique : du mauvais prompt au bon

Pour être bon, un prompt n'a pas besoin d'être long. Il se doit d'être efficace. Les exemples qui suivent, tous tirés du droit pénal, illustrent la transformation d'une requête vague en un prompt opérationnel.

1. La recherche jurisprudentielle

> **Prompt brut**
>
> *« Quelle est la jurisprudence sur la légitime défense ? »*

Ce prompt cumule trois erreurs : l'abstraction sans ancrage concret, l'absence de bornage temporel, et l'absence de contrainte géographique.

✓ Prompt restructuré

« En droit pénal français, analyse l'évolution jurisprudentielle de la légitime défense (article 122-5 du Code pénal) devant la chambre criminelle de la Cour de cassation entre 2018 et 2025. Distingue les conditions de proportionnalité et de simultanéité de la riposte. Pour chaque arrêt cité, indique le numéro de pourvoi, la date et la solution retenue. Signale explicitement toute incertitude sur l'existence d'une décision. Format : note de synthèse structurée. »

Quatre éléments présents : le contexte (droit pénal français, chambre criminelle), l'objectif (évolution jurisprudentielle sur deux critères précis), les contraintes (période, juridiction, obligation de sourcer et de signaler les incertitudes) et le format (note de synthèse). L'instruction de signalement des incertitudes est une parade directe contre le risque d'hallucination.

2. L'analyse de qualification pénale

✗ Prompt brut

« Est-ce que c'est du harcèlement moral ? Mon client se fait insulter par son patron tous les jours. »

Ce prompt pose une question fermée qui active le biais d'acquiescement, et il invite le LLM à qualifier juridiquement des faits. Sur le plan déontologique, il ouvre la porte à la divulgation d'informations confidentielles.

✓ Prompt restructuré

« En droit pénal français, expose les éléments constitutifs du délit de harcèlement moral au sens de l'article 222-33-2 du Code pénal. Distingue l'élément matériel (nature et répétition des agissements) et l'élément intentionnel. Précise les critères retenus par la chambre criminelle pour caractériser la répétition, en citant les arrêts de référence avec numéro de pourvoi. Indique enfin les infractions voisines avec lesquelles une confusion est fréquente et les critères de distinction. Droit français uniquement. Format : fiche d'analyse structurée par élément constitutif. »

Le prompt ne mentionne aucun fait. Il ne demande pas au modèle de qualifier mais d'exposer les critères de qualification, et la différence est fondamentale.

3. Le prompt contradictoire

Prompt brut

« Mon argumentation sur la nullité de la garde à vue est-elle solide ? »

✔ Prompt restructuré

« Tu es un magistrat du parquet expérimenté, spécialisé en procédure pénale. Un avocat soulève la nullité d'une garde à vue au motif que la notification du droit à l'assistance d'un avocat a été effectuée trois heures après le placement, le procès-verbal mentionnant un "retard dû à des circonstances opérationnelles" sans autre précision. Ta mission : démonte cette argumentation. Identifie toutes les faiblesses juridiques de ce moyen de nullité, les contre-arguments que le parquet pourrait opposer, et les jurisprudences de la chambre criminelle qui pourraient faire échec à cette demande. Sois impitoyable. Droit français, jurisprudence postérieure à 2015 uniquement. »

Trois mécanismes sont à l'œuvre. L'assignation de rôle inverse la polarité du modèle. Les faits soumis sont entièrement décontextualisés. L'instruction « sois impitoyable » est un *override* explicite du biais d'acquiescement.

4. La veille normative ciblée

Prompt brut

« Quoi de neuf en procédure pénale ? »

✔ Prompt restructuré

« Identifie les modifications législatives et réglementaires entrées en vigueur en droit de la procédure pénale française entre le 1er janvier 2025 et aujourd'hui. Pour chaque modification, indique : le texte source (loi, décret, ordonnance), sa date de publication au Journal officiel, les articles du Code de procédure pénale modifiés, et une synthèse en trois lignes de la portée pratique du changement. Si tu n'es pas certain qu'une modification est effectivement entrée en vigueur, indique-le explicitement plutôt que de l'affirmer. Format : tableau chronologique. »

Le fil rouge de ces quatre exemples est le même : le prompt efficace ne demande jamais au modèle de penser à la place du juriste. Il lui demande de préparer le terrain sur lequel le juriste exercera son propre jugement. La nuance est celle qui sépare le juriste augmenté du juriste remplacé.

Le réflexe méthodique du praticien malin

Il est assez évident que les techniques de prompting peuvent être longues et rébarbatives. Toutefois, il existe des raccourcis permis notamment par les dernières avancées dans les capacités de raisonnement des LLM. L'une des techniques récentes va être de demander à l'IA d'aller d'abord regarder la méthodologie de ce que vous voulez produire, et ensuite de lui demander le résultat. Par exemple : vous souhaitez obtenir une fiche d'arrêt sur un arrêt spécifique de la Cour de cassation. La méthodologie de la fiche d'arrêt est l'un des exercices les plus codifiés et documentés de la formation juridique : c'est une ressource que l'IA n'aura aucun mal à retrouver une fois sur internet. Ainsi, sans aucune balise, la technique consiste tout simplement à faire un prompt très classique : « *Va regarder la méthodologie de la fiche d'arrêt et ensuite fais-moi l'analyse de l'arrêt Cour de Cassation, Assemblée plénière, du 29 octobre 2004, 03-11.238.* »

E. Les erreurs de formulation les plus courantes

Plusieurs erreurs récurrentes méritent d'être signalées, car elles expliquent l'essentiel des déconvenues rapportées par les utilisateurs professionnels.

La confusion entre date et durée. Demander à un LLM « quelle est la jurisprudence récente sur tel sujet » est une requête ambiguë. « Récent » n'a pas la même signification selon les domaines et les juridictions. Il convient de préciser une période explicite : « entre 2020 et 2025 », « depuis l'entrée en vigueur de la loi du… ».

L'abstrait sans le concret. Les LLM peinent à traiter des requêtes formulées en termes exclusivement abstraits. « Parle-moi de la responsabilité » est un prompt inutilisable. « Analyse les conditions d'engagement de la responsabilité civile délictuelle du fait personnel en droit français, en distinguant les régimes de la faute prouvée et de la faute présumée, avec références aux articles 1240 et 1241 du Code civil » est un prompt opérationnel. La règle est simple : plus la requête est concrète et bornée, plus la réponse sera exploitable.

L'usage de termes subjectifs. Les termes évaluatifs — « bon », « mauvais », « meilleur », « original » — n'ont pas de signification stable pour un LLM. Ils orientent la réponse sans la fonder. Leur usage doit être proscrit au profit d'un vocabulaire descriptif et neutre.

Le verbe « stipuler ». Ce point peut sembler anecdotique, mais il est révélateur. En droit français, « stipuler » s'emploie exclusivement en matière contractuelle. Or, de nombreux utilisateurs — y compris des juristes — écrivent « la loi stipule que… ». Le LLM, fidèle à sa logique statistique, reproduira cette impropriété sans la corriger et construira sa réponse sur un registre potentiellement inadapté. La rigueur terminologique du prompt conditionne la rigueur terminologique de la réponse.

F. La puissance du méta-prompt

Pour conclure cette partie, il convient de revenir sur la notion de *méta-prompt* — ou *prompt système* — que j'ai brièvement évoquée en introduction. Le méta-prompt est un ensemble d'instructions permanentes qui conditionnent le comportement du modèle avant même qu'il ne reçoive une requête spécifique. C'est, en quelque sorte, le cahier des charges du LLM.

Dans un usage juridique, le méta-prompt permet de résoudre en amont la plupart des difficultés exposées dans ce guide. Il peut intégrer le contexte professionnel de l'utilisateur, les contraintes normatives par défaut (droit français, jurisprudence de la Cour de cassation), le format de réponse attendu, l'exigence de citation des sources, l'obligation de signaler les incertitudes, et l'instruction explicite de ne pas acquiescer sans fondement.

Investissement décisif

Un méta-prompt bien conçu transforme chaque interaction ponctuelle en un échange déjà calibré. Il constitue, à mon sens, l'investissement le plus rentable qu'un professionnel du droit puisse faire dans sa maîtrise de l'intelligence artificielle, car il capitalise l'ensemble des enseignements de ce guide en un protocole réutilisable. Faire appel à un professionnel dans ce domaine peut se révéler un investissement décisif.

Conclusion : vers le juriste augmenté

L'intelligence artificielle ne remplacera probablement pas le juriste, du moins aussi longtemps que le juriste sera en mesure de la maîtriser et de lui imposer la rigueur : la pensée, le doute et la vérification.

L'IA est là et elle ne va pas s'en aller. Il est essentiel pour les professionnels du droit de ne pas s'enfermer dans une démarche réfractaire pour en comprendre les mécanismes, non seulement pour en faire usage, mais pour être en mesure de répondre à la question fondamentale que la société leur posera de plus en plus frontalement et certainement bien plus vite qu'on ne le pense : *où voulons-nous qu'il y ait de l'IA, et où doit-elle impérativement être exclue ?*

Le professionnel qui demeure réfractaire s'expose à un risque asymétrique : les autres — et c'est déjà le cas — s'emparent de l'outil avant lui et créent une distance qui pourrait bien être irréversible. Et lorsque la conscience de ce retard s'imposera, il ne lui restera que la résignation : faire usage d'outils métier dont il n'aura qu'une connaissance aveugle.

Le juriste augmenté n'est ni celui qui refuse l'IA, ni celui qui s'y abandonne. C'est celui qui comprend ses mécanismes, en connaît les biais, en maîtrise les limites, et conserve à chaque instant la pleine responsabilité de son raisonnement. C'est celui qui connaît la puissance de ce qu'il a entre les mains et décide d'en faire un usage raisonné. L'outil change ; l'exigence demeure.

Ce guide n'a d'autre ambition que de poser les fondations de cette maîtrise. Les techniques évoluent vite ; les principes de rigueur intellectuelle qui doivent les encadrer, eux, ne changent pas. L'IA crée un monde où l'excellence d'hier est la nouvelle médiocrité. Il appartient aux juristes de redéfinir eux-mêmes les critères de l'excellence dans leur domaine, faute de quoi la machine le fera à leur place.

Guide mis à jour en mars 2026.

* * *

Bibliographie

Ouvrages et rapports institutionnels

S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, PublicAffairs, 2019 (trad. fr. : *L'Âge du capitalisme de surveillance*, Zulma, 2020).

H. Arendt, *The Human Condition*, University of Chicago Press, 1958 (trad. fr. : *Condition de l'homme moderne*, Calmann-Lévy, 1961).

Barreau de Paris, *Livre blanc sur l'Intelligence Artificielle : un an d'innovation au barreau de Paris*, octobre 2025.

Cour de cassation, *Cour de cassation et intelligence artificielle : préparer la Cour de demain*, rapport du groupe de travail, 28 avril 2025.

Paris Place de Droit, *Le droit aux défis de l'IA générative*, Livre blanc, septembre 2025.

CNIL, *Avis sur les techniques d'anonymisation*, Groupe Article 29 / CEPD, avis 05/2014 du 10 avril 2014.

Articles scientifiques

M. Sharma, M. Tong, T. Korbak *et al.*, « Towards Understanding Sycophancy in Language Models », arXiv:2310.13548, 2023 (ICLR 2024).

L. Rocher, J. M. Hendrickx, Y.-A. de Montjoye, « Estimating the success of re-identifications in incomplete datasets using generative models », *Nature Communications*, vol. 10, n° 3069, 2019.

E. L. Thorndike, « A constant error in psychological ratings », *Journal of Applied Psychology*, vol. 4, n° 1, 1920, pp. 25-29.

E. F. Loftus, « Leading questions and the eyewitness report », *Cognitive Psychology*, vol. 7, n° 4, 1975, pp. 560-572.

M. Sherif, *A Study of Some Social Factors in Perception*, 187 Archives of Psychology 1, 1935.

R. B. Zajonc, « Attitudinal Effects of Mere Exposure », *Journal of Personality and Social Psychology*, vol. 9, n° 2 (Monograph Supplement), 1968, pp. 1-27.

E. T. Higgins & W. S. Rholes, « "Saying is Believing": Effects of Message Modification on Memory and Liking for the Person Described », *Journal of Experimental Social Psychology*, vol. 14, n° 4, 1978, pp. 363-378.

T. L. Huon, « User Imprint: Psychological Profiling and Qualified Information in Prolonged Interaction with Large Language Models », SSRN Working Paper, 21 mars 2026. DOI : 10.2139/ssrn.6452038

Doctrine et articles professionnels

P.-H. Levivier, « Les hallucinations d'intelligence artificielle devant les juridictions françaises : premiers cas et implications déontologiques pour les avocats », *Village de la Justice*, févr. 2026.

Jurisprudence

Mata v. Avianca, Inc., 678 F. Supp. 3d 443 (S.D.N.Y. 2023), juge P. Kevin Castel.

Ko v. Li, 2025 ONSC 2766 (Ont. Sup. Ct. J.), juge F. Myers, 6 mai 2025.

Ko v. Li, 2025 ONSC 6785 (Ont. Sup. Ct. J.), juge F. Myers, 4 décembre 2025.

Tribunal administratif de Grenoble, ordonnance du 3 décembre 2025, n° 2509827.

Tribunal administratif de Grenoble, ordonnance du 9 décembre 2025, n° 2512468.

Tribunal administratif d'Orléans, 29 décembre 2025, n° 2506461.

Sources normatives et réglementaires

Règlement (UE) 2024/1689 du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle (« AI Act »).

Règlement intérieur national de la profession d'avocat (RIN), notamment article 1.3.

Code de procédure pénale, article 11 (secret de l'enquête et de l'instruction).

Code civil, articles 1240 et 1241 (responsabilité civile délictuelle).