DOCUMENT DE TRAVAIL

Recherche jurisprudentielle fondée sur les citations, pour juristes

Document de travail sur un outil technique de recherche jurisprudentielle

Heuris, du grec *heuriskein*, « trouver, découvrir », est le nom de l'outil de recherche de jurisprudence étudié dans ce document.

Recherche appliquée · novembre 2026

Thèse. En droit, la méfiance envers l'IA est structurelle : ses résultats sont opaques, peuvent changer d'une exécution à l'autre et peuvent citer des autorités qui n'existent pas. Une intégration généraliste hérite de ces propriétés ; aucun juriste ne peut s'y fier devant un juge. La confiance est donc une question d'architecture, conçue pour les contraintes de la recherche juridique :

1. Chaque phrase affichée repose sur un passage source conservé à des positions de caractères exactes ; l'IA rédige, la source tranche.
2. Un vérificateur supprime toute phrase dont la citation est introuvable ou n'étaye pas l'affirmation, de sorte que le juriste n'a pas à se demander si une citation a été hallucinée.
3. Le système échoue en mode *fail-closed* (sûr par défaut : en cas de doute, le système refuse au lieu de compléter) : lorsque rien ne survit à la vérification, il le dit au lieu d'inventer une réponse.
4. Le dossier du client est traité comme confidentiel dès la conception : il est analysé en mémoire et la recherche s'exécute sur la machine locale ; l'appel au modèle de langage est le seul point où du texte la quitte, et aucun dossier client réel n'est traité avant la mise en place d'un modèle hébergé dans l'UE ou local.

Résumé

Les assistants génératifs rédigent une prose juridique fluide. En recherche juridique, toutefois, la difficulté ne tient pas à la fluidité mais à la vérifiabilité. Une étude préenregistrée portant sur des outils commerciaux de recherche juridique fondés sur l'IA a constaté que les systèmes à *retrieval-augmented generation* (génération conditionnée par des passages extraits d'un corpus) hallucinent moins qu'un chatbot généraliste, mais produisent encore des réponses incorrectes ou mal fondées dans 17–33 % des requêtes.[1] Le *retrieval-augmented generation*[2] change l'origine des mots du modèle. Il ne garantit pas, à lui seul, que le passage cité est le bon, qu'il est cité exactement et qu'il étaye réellement la phrase à laquelle il est rattaché.

Le problème présente un deuxième aspect, qui concerne le temps et les moyens disponibles pour la recherche. La description officielle de Juportal, la base de données publique de jurisprudence du SPF Justice, décrit une recherche par numéro ECLI, numéro de rôle et dates, avec des filtres ; elle ne mentionne aucune recherche sémantique ni en langage naturel.[16] Aucune mesure belge ou française du temps que les juristes consacrent à la recherche de jurisprudence n'a été trouvée. Les moyens de la justice belge sont en outre limités par rapport aux autres États membres de l'UE : selon la Commission européenne, « le budget consacré au système judiciaire et le nombre de juges professionnels pour 100 000 habitants restent inférieurs à la moyenne de l'UE », et le délai de traitement des affaires civiles et commerciales en première

instance était de 246 jours en 2023.[18] Pour 2022, la CEPEJ évalue ce budget à 102,5 € par habitant, au-dessus de la médiane du Conseil de l'Europe, mais à 0,22 % du PIB, en dessous de celle-ci.[17] Dans son communiqué de presse du 10 janvier 2025, la Cour des comptes constate que la transformation numérique de la justice « n'est pas menée de manière satisfaisante ».[19]

Le problème présente un troisième aspect : la confidentialité. La préparation d'un dossier fait entrer les faits du client, couverts par le secret professionnel, dans la requête et dans le matériau que l'outil traite. Un assistant généraliste les transmet à des services tiers, selon des conditions et dans des juridictions que le juriste ne peut examiner, et n'offre aucune garantie que ce secret soit préservé.

Heuris vise à réduire l'effort de recherche de jurisprudence du juriste tout en préservant la traçabilité ; ce gain de temps pour le juriste n'a pas été mesuré (voir l'exigence R8).

Heuris est un outil de recherche de jurisprudence destiné aux juristes, conçu pour toute matière juridique où des données licites sont disponibles. Sa preuve de concept porte sur un premier corpus de plus de 25 000 décisions belges en droit du travail, sous la forme de commentaires en langue française et non des décisions elles-mêmes ; d'autres matières sont prévues, selon l'accès aux données. Le problème est traité au niveau des données et de la vérification : la collecte est limitée aux sources qui l'autorisent ; chaque passage source est conservé à des positions de caractères exactes ; la génération est soumise à un vérificateur qui supprime toute phrase dont la citation ne peut être retrouvée en entier à l'emplacement cité, ou qu'un second contrôle automatisé juge ne pas étayer la phrase. Si aucune phrase ne survit à la vérification, le système l'indique. Pour la confidentialité, le dossier est analysé en mémoire et la recherche s'exécute sur la machine locale ; seul l'appel au modèle de langage la quitte, et la preuve de concept est limitée aux données publiques et à des dossiers fictifs ou anonymisés jusqu'à la mise en place d'un modèle hébergé dans l'UE ou local. Dans ce document, « vérifié » signifie qu'un extrait correspond au texte source aux positions citées ; la source étant un commentaire, un extrait vérifié ne prouve pas que la juridiction l'a dit.

Postulat. Un assistant de recherche juridique n'est fiable que si chaque phrase qu'il affiche peut être rattachée, caractère par caractère, à un passage obtenu licitement qui l'étaye — et s'il répond « rien trouvé » lorsqu'il ne le peut pas.

1. Exposé du problème

1.1 L'hallucination dans la recherche juridique générative

Dans une étude portant sur plus de 800 000 requêtes relatives à la jurisprudence fédérale américaine, des modèles de langage généralistes (GPT-4, GPT-3.5, PaLM 2 et Llama 2) ont halluciné dans au moins 58 % des cas, ont eu du mal à prédire leurs propres hallucinations et ont souvent accepté sans esprit critique les présupposés juridiques erronés des utilisateurs.[3] Les outils spécialisés fondés sur le *retrieval* (recherche des passages pertinents dans le corpus) font mieux sans résoudre le problème : dans l'étude préenregistrée citée plus haut, des outils commerciaux de recherche juridique fondés sur l'IA hallucinaient encore dans 17–33 % des cas.[1] Des travaux plus récents suggèrent que le problème est structurel. Xu et al. formalisent l'hallucination et montrent qu'« il est impossible d'éliminer l'hallucination dans les LLM » ;[4] Kalai et al. observent que les hallucinations « persistent même dans les systèmes les plus avancés » et l'attribuent en partie à des pratiques d'évaluation qui récompensent les réponses au hasard en situation d'incertitude.[5] Au vu de cette littérature, une lecture plausible est que les modèles plus récents réduisent le taux d'hallucination sans en supprimer le mécanisme. Cette lecture est une inférence tirée des travaux cités, et non un résultat mesuré. Les deux études juridiques portent sur le droit américain, et aucune mesure équivalente n'existe pour le droit belge ou français. La langue de la question n'est probablement pas le facteur principal. Les données d'entraînement sont dominées

par l'anglais : dans les données de pré-entraînement de Llama 2, l'anglais représente 89,70 % et le français 0,16 %.[6] Pour une langue bien dotée comme le français, l'écart mesuré sur un benchmark général est faible : sur une version traduite automatiquement du MMLU, GPT-4 a obtenu 83,6 % en français contre 85,5 % en anglais (3-shot).[7] La faiblesse documentée la plus importante concerne les connaissances régionales et de longue traîne : dans Global MMLU, 84,9 % des questions qui exigent des connaissances géographiques portent sur l'Amérique du Nord ou l'Europe,[8] et BLEnD rapporte que les modèles sont plus performants pour les cultures fortement représentées en ligne, avec un écart allant jusqu'à 57,34 % pour GPT-4.[9] *Inférence, et non résultat mesuré :* la jurisprudence belge est à peine représentée dans les données d'entraînement, de sorte que les modèles se souviendront probablement mal de décisions belges précises ; aucun benchmark ne le mesure.

1.2 Conséquences judiciaires

Dans *Mata v. Avianca* (S.D.N.Y., 22 juin 2023), le tribunal a sanctionné des avocats qui avaient « soumis des décisions judiciaires inexistantes, assorties de fausses citations et références créées par l'outil d'intelligence artificielle ChatGPT », en leur infligeant solidairement une amende de 5 000 dollars américains.[10] Des juridictions belges ont depuis été confrontées à la même défaillance. La RTBF a rapporté en décembre 2025 que la cour d'appel d'Anvers avait constaté des conclusions étayées par « une jurisprudence inexistante et des sources juridiques inventées de toutes pièces », et condamné l'avocat concerné à payer 25 000 € au total : une amende de 2 500 € pour abus manifeste de procédure et 7 500 € à chacune des trois parties. Le même reportage indique que le 15 décembre 2025, le *tribunal de l'entreprise* de Gand a examiné des conclusions citant des arrêts de la Cour de cassation et de la Cour constitutionnelle qui n'ont jamais existé.[11] En octobre 2025, l'Orde van Vlaamse Balies a fait état d'un arrêt de la cour d'appel de Gand concernant des conclusions contenant une longue liste de citations inexactes de dispositions légales existantes et des références à une jurisprudence et une doctrine inexistantes.[12] Ces décisions belges sont citées telles que rapportées par la presse et par l'Orde van Vlaamse Balies ; les décisions elles-mêmes, leurs dates et leurs numéros de rôle n'ont pas été consultés. Aucune n'a été rendue par une juridiction du travail ; elles sont citées pour le devoir de prudence de l'avocat dans l'usage de l'IA.

1.3 L'accès à la jurisprudence belge

Environ 1,4 million de décisions sont rendues chaque année en Belgique, selon le SPF Justice (2026).[13] Seule une faible part est publiée. Une estimation académique la situe à environ 1 % ;[14] aucune mesure officielle de cette part n'a été trouvée. Juportal, la base de données officielle de jurisprudence du SPF Justice, est une sélection : chaque juridiction décide des décisions qu'elle anonymise et y publie. À son lancement en 2020, les cours du travail ont fourni une sélection d'arrêts (plus de 21 000 publications).[15] Le registre central des décisions judiciaires (JustJudgment) existe, mais à la mi-2026 seules les parties concernées et leurs avocats ont accès à leurs propres décisions, et les arrêts des cours du travail devaient encore y être ajoutés.[13] Comme l'explique le § 6, le fichier robots.txt de Juportal le ferme aux agents automatisés autres que le robot d'indexation du ministère lui-même, et Heuris n'y collecte rien.

1.4 Temps et moyens de recherche

Juportal a remplacé le moteur de recherche Jure-Juridat en décembre 2020, parce que la technologie Flash sur laquelle Jure-Juridat avait été construit ne serait plus prise en charge. La description officielle la présente comme « la seule banque de données publique comportant une sélection de jugements et d'arrêts anonymisés » et décrit une recherche par numéro ECLI, numéro de rôle (*numéro de rôle*) et dates, avec des filtres par juridiction, matière, type et année de publication, ainsi que des flux RSS.[16] Cette description ne mentionne aucune recherche sémantique ni en langage naturel ; il s'agit d'un constat portant sur la description officielle, et non de la preuve qu'une telle fonction n'existe pas.

Aucune mesure belge ou française du temps que les juristes consacrent à la recherche de jurisprudence n'a été trouvée. Les chiffres disponibles pour d'autres juridictions proviennent d'enquêtes de fournisseurs ou n'ont pu être vérifiés dans leur source primaire, et ne sont pas utilisés ici.

Les chiffres comparatifs sur les moyens de la justice belge dépendent du groupe de comparaison. La CEPEJ (Conseil de l'Europe) rapporte pour 2022 un budget exécuté du système judiciaire d'environ 1,2 milliard d'euros, soit 102,5 € par habitant, « supérieur à la médiane du CdE », alors qu'en pourcentage du PIB (0,22 %) le budget est « inférieur à la médiane du CdE » ; elle ajoute que le nombre d'affaires civiles et pénales pendantes « n'est pas disponible pour la 1re et la 2e instance ».[17] Le rapport 2025 de la Commission européenne sur l'état de droit, qui compare la Belgique aux autres États membres de l'UE, indique que « le budget consacré au système judiciaire et le nombre de juges professionnels pour 100 000 habitants restent inférieurs à la moyenne de l'UE », en se référant au tableau de bord 2025 de la justice dans l'UE. Il fait état d'un délai de traitement de 246 jours en 2023 pour les affaires civiles et commerciales en première instance, observe qu'« une vue d'ensemble complète de l'efficacité de la justice reste indisponible en raison d'un manque persistant de données sur les procédures judiciaires », et cite « l'accès en ligne aux décisions publiées et aux décisions judiciaires lisibles par machine » parmi les points soulevés par les parties prenantes.[18] Les deux constats ne sont pas contradictoires : ils reposent sur des groupes de référence différents (les États membres du Conseil de l'Europe ; les États membres de l'UE) et sur des années différentes. Dans son communiqué de presse du 10 janvier 2025 sur le pilotage par l'État fédéral de la transformation numérique de la justice, la Cour des comptes (*Cour des comptes*) constate que cette transformation « n'est pas menée de manière satisfaisante » et que la numérisation de la justice « reste en chantier depuis plusieurs décennies ».[19]

1.5 Limites des connaissances actuelles

L'absence de mesure belge ou française du temps de recherche est constatée au § 1.4. Les taux d'hallucination ci-dessus n'ont été mesurés que sur le droit américain. La part de 1 % est une estimation dont la méthode n'est pas indiquée.

2. La situation du juriste

Les sources disponibles sur ce point sont limitées. Cette section distingue ce qui est documenté de ce qui ne l'est pas.

2.1 La place de la recherche de jurisprudence dans la préparation d'un dossier

Le cas d'usage consigné pour ce projet décrit la séquence suivante. Un juriste reçoit un dossier. Les faits, la question juridique, les dispositions invoquées, les parties et les dates clés sont identifiés ; les questions du client et les arguments de la partie adverse sont listés ; de la jurisprudence est recherchée pour chacun d'eux ; et le juriste rédige une réponse qui cite les décisions retenues. Heuris suit cette séquence : aucune recherche n'est lancée avant que le juriste ait validé la fiche du dossier, les questions déduites du dossier ne sont recherchées qu'après validation par le juriste, et la réponse reste l'œuvre du juriste (D46–D53 ; voir l'annexe B). (Sources : *documentation du projet (non publique)* ; *documentation du projet (non publique)*, étapes 6–12g.)

Cette séquence est le cas d'usage consigné par le concepteur le 2 novembre 2026. Elle ne résulte pas d'une étude sur la manière dont travaillent les juristes belges, et aucune étude de ce type n'a été trouvée. Aucun chiffre belge n'a non plus été trouvé sur le temps que les juristes consacrent à la recherche de jurisprudence, ni sur la fréquence à laquelle ils vérifient les citations sur lesquelles ils s'appuient (§ 1.5).

2.2 La portée professionnelle d'une provenance vérifiable

Le juriste doit savoir ce qu'une juridiction a décidé, sur quels faits, et où cela est écrit : juridiction, date, numéro de rôle (*R.G.*) et un passage qui peut être ouvert et lu dans son contexte (*documentation du projet (non publique)*). Les décisions rapportées au § 1.2 montrent les conséquences professionnelles de citations qui ne peuvent être retracées. Dans *Mata v. Avianca*, le tribunal a estimé que les avocats avaient « abandonné leurs responsabilités » en soumettant des décisions inexistantes.[10] En Belgique, la cour d'appel d'Anvers a condamné un avocat à payer une amende et une indemnité à chacune des trois parties, les conclusions reposant sur une jurisprudence inexistante,[11] et l'Orde van Vlaamse Balies a relevé que de telles erreurs font perdre du temps aux juridictions et aux confrères.[12] La préparation d'un dossier implique aussi les faits du client, couverts par le secret professionnel ; dans un flux de travail assisté, ces faits entrent dans la requête et dans le matériau soumis au modèle de langage (*documentation du projet (non publique)* ; *documentation du projet (non publique)*). Les règles applicables ne sont pas analysées ici (§ 13, limite 20). Les règles déontologiques applicables en Belgique à l'usage de l'IA ne sont pas analysées ici. L'Orde van Vlaamse Balies renvoie à des lignes directrices élaborées avec Avocats.be ;[12] elles n'ont pas été consultées.

3. Exigences de fiabilité pour un outil de recherche jurisprudentielle

Le problème exposé aux §§ 1 et 2 implique des exigences pour tout système qui assiste la recherche de jurisprudence. Elles sont formulées indépendamment de toute implémentation, et chacune n'est retenue que parce que le § 1 ou le § 2 la justifie.

- **R1 — Traçabilité littérale.** Toute affirmation présentée au juriste doit pouvoir être rattachée à un passage d'une source identifiée, qu'elle cite littéralement et qui peut être ouvert dans son contexte. *Justifiée par* le § 1.1 (une citation qui n'étaye pas sa proposition compte comme une hallucination) et le § 2.2 (juridiction, date, numéro de rôle et un passage consultable).
- **R2 — Vérification indépendante de la génération.** Le fait qu'un passage cité étaye une affirmation doit être contrôlé par une étape distincte de celle qui a généré l'affirmation ; le *retrieval* seul ne suffit pas. *Justifiée par* le § 1.1 (les outils juridiques fondés sur le *retrieval* hallucinent encore ; l'hallucination n'est pas éliminée dans les systèmes récents).

- **R3 — Le refus plutôt qu'une réponse non étayée.** Lorsqu'aucune affirmation ne peut être étayée, le système doit le dire au lieu de répondre. *Justifiée par* le § 1.2 (sanctions pour des conclusions reposant sur une jurisprudence inexistante).
- **R4 — Périmètre et lacunes du corpus visibles.** Le système doit indiquer ce que couvre son corpus et ce qu'il omet, afin qu'une absence de résultats ne soit pas lue comme une absence de jurisprudence. *Justifiée par* le § 1.3 (environ 1 % des décisions publiées, selon une estimation ; Juportal est une sélection) et le § 1.5 (les lacunes constatées).
- **R5 — Clarté sur la nature et l'autorité de la source.** Chaque résultat doit identifier la juridiction, la date et le numéro de rôle, et indiquer si la source est la décision elle-même ou un texte secondaire qui la commente. *Justifiée par* le § 2.2 (ce qu'une juridiction a décidé, sur quels faits, et où cela est écrit) et le § 1.3 (textes officiels sélectionnés et accès restreint).
- **R6 — Le juriste garde la décision.** Le système peut proposer ; le juriste valide, rédige et reste l'auteur de ce qui est déposé. *Justifiée par* le § 2.1 (le juriste rédige la réponse) et le § 2.2 (la responsabilité de l'avocat pour ce qu'il soumet).
- **R7 — Confidentialité des faits du client.** La confidentialité des faits du client doit être protégée tout au long du traitement, y compris lorsqu'ils sont soumis à un modèle de langage et lorsque les résultats sont conservés. *Justifiée par* le § 2.2 (les faits du client, couverts par le secret professionnel, entrent dans la requête).
- **R8 — Un effort de recherche réduit sans perte de traçabilité.** Le système devrait réduire l'effort nécessaire pour trouver la jurisprudence pertinente par rapport à une recherche par numéro, date et filtres, sans affaiblir R1. *Justifiée par* le § 1.4 (recherche par ECLI, numéro de rôle, date et filtres, sans recherche sémantique dans la description officielle ; aucune mesure du temps de recherche ; moyens contraints).

4. Travaux connexes

4.1 Les outils de recherche juridique à *retrieval-augmented generation*

Le *retrieval-augmented generation* conditionne un modèle de langage par des passages extraits d'un corpus externe.[2] Magesh et al. ont évalué des outils commerciaux de recherche juridique fondés sur cette approche, sur 202 requêtes préenregistrées portant sur le droit américain, et rapportent des taux d'hallucination de 17–33 %. Leur définition de l'hallucination inclut une réponse dont la citation n'étaye pas la proposition à laquelle elle est rattachée.[1] Dahl et al. avaient documenté des taux plus élevés pour des modèles généralistes utilisés sans *retrieval*.[3]

4.2 L'évaluation de l'attribution et des citations

Plusieurs travaux traitent la question de savoir si un texte généré est étayé par une source comme un problème d'évaluation. Rashkin et al. définissent « Attributable to Identified Sources » (AIS), un cadre et une procédure d'annotation humaine en deux étapes pour juger si des énoncés générés par un modèle sont étayés par les sources sous-jacentes.[20] Bohnet et al. formulent la tâche de réponse attribuée aux questions, dans laquelle un système renvoie une réponse accompagnée d'une source qui l'étaye, et proposent un cadre d'évaluation reproductible dans lequel une métrique automatique corrélée aux jugements humains sert au développement.[21] Gao et al. introduisent ALCE, un benchmark pour les systèmes qui recherchent des éléments de preuve et génèrent des réponses avec citations, assorti de métriques automatiques de fluidité, d'exactitude et de qualité des citations.[22]

4.3 Décomposition en affirmations et *entailment check*

Min et al. (FActScore) décomposent un texte généré en faits atomiques et calculent la part de ces faits étayée par une source de connaissances ; ils proposent aussi un estimateur automatisé fondé sur le *retrieval* et un modèle de langage.[23] Honovich et al. (TRUE) comparent des métriques de cohérence factuelle sur une collection standardisée de textes annotés et rapportent que les approches à grande échelle d'inférence en langage naturel (implication) et de génération et réponse de questions obtiennent des résultats solides et complémentaires.[24]

4.4 Positionnement de Heuris

Heuris combine des éléments présents dans cette littérature : la recherche de passages (§ 8), la décomposition des réponses en affirmations et un *entailment check* (vérifier que le passage cité implique bien l'affirmation) par un modèle de langage (§ 9). Ses choix propres consistent à exiger que chaque citation soit retrouvée en entier aux positions de caractères citées, et à retenir, plutôt qu'à noter, toute affirmation qui échoue à un contrôle (D9, D15). Ce sont des choix de conception, non des résultats. Heuris n'a été évalué sur aucun des benchmarks cités ci-dessus, et aucune comparaison d'efficacité n'est faite ni suggérée.

5. Fluidité et provenance vérifiable

Les exigences du § 3 servent ci-dessous à situer Heuris. Le statut de chacune n'est apprécié qu'au regard des faits exposés dans ce document ; il ne s'agit pas d'une évaluation et cela n'implique aucune affirmation de performance.

R	Comment Heuris y répond	Statut

Tableau 5.1 · Exigences du § 3 et manière dont Heuris y répond

R	Comment Heuris y répond	Statut
R1	Passages conservés à des positions exactes ; chaque affirmation doit citer son passage en entier aux positions citées ; en mode juriste, les citations sont découpées par le serveur (D2, D9, D24, D53)	Satisfaite, au regard du texte du commentaire (voir R5)
R2	Le *verbatim check* (le passage cité figure mot pour mot dans la source) est du code déterministe ; le contrôle d'étayage est un second appel au même modèle de langage que celui qui a rédigé l'affirmation (D9 ; *documentation du projet (non publique)*) ; les phrases du mode juriste ne font pas l'objet d'un contrôle d'étayage (D53)	Partiellement satisfaite
R3	Les affirmations qui échouent sont supprimées ; au moins une affirmation doit être vérifiée par rapport à un passage de décision, sinon « Absence de réponse » (D9, D15, D16 ; § 9.4)	Partiellement satisfaite : implémentée, mais sa fiabilité dépend du contrôle d'étayage, qui n'a pas été mesuré
R4	Le périmètre et les lacunes sont indiqués dans ce document (§ 1.3 ; § 13, limite 19) ; l'interface n'affiche aucune mention du périmètre du corpus (*documentation du projet (non publique)*, `app.js`)	Partiellement satisfaite
R5	Juridiction (avec la chambre lorsqu'elle est enregistrée), date et R.G., avec un lien vers le commentaire ; mais le corpus est composé de commentaires et l'interface désigne les éléments comme « décision » (§ 6.2 ; § 13, limite 12)	Partiellement satisfaite
R6	Fiche du dossier validée avant toute recherche ; les éléments déduits exigent une validation ; la rédaction est optionnelle ; le juriste valide les réponses (D46–D53)	Satisfaite dans la conception ; exercée en mode simulé uniquement (§ 13, limite 2)
R7	Les passages extraits, la fiche du dossier et des extraits du dossier sont envoyés à l'API DeepSeek, exploitée hors de l'UE (D11 ; note 28) ; les brouillons et les réponses validées sont conservés sans chiffrement et sans limite de durée (`06_confidentiality.md` ; § 13, limite 7) ; un modèle local ou hébergé dans l'UE est prévu, non implémenté (§ 11)	Non satisfaite

R	Comment Heuris y répond	Statut
R8	La recherche hybride combine recherche lexicale (BM25) et sémantique (bge-m3), fusionnées par RRF et reclassées sur la machine locale ; chaque sous-question d'un dossier reçoit au moins un emplacement de résultat (D4–D7, D17, D18, D27) ; chaque affirmation reste soumise au *verbatim check* (R1)	Non démontrée : l'effort de recherche n'a pas été mesuré et aucune étude de temps n'existe ; quinze questions rédigées par le concepteur ne permettent pas de départager les modes de recherche (§ 13, limite 6)

5.1 Ce que le juriste exige d'une réponse

Un paragraphe plausible mais non sourcé est d'une utilité limitée pour un juriste, en droit du travail comme dans toute autre matière. Le juriste doit savoir ce qu'une juridiction a décidé, sur quels faits, et où cela est écrit : juridiction, date, numéro de rôle (*R.G.*), et un passage qu'il peut ouvrir et lire dans son contexte. Devant un juge, une citation non étayée est plus dommageable que l'absence de réponse. La conception repose sur une contrainte : chaque énoncé doit renvoyer à un passage exact et consultable, et le système doit refuser lorsqu'il ne le peut pas. (Source : *documentation du projet (non publique).*)

5.2 Les limites du *retrieval* seul

Le *retrieval* restreint le matériau que le modèle lit ; il ne contraint pas ce que le modèle écrit. Trois défaillances restent possibles après la recherche : le bon document mais le mauvais passage ; le bon passage, paraphrasé en quelque chose qu'il ne dit pas ; et une citation d'apparence correcte qui n'étaye pas l'affirmation à laquelle elle est rattachée. L'étude citée plus haut classe le dernier cas, une citation qui n'étaye pas sa proposition, comme une hallucination à part entière.[1] Heuris traite chacun des trois cas comme un problème de vérification, et non comme un problème de formulation des instructions.

6. Licéité des données

Les fiches de décision citées dans cette section sont expliquées à l'intention des juristes à l'annexe B.

6.1 Sources exclues

Juportal est la base de données officielle de jurisprudence du SPF Justice. Son fichier robots.txt interdit tous les agents automatisés à l'exception du robot d'indexation du ministère lui-même.[25] Heuris n'explore pas Juportal : aucune exploration, et aucun contournement par navigateur (fiche de décision D1). Les décisions officielles en texte intégral et leurs identifiants ECLI ne font donc pas partie du corpus.

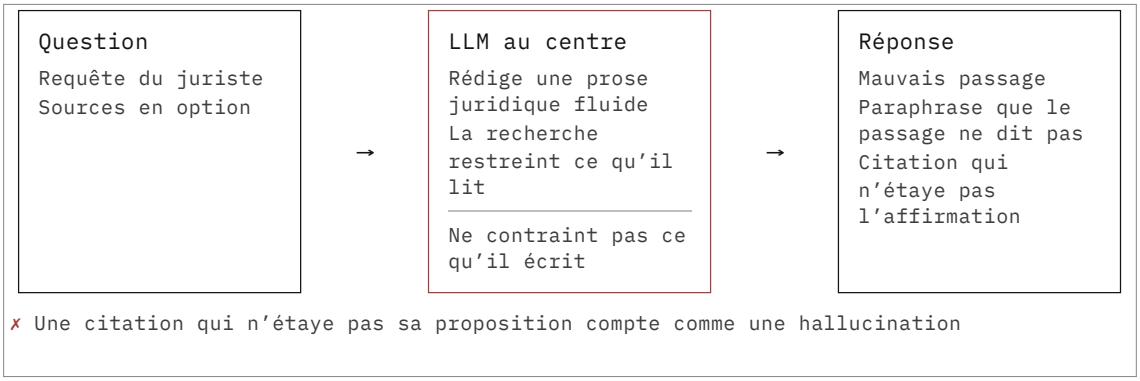


Figure 1 · Schéma par défaut : le LLM au centre, les sources en option

Le *retrieval* restreint le matériau que le modèle lit ; il ne contraint pas ce que le modèle écrit.

Trois défaillances restent possibles après la recherche : le bon document mais le mauvais passage ; le bon passage, paraphrasé en quelque chose qu'il ne dit pas ; et une citation d'apparence correcte qui n'étaye pas l'affirmation à laquelle elle est rattachée.

6.2 Sources retenues (premier corpus : droit du travail)

Le premier corpus se compose de commentaires en langue française de décisions des *tribunaux du travail* et des *cours du travail* belges. Le fichier robots.txt de l'éditeur autorise les pages d'articles. Ses conditions d'utilisation autorisent la reproduction des informations textuelles à condition que la source soit clairement citée, et interdisent toute exploitation commerciale.[26] Le collecteur respecte un intervalle de 1,5 seconde entre les requêtes, supérieur au délai d'exploration déclaré, et conserve les conditions d'utilisation à l'identique dans chaque enregistrement. Chaque passage renvoyé par le système porte l'URL de sa source et la référence de la décision. (Sources : *documentation du projet (non publique)* § 3 ; *documentation du projet (non publique)* ; *documentation du projet (non publique)* § 5.)

Les documents sont des commentaires, et non des décisions : le collecteur type chaque enregistrement comme commentaire (*documentation du projet (non publique)*). Ils portent sur des *arrêts* de cours du travail et sur des *jugements* de tribunaux du travail. Chacun comprend un titre, une ligne de référence, un résumé introductif, les dispositions légales et une analyse dans laquelle le commentateur rapporte le raisonnement de la juridiction, en partie par paraphrase et en partie par extraits cités, qu'il s'agisse de la juridiction ou d'autres textes. Le *dispositif* est le plus souvent absent. Les commentaires ne portent pas d'ECLI ; une décision est identifiée par la juridiction (avec la chambre lorsqu'elle est enregistrée), la date et le numéro de rôle, et chaque résultat renvoie au commentaire sur le site de l'éditeur. (Sources : *documentation du projet (non publique)*, *documentation du projet (non publique)* ; `01_data-sources.md`.)

Définition. Dans tout ce document, « vérifié » signifie que l'extrait correspond au texte source — ici, le commentaire — aux positions citées. Cela ne signifie pas que la juridiction l'a dit : un passage vérifié peut être la paraphrase ou l'analyse du commentateur.

Source	Contenu	Statut dans Heuris	Fondement

Tableau 6.1 · Comparaison des sources

Source	Contenu	Statut dans Heuris	Fondement
Juportal (SPF Justice)	Décisions officielles en texte intégral, identifiées par ECLI	Non collecté	robots.txt interdit tous les robots sauf celui du ministère[25]
Recherche ECLI du portail e-Justice européen	Index ECLI comprenant des décisions belges	Non utilisé	Signalé comme fermé aux robots (01 § 1 ; non revérifié)
Éditeur de commentaires de jurisprudence	Commentaires de décisions des tribunaux et cours du travail	Utilisé	robots.txt autorise les pages d'articles ; les conditions permettent la reproduction avec mention de la source et interdisent l'usage commercial[26]

6.3 Corpus

Heuris fonctionne avec un corpus de plus de 25 000 décisions. L'éditeur publie parfois un même commentaire sous plusieurs identifiants d'article ; le dédoublonnage par empreinte du texte normalisé conserve le plus petit identifiant. Les enregistrements qui partagent une référence de décision mais diffèrent par le corps ou le titre sont conservés. Le choix de calculer l'empreinte sur le contenu plutôt que sur la référence était délibéré : une référence commune ne prouve pas l'identité du texte (D3). Les quasi-doublons qui ne diffèrent que par la ponctuation constituent une lacune résiduelle connue. (Sources : `01_data-sources.md` §§ 4–5 ; *documentation du projet (non publique)* ; *documentation du projet (non publique).*)

7. Segmentation respectueuse de la structure

Les fiches de décision citées dans cette section sont expliquées à l'intention des juristes à l'annexe B.

Un vérificateur ne peut contrôler une citation par rapport à un passage que si ce passage est exactement ce que dit la source. Heuris ne découpe donc chaque commentaire qu'aux endroits où son auteur l'a fait : aux intertitres et aux limites de paragraphe. Les paragraphes trop longs (plus de 1 500 caractères) ne sont scindés qu'à une fin de phrase située hors d'une citation entre « ». Aucun passage ne se termine au milieu d'une phrase et aucun passage ne chevauche un intertitre. Pour chaque passage, l'égalité `source_text[char_start:char_end] == passage` est vérifiée à la construction et dans les tests. Un document comportant moins de deux types de section reconnus est signalé et ne produit aucun passage plutôt que des passages mal découpés (D2).

Les types de section suivent la structure des commentaires : résumé, dispositions légales, faits, procédure, positions des parties, décision du premier juge, décision de la juridiction, avis du ministère public et notes. Aucun passage ne provient d'un document signalé. (Sources : *documentation du projet (non publique)* §§ 1–5 ; *documentation du projet (non publique)* ; *documentation du projet (non publique)* ; *documentation du projet (non publique).*)

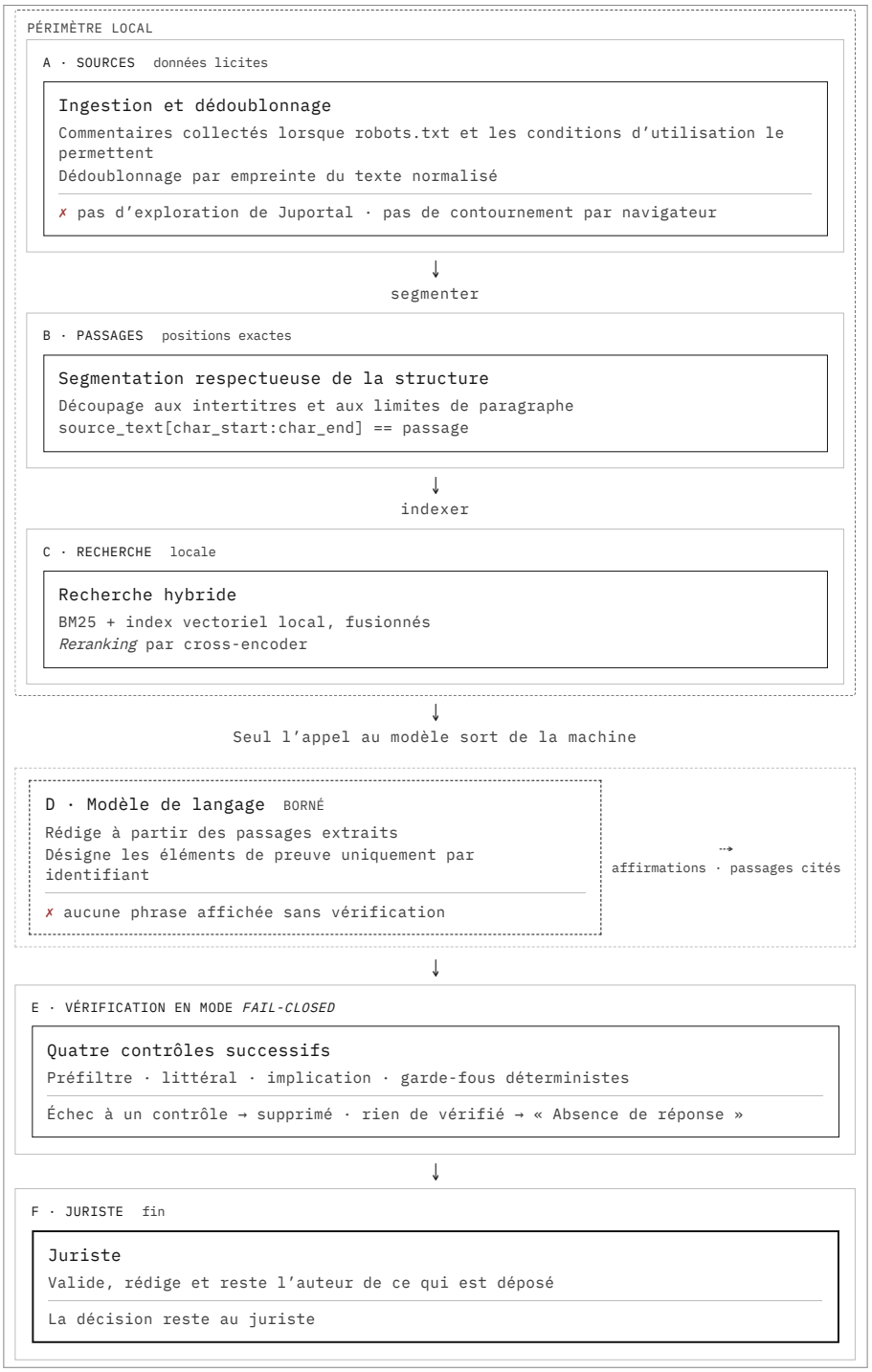


Figure 2 · *Spine* (ossature de données et de moteurs qui fait référence) de Heuris : sources, passages, recherche hybride (BM25 + index vectoriel local), vérification en mode *fail-closed*, juriste

Chaque énoncé doit renvoyer à un passage exact et consultable, et le système doit refuser lorsqu'il ne le peut pas. Le modèle ne peut désigner les éléments de preuve que par identifiant. Chaque phrase générée est décomposée en affirmations ; une affirmation qui échoue à un contrôle est supprimée. La décision reste au juriste.

Tableau 7.1 · Blocs de taille fixe et passages respectueux de la structure

	Blocs de taille fixe	Passages respectueux de la structure (Heuris)
Points de coupe	Tous les *n* jetons, n'importe où	Intertitres, paragraphes, fins de phrase hors citations
Identité de la section	Perdue	Conservée (résumé, faits, décision de la juridiction…)
Positions exactes	Généralement approximatives	Vérifiées pour chaque passage
Entrée non structurée	Découpée en silence	Signalée, aucun passage
Adapté à la vérification littérale	Faiblement	Par construction

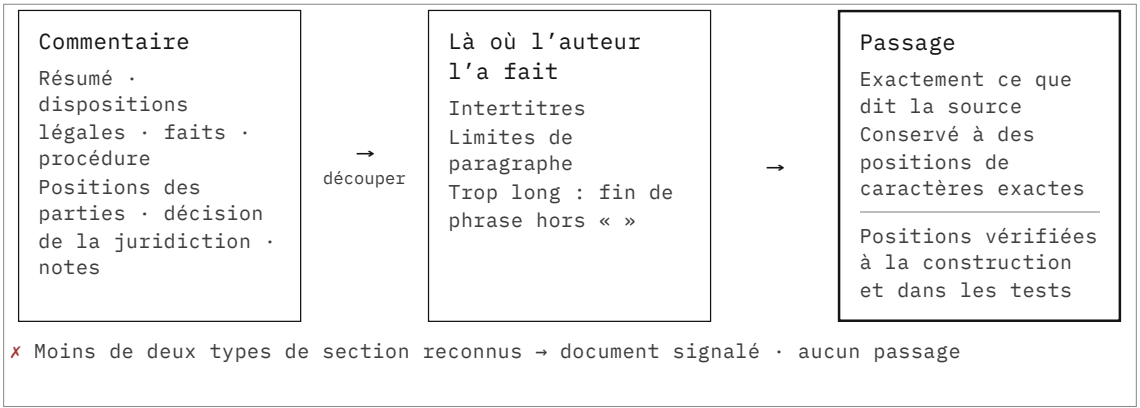


Figure 3 · Segmentation respectueuse de la structure : des passages découpés là où l'auteur l'a fait

Un vérificateur ne peut contrôler une citation par rapport à un passage que si ce passage est exactement ce que dit la source.

Aucun passage ne se termine au milieu d'une phrase et aucun passage ne chevauche un intertitre. Un document comportant moins de deux types de section reconnus est signalé et ne produit aucun passage plutôt que des passages mal découpés.

8. *Retrieval*

Les fiches de décision citées dans cette section sont expliquées à l'intention des juristes à l'annexe B.

En droit du travail, les termes exacts comptent — « motif grave », un numéro d'article, un numéro de rôle — et le sens aussi. Heuris combine les deux. Un index creux (BM25 avec un racinisateur français, qui conserve comme jetons les références juridiques telles que « art. 35 ») et un index dense (le modèle multilingue bge-m3, exécuté localement et stocké dans une table LanceDB locale) sont fusionnés par *reciprocal rank fusion* (fusion des classements issus de plusieurs recherches) avec k = 60.[27] Les deux index, lexical et vectoriel, s'exécutent sur la machine locale. Les 50 premiers passages fusionnés sont reclassés par le cross-encoder bge-reranker-v2-m3. Lorsqu'un dossier est analysé, chaque sous-question produit sa propre requête, et les rangs par décision sont à nouveau fusionnés par *reciprocal rank fusion* (D4–D7, D17). Le calcul des *embeddings* (représentations vectorielles du texte) et le *reranking* (réordonnancement des résultats par un modèle) s'exécutent sur la machine locale ; aucun texte du corpus n'est envoyé à un ser-

vice d'*embeddings* hébergé. La génération est différente : les passages extraits sont envoyés, avec la question, à l'API LLM externe du fournisseur configuré — DeepSeek par défaut — qui rédige la réponse (D11 ; *documentation du projet (non publique)*).

L'évaluation de la recherche est réduite et présentée comme telle. Quinze questions provisoires rédigées par le concepteur donnent un recall@5 de 1,00 dans tous les modes, et un rang réciproque moyen de 1,00 (BM25), 0,92 (dense), 0,97 (hybride) et 0,96 (hybride avec *reranking*). Ce jeu ne permet pas de départager les modes ; il n'étaye donc aucune affirmation selon laquelle un mode serait meilleur. La conception hybride repose sur l'argument d'ingénierie exposé ci-dessus, et non sur ces chiffres. (Sources : *documentation du projet (non publique)* ; *documentation du projet (non publique)* ; *documentation du projet (non publique)* ; *documentation du projet (non publique)*, *documentation du projet (non publique).*)

Tableau 8.1 · Modes de recherche

Mode	Point fort	Défaillance typique	Rôle dans Heuris
BM25 seul	Termes juridiques exacts, références	Manque les paraphrases	Composant
Dense seul	Sens, paraphrase	Brouille les termes exacts et les numéros	Composant
Hybride (RRF)	Les deux signaux, fusion fondée sur les rangs	Hérite des angles morts des deux index	Liste de candidats par défaut
Hybride + *reranking*	Les meilleurs passages en premier	Plus lent sur CPU	Classement par défaut

9. Vérification en mode *fail-closed*

Les fiches de décision citées dans cette section sont expliquées à l'intention des juristes à l'annexe B.

La principale différence avec un assistant classique à *retrieval-augmented generation* tient à la vérification. Le modèle ne peut désigner les éléments de preuve que par identifiant : des passages de décisions, et des paragraphes du dossier dont la citation a été localisée littéralement dans le dossier. La vérification porte sur les phrases générées par le modèle : chacune est décomposée en affirmations, et chaque affirmation est soumise à quatre contrôles successifs. Si elle échoue à l'un d'eux, elle est supprimée. Les phrases écrites par le juriste en mode juriste ne sont pas vérifiées de cette manière : leurs citations sont découpées littéralement par le serveur, donc exactes, mais aucun *entailment check* n'est effectué sur elles (D53 ; § 10). (Source : *documentation du projet (non publique)* ; *documentation du projet (non publique).*)

9.1 Contrôles de vérification

1. **Préfiltre.** Suppression, par des règles, des affirmations qui ne peuvent être vérifiées en tant que telles, par exemple les affirmations négatives ou les méta-affirmations sur une décision (D16).
2. ***Verbatim check.*** La citation de l'affirmation doit être retrouvée en entier aux positions citées, en tolérant uniquement les variantes d'espaces et d'apostrophes, et elle est affichée avec les caractères d'origine de la source (D9, D24).
3. ***Entailment check.*** Un second appel au LLM doit juger que le passage cité étaye l'affirmation (D9).
4. **Garde-fous déterministes.** Chaque nombre, mois, date ou montant figurant dans une affirmation relative aux faits du dossier, et chaque date ou montant figurant dans une conclusion, doit provenir de la citation ou d'une prémisse vérifiée (D23, D29).

9.2 Règles au niveau de la réponse

- Un résultat exige au moins une affirmation vérifiée par rapport à un passage de **décision** ; des affirmations portant sur le seul dossier ne constituent pas une réponse de jurisprudence (D15).
- Pour chaque décision, le système tente d'établir un **lien fait → règle** : un fait du dossier et un passage du commentaire rapportant le motif sur lequel la juridiction s'est fondée, tous deux cités, tous deux vérifiés. Le passage peut être formulé par le commentateur ; l'étiquette « fait → règle » n'implique pas que la décision crée une règle. Si l'une des deux moitiés échoue, l'interface affiche « Lien avec le dossier non établi » (D30, D31).
- L'**étape d'application**, qui rapporte une règle au dossier, n'utilise que des prémisses vérifiées et est formulée comme une indication. Lorsque des prémisses manquent, elle ne conclut pas : l'interface affiche « Application au dossier — absence de conclusion », suivi de « Éléments manquants : … », avec la liste de ce qui manque (valeur de statut interne `insufficient` ; D19, D25).

9.3 Un score de confiance heuristique

Chaque résultat porte un score combinant la pertinence selon le *reranker* (poids 0,40), la part des affirmations rédigées qui ont passé la vérification (0,30), la similarité factuelle entre le dossier et la décision (0,20) et l'ancienneté (0,10, divisée par deux tous les cinq ans). Il est forcé à 0 lorsque rien n'est vérifié. Des paliers en langage courant traduisent le score : « Lien fort » exige un score d'au moins 0,78 **et** un lien fait → règle vérifié ; « Lien partiel » commence à 0,65 ; « Lien faible » en dessous. Les poids et les seuils sont fixés à la main et non calibrés, et l'interface les présente comme tels (D10, D18, D34 ; *documentation du projet (non publique)* ; *documentation du projet (non publique)*).

	Génération libre	***Retrieval-augmented generation***	**Ancrage en mode *fail-closed* (Heuris)**

Tableau 9.1 · Ce que chaque approche garantit au sujet d'un passage cité

	Génération libre	***Retrieval-augmented generation***	**Ancrage en mode *fail-closed* (Heuris)**
Le passage existe	Aucune garantie	Extrait, en général	Identifié par un identifiant détenu par le serveur
Citation exacte	Non	Non contrôlée	Segment complet contrôlé aux positions
Le passage étaye l'affirmation	Non	Non contrôlée	*Entailment check*, sinon supprimée
Nombres et dates	Non contrôlés	Non contrôlés	Garde-fou déterministe
Lorsque les preuves manquent	Répond quand même	Répond souvent quand même	Supprime l'affirmation, ou refuse

Le refus par défaut a un coût : certains énoncés corrects sont supprimés parce qu'ils n'ont pu être vérifiés. Ce coût est accepté délibérément, car un juriste peut lire une décision que le système n'a pas résumée, alors qu'une phrase que le système n'aurait pas dû écrire est nettement plus difficile à détecter.

9.4 Exemple détaillé (mode simulé, pas une évaluation)

Pour illustrer les étapes, l'une des quinze questions de recherche rédigées par le concepteur (q02 : « Les chauffeurs Uber sont-ils liés par un contrat de travail ? ») a été soumise à la recherche et à la vérification le 4 novembre 2026 en mode simulé. Le modèle de langage a été remplacé par le substitut déterministe utilisé dans les tests automatisés ; aucune API externe n'a été appelée et aucun dossier n'a été fourni. Dans ce mode, l'*entailment check* réussit toujours, et le substitut émet, par conception, une citation invérifiable lorsqu'il ne trouve aucun passage utilisable. L'exécution illustre donc uniquement le mécanisme ; ce n'est pas une évaluation et elle n'étaye aucune affirmation de performance. (Commande, script et journal : *documentation du projet (non publique).*) Instruction d'ancrage `legal-ground-v4` ; schéma de sortie `legal-grounding-v1`. L'exemple utilise la même chaîne de traitement avec un modèle simulé (implication toujours vraie) ; il montre donc le flux, et non la qualité de la vérification (*documentation du projet (non publique)*, `schema.py:7`). L'exemple ne dit rien du cadre juridique régissant la qualification d'une relation de travail, que ce document ne décrit pas.

La recherche a renvoyé huit passages. Six, dont les cinq premiers, provenaient du commentaire désigné comme document attendu provisoire pour cette question (`4200`) ; les deux autres provenaient de deux autres commentaires. Le substitut a rédigé une seule affirmation, citant le passage classé premier, dont la citation ne figure pas dans ce passage. La vérification l'a rejetée (`verify: fail` dans la trace d'audit), aucune affirmation n'a subsisté, et le système a renvoyé « Absence de réponse : aucune affirmation vérifiable dans les sources. » L'exemple montre ainsi la règle du mode *fail-closed* à l'œuvre : une phrase qui ne peut être rattachée à sa source est retenue, et non affichée.

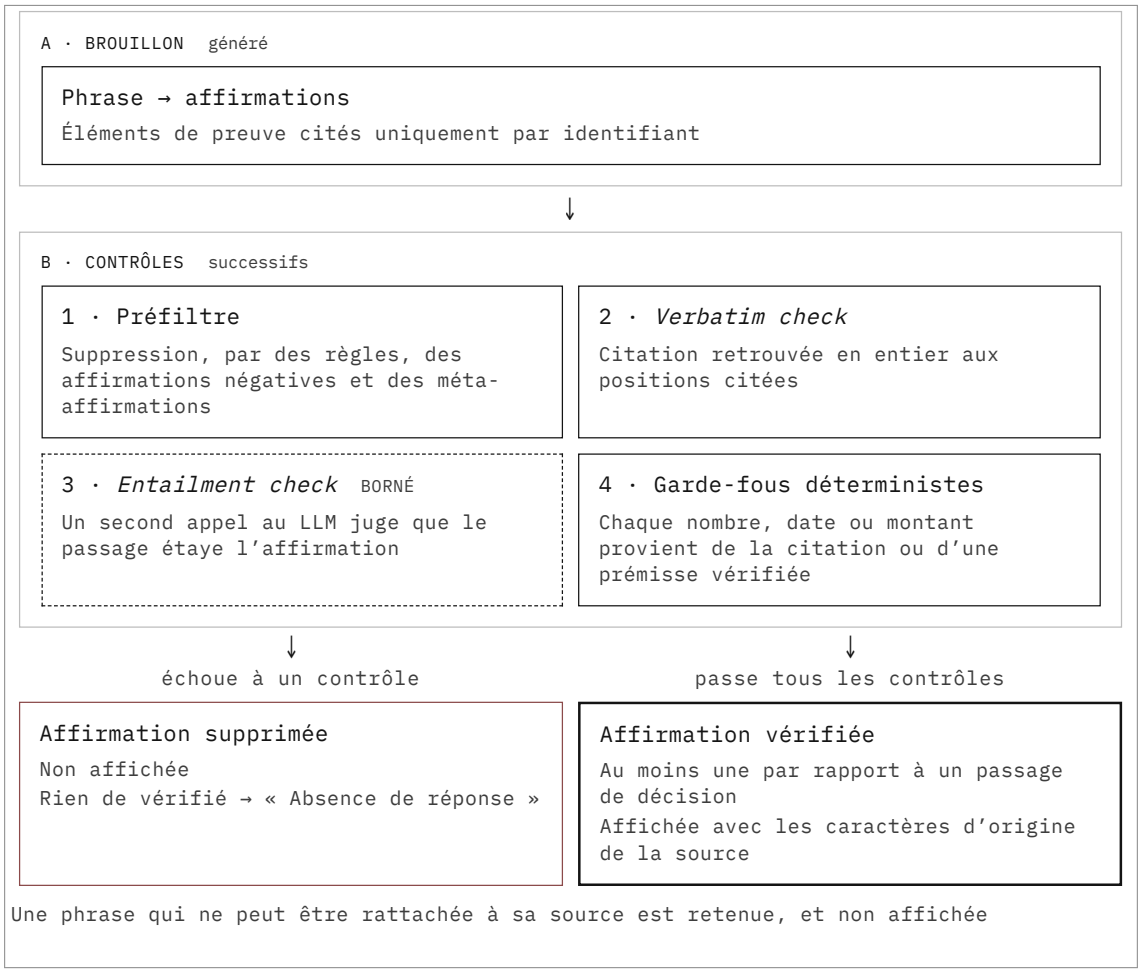


Figure 4 · Vérification en mode *fail-closed* : quatre contrôles par affirmation, refus par défaut

Chaque phrase générée par le modèle est décomposée en affirmations, et chaque affirmation est soumise à quatre contrôles successifs. Si elle échoue à l'un d'eux, elle est supprimée.

Un résultat exige au moins une affirmation vérifiée par rapport à un passage de décision. Lorsqu'aucune affirmation ne peut être étayée, le système le dit au lieu de répondre.

10. L'autorité du juriste et le *workbench*

Les fiches de décision citées dans cette section sont expliquées à l'intention des juristes à l'annexe B.

Les résultats du système sont soumis à la validation du juriste ; la décision reste au juriste. L'interface est un *workbench* (interface de travail du juriste) en trois colonnes : la fiche du dossier et le dossier à gauche, les résultats au centre, le texte intégral de la décision à droite, avec des repères de passage numérotés (D13, D43, D45).

- **Fiche du dossier.** Un résumé typé et sourcé du dossier — faits, dispositions, événements datés — dans lequel chaque champ cite le dossier. Le juriste la valide avant la recherche (D28, D35).
- **Questions et arguments.** Les questions du client et les arguments adverses deviennent des éléments cibles. Les éléments que le système déduit doivent être validés par le juriste avant d'être recherchés (D46, D47).
- **Positions indicatives.** Pour chaque élément, une décision peut être étiquetée comme tendant à le soutenir ou à le contredire, mais seulement lorsque l'appariement est vérifié. L'étiquette

décrit le sens du raisonnement d'une décision tel que rapporté dans un commentaire et n'a aucune valeur de précédent. Les synthèses par groupe comptent ces étiquettes (par exemple « 2 tendent à soutenir » ; *documentation du projet (non publique)*) ; un tel décompte ne mesure pas l'état de la jurisprudence. Sinon, l'étiquette est « non établi ». Une position est une indication pour le juriste, jamais une conclusion.

- **Assistant de rédaction (optionnel).** Le projet est affiché sous les décisions extraites et ne les remplace pas. Un projet de réponse compte au plus six phrases, chacune vérifiée comme toute autre affirmation. Lorsque les décisions rattachées ne contiennent aucun passage suffisamment vérifié, le panneau affiche « Les décisions rattachées à cette question ne comportent pas de passage vérifié suffisant pour proposer un projet de réponse. » et « Cela ne préjuge pas de leur pertinence. », avec « Consulter les décisions » et « Rédiger soi-même » (valeur de statut interne `insufficient` ; D48, D51, D52).
- **Mode juriste.** Le juriste rédige la réponse et y rattache des citations en sélectionnant du texte dans la décision ou dans le dossier ; le serveur découpe la citation exacte à partir des positions. Ces citations sont exactes, mais aucun *entailment check* n'est effectué sur elles (D53).
- **Réponses validées.** La validation crée une version immuable ; un retrait est consigné, non effacé. Les réponses s'exportent vers Word avec leurs citations ; le statut de vérification par phrase n'apparaît que dans l'annexe de l'export consacrée aux projets initiaux, et l'export n'indique pas que les citations du mode juriste n'ont pas fait l'objet d'un *entailment check* (*documentation du projet (non publique)*) ; une trace d'audit consigne chaque étape (D49, D50).

L'assistant de rédaction et les fonctions du *workbench* décrites ci-dessus n'ont été exercés qu'en mode LLM simulé (§ 13). (Sources : *documentation du projet (non publique)* D12–D13, D35–D54 ; *documentation du projet (non publique)* ; *documentation du projet (non publique)*, *documentation du projet (non publique)*, *documentation du projet (non publique)*.)

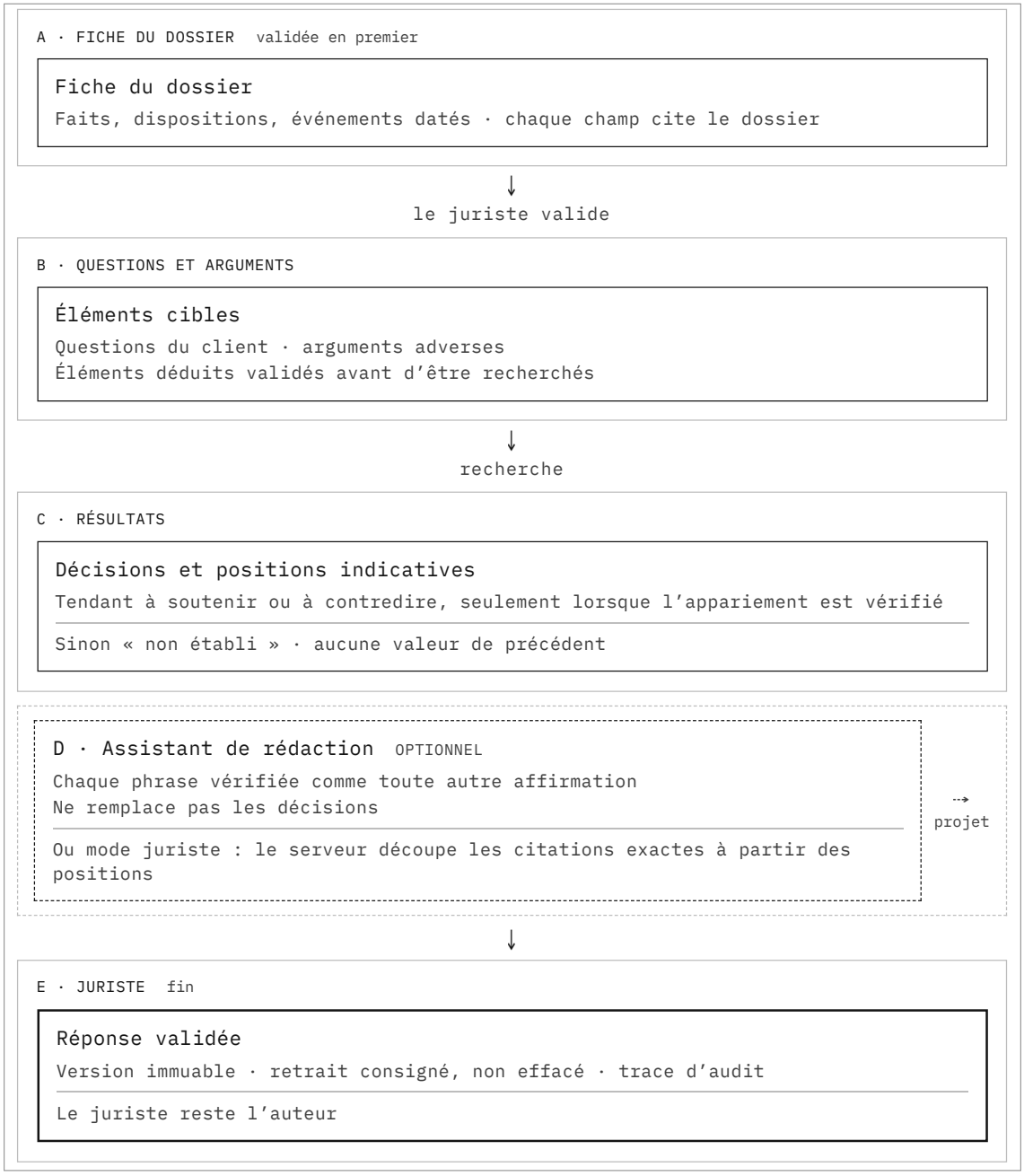


Figure 5 · Le *workbench* : le système propose, le juriste valide

Les résultats du système sont soumis à la validation du juriste ; la décision reste au juriste.

Aucune recherche n'est lancée avant que le juriste ait validé la fiche du dossier, et les questions déduites du dossier ne sont recherchées qu'après validation par le juriste. Une position est une indication pour le juriste, jamais une conclusion.

11. Confidentialité et protection des données

La préparation d'un dossier fait entrer les faits du client dans la requête et dans le matériau que le système traite (§ 2.2). Le secret professionnel et le RGPD[29] font donc du chemin suivi par le texte du dossier une contrainte de conception, et non un détail d'exploitation. L'architecture générique à *retrieval-augmented generation* va à l'encontre de cette contrainte : les passages sont vectorisés via une API hébergée, stockés dans un service géré de base de données vectorielle et envoyés à une API distante de modèle de langage. Chacune de ces étapes transmet des passages, des vecteurs ou des requêtes à un tiers, selon les conditions de ce tiers et dans sa juridiction.

Heuris est construit à l'inverse pour chaque étape, à l'exception de la génération. Les mesures sont énumérées ci-dessous, avec deux exceptions indiquées d'emblée : les brouillons et les réponses validées, qui peuvent citer le dossier, sont conservés sur disque sans chiffrement et sans limite de durée ; et le modèle de langage utilisé par la preuve de concept est exploité hors de l'Union européenne.

- Le dossier est analysé en mémoire au moyen de bibliothèques locales (`pypdf`, `python-docx`) ; aucun service d'analyse hébergé n'est utilisé (D8).
- Le texte du dossier n'est jamais écrit sur disque pour l'indexation et n'est jamais indexé dans le corpus. Ses paragraphes ne sont des éléments de preuve citables qu'au sein de la session.
- Les sessions résident en mémoire pendant quatre heures au plus et sont purgées sur demande, à expiration et à l'arrêt. Les journaux consignent les identifiants et les tailles des sessions, jamais les instructions, les réponses ou le contenu du dossier.
- La recherche s'exécute sur la machine locale : l'index BM25, les vecteurs bge-m3 stockés dans un fichier LanceDB local et le cross-encoder bge-reranker-v2-m3 (§ 8). Aucun texte du corpus, aucun vecteur et aucune requête ne quitte la machine pour la recherche. Un serveur de base de données vectorielle n'a pas été retenu : à la taille actuelle du corpus, il n'apporte rien sur le plan opérationnel (D5). Les vecteurs sont calculés localement, de sorte qu'aucun texte du corpus n'est envoyé à un service d'*embeddings* hébergé (D4).
- Le serveur n'écoute que sur la machine locale. Il n'y a pas d'authentification.

Une exception documentée. Les brouillons, les modifications du juriste et les réponses validées sont écrits sur disque au format JSON, un fichier par session. Ils peuvent contenir des citations du dossier, ne sont pas chiffrés, n'ont pas de limite de durée de conservation et survivent à la session de quatre heures ; ils ne sont supprimés que lorsque la session est explicitement fermée (06 § 1b).

Le modèle de langage. La génération et l'*entailment check* (§ 9) forment une frontière distincte. Tous deux sont des appels à un modèle de langage externe, et tous deux lui envoient les passages extraits, la question, la fiche du dossier et des extraits du dossier (D11). Le fournisseur est défini par un paramètre de configuration. La preuve de concept utilise DeepSeek, pour lequel un crédit d'API était déjà disponible (D11). Sa politique de confidentialité indique qu'il peut collecter « your text input, voice input, prompt, uploaded files, photos, feedback, chat history, or other content that you provide to our model and Services » ; qu'il conserve les données personnelles « for as long as necessary to provide our Services and for the other purposes set out in this Privacy Policy », finalités qui incluent « to train and improve our technology » ; et que « we directly collect, process and store your Personal Data in People's Republic of China ».[28] Les conditions distinctes de la plateforme API de DeepSeek n'ont pas été examinées, et aucun engagement de non-conservation n'a été trouvé. Le projet est donc limité aux données publiques et à des dossiers

fictifs ou anonymisés. Lorsque le fournisseur est DeepSeek, l'interface demande une confirmation explicite avant la transmission du dossier, et rappelle que seuls des dossiers fictifs ou anonymisés peuvent être soumis (06 § 1).

L'usage réel est prévu sur un fournisseur hébergé dans l'UE (par exemple Mistral) ou sur un modèle local exécuté sur une infrastructure louée (par exemple un serveur GPU H200), sans conservation et avec des conditions de traitement compatibles avec le RGPD et le secret professionnel. C'est l'objectif de conception, et non le réglage par défaut de la preuve de concept (D11 ; `06_confidentiality.md` § 4 ; prévu, non implémenté). Les conditions des fournisseurs cités ici n'ont pas été examinées.

Le règlement européen sur l'IA classe comme systèmes d'IA à haut risque les systèmes « destinés à être utilisés par une autorité judiciaire ou en son nom pour aider les autorités judiciaires à rechercher et à interpréter les faits ou la loi, et à appliquer la loi à un ensemble concret de faits » (considérant 61 ; annexe III, point 8, a)).[30] Heuris est conçu pour les juristes en général. La question de savoir si un déploiement donné relève de cette catégorie dépend de qui l'utilise et pour le compte de qui, et n'est pas examinée ici. L'analyse réglementaire plus large reste en dehors de ce document.

12. Comparaison par catégorie

Heuris ne remplace pas les outils que les juristes utilisent déjà, dans quelque matière que ce soit, et ce document ne prétend pas le contraire.

Les assistants généralistes offrent l'étendue et la rapidité de rédaction. Les bases de données juridiques commerciales offrent ce qui manque entièrement à Heuris : des textes officiels, une sélection éditoriale, une large couverture et, dans certains cas, leurs propres couches d'IA. L'étude citée aux §§ 1 et 5 a constaté que les outils spécifiquement juridiques hallucinaient moins qu'un chatbot généraliste, et que les meilleurs d'entre eux répondaient correctement à une majorité de requêtes.[1]

La preuve de concept applique de bout en bout une méthode plus étroite à un corpus restreint et licite — aujourd'hui un seul corpus de droit du travail. La méthode n'est pas propre au droit du travail ; son extension à d'autres matières est prévue, selon l'accès aux données, et n'est pas implémentée. Une phrase n'est affichée que si elle est rattachée à sa source. La comparaison ci-dessous se fait par catégorie ; aucun produit n'est cité ni évalué.

L'architecture de recherche, de vérification et de traçabilité n'est pas propre au droit social. Passer à une autre matière supposerait de remplacer le corpus ainsi que son adaptateur d'ingestion et de découpage, et d'adapter une couche métier : le schéma de la fiche du dossier (types de litige, rôles des parties et des dates), les instructions données au LLM, les termes d'expansion des requêtes et les libellés de l'interface. La version actuelle est aussi liée au français (racinisation, analyse des dates, instructions) et aux conventions des juridictions belges. Aucune mesure entre matières n'existe : la seule évaluation de la recherche, sur 15 questions rédigées par le concepteur, porte uniquement sur le corpus de droit du travail et ne permet pas de distinguer les modes de recherche (§ 8, § 13). L'extension à d'autres matières reste prévue, selon l'accès aux données. (Source : *documentation du projet (non publique)*, *documentation du projet (non publique)*, *documentation du projet (non publique)*, *documentation du projet (non publique).*)

Tableau 12.1 · Comparaison par catégorie

	Assistant généraliste	**Base de données juridique commerciale**	**Heuris (preuve de concept)**
Couverture des sources	Web ouvert / données d'entraînement	Large, sélectionnée, textes officiels, nombreuses matières	Une matière (droit du travail) : plus de 25 000 décisions, aucun texte officiel
Provenance par phrase	Pas par conception	Variable selon le produit	Exigée, vérifiée aux positions
Lorsque les preuves manquent	Tend à répondre	Variable selon le produit	Supprime les affirmations ou refuse
Lieu de traitement des données	Infrastructure du fournisseur	Infrastructure du fournisseur	Local, sauf l'appel au LLM (§ 11)
Disponibilité commerciale	Oui	Oui	Non

Une preuve de concept ne remplace pas des outils établis ; les deux approches peuvent être complémentaires.

13. Limites de l'étude

Les limites de l'étude sont exposées ci-dessous.

1. **Aucun chiffre d'exactitude.** Aucun taux d'exactitude ni d'hallucination n'est revendiqué. Aucun avocat n'a évalué une quelconque réponse.
2. **Tests en mode simulé.** Les 105 tests automatisés réussissent avec des simulacres, et non avec un modèle réel. L'interface et l'assistant de rédaction (D46–D54) n'ont été exercés qu'en mode LLM simulé.
3. **Exécutions réelles uniques.** Le modèle réel n'a été exécuté que lors d'exécutions uniques sur des dossiers de test fictifs. La variance d'une exécution à l'autre est inconnue. Ces exécutions ne constituent pas un benchmark.
4. **Étape d'application.** Lors des exécutions de test réelles de la dernière version d'ancrage (v4), l'étape d'application a conclu sur 0 des 3 questions juridiques, contre 3 sur 3 pour des questions similaires avec la version précédente (v3). La cause n'a pas été établie.
5. **Scores non calibrés.** Les scores de confiance et les paliers « Lien fort / partiel / faible » sont heuristiques, avec des seuils fixés à la main.
6. **Évaluation de la recherche.** Quinze questions rédigées par le concepteur ne permettent pas de départager les modes de recherche et ne prouvent pas la qualité de la recherche.
7. **Réponses conservées.** Les brouillons et les réponses validées conservés sur disque ne sont pas chiffrés, n'ont pas de limite de durée de conservation, survivent à la session et peuvent contenir des citations du dossier.
8. **Aucun contrôle d'accès.** L'API n'a pas d'authentification ; elle n'écoute que sur la machine locale.
9. **Modèle hors UE.** DeepSeek opère hors de l'UE, et les passages extraits sont envoyés à son API pour la génération (D11).[28] Seuls des données publiques et des dossiers fictifs ou anonymisés peuvent être utilisés.

10. **Licence.** Les conditions de l'éditeur indiquent : « La reproduction des informations textuelles du site est autorisée pour autant que la source soit clairement citée. » et « Toute exploitation commerciale des informations est interdite. »[26] Elles ne disent rien de l'usage professionnel non commercial ; la question de savoir si l'usage au sein d'un cabinet qui facture ses clients constitue une exploitation commerciale n'est pas tranchée par ces conditions. Toute offre payante exigerait un accord avec l'éditeur ou une autre source.
11. **Aucun texte officiel.** Heuris n'explore pas Juportal, la base de données officielle du SPF Justice (robots.txt). Il n'y a ni décision officielle en texte intégral, ni ECLI.
12. **Des commentaires, non des décisions.** Le corpus se compose de commentaires qui citent des extraits du raisonnement ; le *dispositif* est le plus souvent absent. L'interface désigne néanmoins les éléments comme « décision » et le panneau source par le nom de l'éditeur (*documentation du projet (non publique)*), jamais comme « commentaire » ; un changement de libellé est une proposition non encore tranchée, et aucune fiche de décision n'existe à ce sujet.
13. **Quasi-doublons.** Les enregistrements qui ne diffèrent que par la ponctuation survivent au dédoublonnage.
14. **Citations du mode juriste.** Les citations insérées en mode juriste sont exactes mais ne font pas l'objet d'un *entailment check*.
15. **Nom et statut.** Les vérifications de marque et de nom de domaine pour « Heuris » n'ont pas été effectuées. Heuris n'est pas disponible commercialement. Rien dans ce document ne constitue un avis juridique.
16. **Une seule matière.** Seul le corpus belge de droit du travail est implémenté ; tous les chiffres le concernent. Aucune autre matière juridique n'est implémentée.
17. **Aucun suivi d'appel ni de cassation.** Le corpus n'enregistre rien sur l'appel, la cassation ou les développements ultérieurs, ni sur les modifications de la législation appliquée ; un résultat peut concerner une décision réformée par la suite.
18. **L'ancienneté dans le score.** Le score de confiance inclut l'ancienneté (poids 0,10, divisé par deux tous les cinq ans ; *documentation du projet (non publique)*, *documentation du projet (non publique)*). La récence n'est pas l'autorité ; cette pondération est un choix de conception qu'aucune mesure n'étaye.
19. **Biais d'échantillon.** Le corpus de plus de 25 000 décisions provient de commentaires sélectionnés par un seul éditeur, uniquement en français ; la jurisprudence néerlandophone est absente, et la jurisprudence publiée est elle-même une sélection (§ 1.3). Une réponse, ou son absence, ne dit rien de l'état de la jurisprudence.
20. **Secret professionnel et protection des données.** Le secret professionnel et le RGPD[29] sont mentionnés mais non analysés. Ce document n'examine pas à partir de quand un dossier peut être considéré comme anonymisé au sens du RGPD. Aucun dossier client réel ne peut être traité avant la mise en place d'un modèle local ou hébergé dans l'UE, sans conservation (§ 11).
21. **Sens de « vérifié ».** « Vérifié » signifie que l'extrait correspond au texte source, ici un commentaire ; l'*entailment check* est un jugement du modèle et n'a pas été audité par un humain.
22. **Export.** L'export Word ne reprend pas le statut de vérification des citations de la réponse validée.

(Sources : `05_evaluation-protocol.md` ; `04_grounding-verification.md` § 13 ; `06_confidentiality.md` ; `01_data-sources.md` ; fiches de décision D10, D11, D34, D46–D54 ; exécution de pytest au commit `9ceeefe`.)

14. Perspectives et travaux prévus (tous les éléments sont prévus)

- **Autres matières juridiques.** De nouveaux corpus dans d'autres matières où des données licites sont disponibles — prévu, selon l'accès aux données.
- **Évaluation.** Un jeu de questions rédigé par des avocats en droit du travail pour le premier corpus ; des métriques d'ancrage et de refus ; un calibrage du score de confiance par rapport aux jugements d'avocats.
- **Modèle.** Un fournisseur hébergé dans l'UE ou un modèle exécuté sur une infrastructure louée, sans conservation, avant tout dossier réel.
- **Renforcement.** Authentification de l'API ; OCR local pour les dossiers numérisés ; détection des quasi-doublons ; seuil de refus sur les scores du *reranker* ; chiffrement et limites de conservation pour les réponses conservées.
- **Ce document.** Version française — réalisée (`heuris_whitepaper_fr.md`).
- **Licence.** Toute démarche commerciale uniquement avec l'accord de l'éditeur ou une autre source licite.

(Sources : *documentation du projet (non publique)*, lignes prévues ; `05_evaluation-protocol.md` ; `06_confidentiality.md` § 4 ; `01_data-sources.md` § 7.)

ID	Contenu	Source

Figures et tableaux

ID	Contenu	Source
F1	Schéma par défaut : le LLM au centre, les sources en option	*documentation du projet (non publique)* (plan des figures)
F2	*Spine* de Heuris : Sources → Ingestion et dédoublonnage → Passages → Recherche → Vérification en mode *fail-closed* → Juriste	Chaîne de traitement de `00_architecture.md`
T1	Corpus : plus de 25 000 décisions · français · dédoublonnage par empreinte du texte normalisé	`01` §§ 4–5 ; `02` § 5 ; `dedupe_report.md`
T2	Évaluation de la recherche (15 questions rédigées par le concepteur) : recall@5 de 1,00 dans tous les modes ; MRR BM25 1,00 · dense 0,92 · hybride 0,97 · hybride + *reranking* 0,96 — non discriminant	`retrieval_report.md` ; `questions.jsonl`
P1	Le problème (§ 1) : LLM généralistes ≥ 58 % d'hallucination (droit américain) ; outils RAG juridiques 17–33 % (droit américain) ; environ 1,4 million de décisions belges par an ; environ 1 % publiées (estimation) ; sanction d'Anvers 25 000 € ; budget de la justice belge 102,5 € par habitant (2022, au-dessus de la médiane du CdE), inférieur à la moyenne de l'UE (Commission 2025) ; 246 jours en première instance (2023) ; communiqué de presse de la Cour des comptes (10 janvier 2025)	Notes 1, 3, 11, 13, 14, 16–19
T3	Tests automatisés : 105 réussis au commit `9ceeefe`, simulacres uniquement	pytest, réexécuté le 2026-11-04
Tableau 5.1	Exigences R1–R6 et leur statut pour Heuris (appréciés au regard des faits exposés dans ce document)	§ 3, § 5
Tableaux 6.1, 7.1, 8.1, 9.1, 12.1	Comparaisons qualitatives	§§ 6–12

Les illustrations d'interface, si elles sont ajoutées, portent la mention « Illustration d'interface — LLM simulé, données fictives ».

Glossaire

Terme	Sens dans ce document
Passage	Segment d'une source respectueux de sa structure, avec des positions de caractères exactes
Commentaire	Texte de l'éditeur consacré à une décision ; non la décision elle-même

Terme	Sens dans ce document
Référence de la décision	Juridiction (avec la chambre lorsqu'elle est enregistrée), date et numéro de rôle *R.G.*, tels qu'enregistrés, par exemple « C. trav. Liège (div. Liège), 14 janvier 2026, R.G. 2025/AL/438 » ; la chambre n'est pas enregistrée
Affirmation	Un énoncé atomique dans une réponse générée
Verbatim check	Correspondance complète de la citation d'une affirmation aux positions citées (règle : l'extrait doit figurer littéralement dans la source)
Entailment check	Contrôle, par le modèle, que le passage cité étaye l'affirmation
Mode *fail-closed*	En cas d'échec de la vérification, l'affirmation est supprimée ou la réponse refusée ; jamais affichée
Fiche du dossier	Résumé typé et sourcé d'un dossier, validé par le juriste
Lien fait ⟶ règle	Une paire vérifiée formée d'un fait du dossier et d'un passage de commentaire rapportant le motif sur lequel la juridiction s'est fondée
Vérifié	L'extrait correspond au texte source (ici, un commentaire) aux positions citées ; ne prouve pas que la juridiction l'a dit
Tribunal du travail (*tribunal du travail*)	Juridiction du travail de première instance ; rend des *jugements*
Cour du travail (*cour du travail*)	Juridiction d'appel en matière sociale ; rend des *arrêts*
Cour de cassation (*Cour de cassation*)	La juridiction suprême de l'ordre judiciaire belge
Étape d'application	Rapporter une règle au dossier à partir de prémisses vérifiées uniquement ; à défaut, aucune conclusion (« Application au dossier — absence de conclusion », avec les éléments manquants)
Position indicative	Étiquette vérifiée « tend à soutenir / à contredire » ; jamais une conclusion
LLM simulé (mode simulé)	Modèle de substitution déterministe pour les tests et les démonstrations
Illustration d'interface	Une illustration produite en mode simulé avec des données fictives
Modèle de langage (LLM)	Un modèle statistique qui génère du texte ; dans Heuris, il rédige et contrôle, mais ne décide jamais de ce qui est affiché
Retrieval-augmented generation (RAG)	Génération conditionnée par des passages extraits d'un corpus[2]
Hallucination	Chez Magesh et al., une réponse incorrecte, ou dont la citation n'étaye pas la proposition[1]
Attribution	Le fait qu'un énoncé généré soit étayé par une source identifiée (cf. AIS)[20]
BM25	Une fonction de classement fondée sur les mots-clés, utilisée pour l'index creux
Vecteur dense (*embeddings*)	Une représentation numérique du sens d'un passage (ici, bge-m3), utilisée pour la recherche sémantique

Terme	Sens dans ce document
Base de données vectorielle	Un stockage de vecteurs denses interrogé par similarité ; dans Heuris, un fichier LanceDB local
Reciprocal rank fusion (RRF)	Une méthode qui fusionne plusieurs classements en utilisant le rang de chaque élément, avec la constante k = 60
Reranker (cross-encoder)	Un modèle qui renote conjointement une paire question–passage (ici, bge-reranker-v2-m3)
Position (offset)	La position d'un caractère dans un texte source ; les citations sont localisées par leurs positions de début et de fin
Implication (inférence en langage naturel)	La relation par laquelle un texte étaye la vérité d'un autre[24]
Instruction (prompt)	Les instructions et les sources envoyées au modèle de langage lors d'un appel
API	Interface de programmation par laquelle le logiciel appelle le fournisseur du modèle de langage
robots.txt	Un fichier par lequel un site web indique quels agents automatisés peuvent collecter ses pages
ECLI	Identifiant européen de la jurisprudence (European Case Law Identifier)
Registre central (JustJudgment)	Le registre central belge des décisions judiciaires ; non ouvert au public à la mi-2026[13]
Fiche de décision (fiche D)	Une décision d'architecture numérotée du projet, D1–D54 ; voir l'annexe B

Notes

Les sources internes sont indiquées comme documentation du projet (non publique).
Chaque note indique ce que la source étaye.

1. V. Magesh, F. Surani, M. Dahl, M. Suzgun, C. D. Manning, D. E. Ho, « Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools », arXiv:2405.20362, https://arxiv.org/abs/2405.20362 (consulté le 2026-11-04). *Étaye :* une évaluation préenregistrée d'outils commerciaux de recherche juridique fondés sur le RAG ; ils hallucinent encore dans 17–33 % des cas, et les auteurs les jugent « less prone to hallucination than GPT-4 » (§ 5) ; une citation qui n'étaye pas sa proposition compte comme une hallucination ; l'outil le plus performant testé, Lexis+ AI, a répondu correctement à 65 % des requêtes (§ 1, § 5). (Étude portant sur le droit américain ; citée pour la catégorie, non pour le droit belge.)
2. P. Lewis et al., « Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks », arXiv:2005.11401, https://arxiv.org/abs/2005.11401 (consulté le 2026-11-04). *Étaye :* la définition du *retrieval-augmented generation*, et le fait que la provenance des résultats des modèles était un problème ouvert que la technique ne traitait qu'en partie.
3. M. Dahl, V. Magesh, M. Suzgun, D. E. Ho, « Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models », *Journal of Legal Analysis* (2024), arXiv:2401.01301, https://arxiv.org/abs/2401.01301 (consulté le 2026-11-04). *Étaye :* les LLM « hallucinate at least 58% of the time, struggle to predict their own hallucinations, and often uncritically accept users' incorrect

legal assumptions » ; plus de 800 000 requêtes sur GPT-4, GPT-3.5, PaLM 2 et Llama 2, portant sur la jurisprudence fédérale américaine.

4. Z. Xu, S. Jain, M. Kankanhalli, « Hallucination is Inevitable: An Innate Limitation of Large Language Models », arXiv:2401.11817, https://arxiv.org/abs/2401.11817 (consulté le 2026-11-04). *Étaye :* « we formalize the problem and show that it is impossible to eliminate hallucination in LLMs ».
5. A. T. Kalai, O. Nachum, S. S. Vempala, E. Zhang, « Why Language Models Hallucinate », arXiv:2509.04664, https://arxiv.org/abs/2509.04664 (consulté le 2026-11-04). *Étaye :* les hallucinations « persist even in state-of-the-art systems » ; elles persistent « due to the way most evaluations are graded », répondre au hasard en situation d'incertitude améliorant les résultats aux tests.
6. H. Touvron et al., « Llama 2: Open Foundation and Fine-Tuned Chat Models », arXiv:2307.09288, https://arxiv.org/abs/2307.09288 (consulté le 2026-11-04). *Étaye :* tableau 10, parts linguistiques des données de pré-entraînement « en 89.70% » et « fr 0.16% » ; « Most data is in English ».
7. OpenAI, « GPT-4 Technical Report », arXiv:2303.08774, https://arxiv.org/abs/2303.08774 (consulté le 2026-11-04). *Étaye :* la figure « GPT-4 3-shot accuracy on MMLU across languages » (MMLU traduit automatiquement) : anglais 85,5 %, français 83,6 %.
8. S. Singh et al., « Global MMLU: Understanding and Addressing Cultural and Linguistic Biases in Multilingual Evaluation », arXiv:2412.03304, https://arxiv.org/abs/2412.03304 (consulté le 2026-11-04). *Étaye :* « for questions requiring geographic knowledge, an astounding 84.9% focus on either North American or European regions ».
9. J. Myung et al., « BLEnD: A Benchmark for LLMs on Everyday Knowledge in Diverse Cultures and Languages », arXiv:2406.09948, https://arxiv.org/abs/2406.09948 (consulté le 2026-11-04). *Étaye :* « LLMs perform better for cultures that are highly represented online, with a maximum 57.34% difference in GPT-4 ».
10. *Mata v. Avianca, Inc.*, No. 22-cv-1461 (PKC) (S.D.N.Y. 22 juin 2023), Opinion and Order on Sanctions, https://www.nhd.uscourts.gov/sites/default/files/pdf/Mata-v-Avianca-sanctions-order.PDF (consulté le 2026-11-04). *Étaye :* la soumission de « non-existent judicial opinions with fake quotes and citations created by the artificial intelligence tool ChatGPT » ; « A penalty of $5,000 is jointly and severally imposed on Respondents ».
11. M. Joris, « Un avocat condamné pour avoir remis des conclusions fantaisistes rédigées par l'intelligence artificielle », RTBF, 25 décembre 2025, https://www.rtbf.be/article/un-avocat-condamne-pour-avoir-remis-des-conclusions-fantaisistes-redigees-par-l-intelligence-artificielle-11653006 (consulté le 2026-11-04). *Étaye :* l'arrêt de la cour d'appel d'Anvers (amende de 2 500 €, 7 500 € par partie à trois parties, 25 000 € au total ; « jurisprudence inexistante ») ; l'affaire du tribunal de l'entreprise de Gand du 15 décembre 2025. Article de presse ; les décisions n'ont pas été consultées.
12. A. Ockerman, « Arrest Gentse hof van beroep benadrukt noodzaak van zorgvuldigheid bij gebruik van artificiële intelligentie », Orde van Vlaamse Balies, 17 octobre 2025, https://www.ordevanvlaamsebalies.be/nl/nieuws-en-events/arrest-hof-van-beroep-onderstreept-de-noodzaak-voor-advocaten-om-zorgvuldig-om-te-springen-met-artifici%C3%ABle-intelligentie (consulté le 2026-11-04). *Étaye :* des conclusions comportant « een waslijst aan verkeerde citaten uit bestaande wetsartikelen en verwijzingen naar onbestaande rechtspraak en rechtsleer ». Communication d'un ordre professionnel ; l'arrêt n'a pas été consulté.
13. SPF Justice, « Déjà 800 000 jugements disponibles sur Just-on-web » (2026), relayé sur https://blog.forumforthefuture.be/fr/article/deja-800-000-jugements-disponibles-sur-just-on-web/31929 (consulté le 2026-11-04). *Étaye :* « près de 1,4 million de décisions sont prononcées chaque année en Belgique » ; seules les parties concernées et leurs avocats ont accès à leurs propres décisions ; les *arrêts* des *cours du travail* doivent être ajoutés à JustJudgment « dans les prochains mois ».
14. V. Hendrickx, « The Belgian database for case law – path towards legal analytics? », blog du CiTiP de la KU Leuven, 3 février 2023, https://www.law.kuleuven.be/citip/blog/the-belgian-database-for-

case-law-path-towards-legal-analytics/ (consulté le 2026-11-04). *Étaye :* « only a limited number of all judicial decisions are made public (e.g. 1% in Belgium or 4% in the Netherlands) ». Une estimation ; méthode non indiquée.

15. SPF Justice, « Nieuwe JUPORTAL zoekmachine vervangt Jure-Juridat » (décembre 2020), https://justitie.belgium.be/nl/nieuws/andere_berichten/nieuwe_juportal_zoekmachine_vervangt_jure_juridat (consulté le 2026-11-04). *Étaye :* « de individuele rechtscolleges bepalen welke vonnissen en arresten zij anonimiseren en publiceren in JUPORTAL » ; cours du travail (*arbeidshoven*) : « Een selectie van arresten (meer dan 21.000 publicaties) ».
16. SPF Justice, « Le nouveau moteur de recherche JUPORTAL remplace Jure-Juridat » (décembre 2020), https://justice.belgium.be/fr/nouvelles/autres_communiques/le_nouveau_moteur_de_recherche_juportal_remplace_jure_juridat (consulté le 2026-11-04). *Étaye :* un remplacement parce que « la technologie flash [...] ne sera effectivement plus supportée » ; « la seule banque de données publique comportant une sélection de jugements et d'arrêts anonymisés » ; une recherche « via n° ECLI, n° de rôle et dates » et « Options - Sélections Juridiction - Domaine juridique - Type - Année de publication » ; flux RSS. Aucune mention d'une recherche sémantique.
17. CEPEJ (Conseil de l'Europe), *European judicial systems — CEPEJ Evaluation Report, 2024 Evaluation cycle (2022 data): Belgium*, https://rm.coe.int/belgium/1680b1df5c (consulté le 2026-11-04). *Étaye :* « In 2022, the implemented budget of the judicial system is €1,199,563,192, representing 102,5 € per inhabitant and exceeding the CoE median. Conversely, as a percentage of GDP (0,22%), this budget is below the CoE median » ; « the number of pending civil and criminal cases is not available for 1st and 2nd instances ».
18. Commission européenne, *2025 Rule of Law Report — Country Chapter on the rule of law situation in Belgium*, SWD(2025) 901 final, 8 juillet 2025, https://commission.europa.eu/document/download/5fe5bb71-2898-4079-ad5a-19a3ff595a62_en?filename=5_1_63936_coun_chap_belgium_en.pdf (consulté le 2026-11-04). *Étaye :* « remain below the EU average » (budget, juges pour 100 000 habitants) ; « 246 days in 2023 » ; « persistent lack of data on court proceedings » ; « online access to published judgments and machine-readable judicial decisions ».
19. Cour des comptes, « Pilotage de la transformation numérique de la justice par l'État fédéral », communiqué de presse, 10 janvier 2025, https://www.ccrek.be/sites/default/files/Docs/2025_01_PilotageTransformationNumeriqueJustice_Communique.pdf (consulté le 2026-11-04). *Étaye :* « la transformation numérique de la justice n'est pas menée de manière satisfaisante » ; « la numérisation de la justice reste en chantier depuis plusieurs décennies ».
20. H. Rashkin, V. Nikolaev, M. Lamm, L. Aroyo, M. Collins, D. Das, S. Petrov, G. S. Tomar, I. Turc, D. Reitter, « Measuring Attribution in Natural Language Generation Models », arXiv:2112.12870, https://arxiv.org/abs/2112.12870 (consulté le 2026-11-04). *Étaye :* le cadre AIS (« Attributable to Identified Sources ») et sa procédure d'annotation en deux étapes pour juger si des énoncés générés par un modèle sont étayés par les sources sous-jacentes.
21. B. Bohnet et al., « Attributed Question Answering: Evaluation and Modeling for Attributed Large Language Models », arXiv:2212.08037, https://arxiv.org/abs/2212.08037 (consulté le 2026-11-04). *Étaye :* la formulation de la tâche de réponse attribuée aux questions ; un cadre d'évaluation reproductible ; des annotations humaines comme référence et une métrique automatique corrélée.
22. T. Gao, H. Yen, J. Yu, D. Chen, « Enabling Large Language Models to Generate Text with Citations », EMNLP 2023, arXiv:2305.14627, https://arxiv.org/abs/2305.14627 (consulté le 2026-11-04). *Étaye :* le benchmark ALCE et ses métriques automatiques de fluidité, d'exactitude et de qualité des citations.
23. S. Min et al., « FActScore: Fine-grained Atomic Evaluation of Factual Precision in Long Form Text Generation », EMNLP 2023, arXiv:2305.14251, https://arxiv.org/abs/2305.14251 (consulté le 2026-11-04). *Étaye :* la décomposition d'une génération en faits atomiques et la part étayée par une source de connaissances ; un estimateur automatisé utilisant le *retrieval* et un modèle de langage.

24. O. Honovich et al., « TRUE: Re-evaluating Factual Consistency Evaluation », NAACL 2022, arXiv:2204.04991, https://arxiv.org/abs/2204.04991 (consulté le 2026-11-04). *Étaye :* sur 11 jeux de données, « large-scale NLI and question generation-and-answering-based approaches achieve strong and complementary results ».
25. Juportal, robots.txt, https://juportal.be/robots.txt (consulté le 2026-11-04). *Étaye :* `User-agent: * / Disallow: /`, avec `Allow: /` uniquement pour DG_JUSTICE_CRAWLER.
26. Conditions d'utilisation du site de l'éditeur, « Copyright et conditions d'utilisation du site » (consulté le 2026-11-04). *Étaye :* la reproduction des informations textuelles est autorisée à condition que la source soit clairement citée ; toute exploitation commerciale est interdite.
27. G. V. Cormack, C. L. A. Clarke, S. Büttcher, « Reciprocal Rank Fusion Outperforms Condorcet and Individual Rank Learning Methods », SIGIR 2009 — **référence en attente** (page de l'éditeur inaccessible pour vérification). *Étayerait :* la méthode RRF et la constante k = 60.
28. DeepSeek, « DeepSeek Privacy Policy », https://cdn.deepseek.com/policies/en-US/deepseek-privacy-policy.html (consulté le 2026-11-04). *Étaye :* la collecte de « text input, voice input, prompt, uploaded files » ; la conservation « for as long as necessary to provide our Services and for the other purposes set out in this Privacy Policy », y compris « to train and improve our technology » ; le stockage « in People's Republic of China ». Les conditions distinctes de la plateforme API n'ont pas été examinées.
29. Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 (règlement général sur la protection des données), JO L 119 du 4.5.2016, p. 1, https://eur-lex.europa.eu/eli/reg/2016/679/oj (consulté le 2026-11-17). *Étaye :* le cadre juridique de l'UE relatif au traitement des données à caractère personnel mentionné aux §§ 11 et 13 ; non analysé dans ce document.
30. Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 (règlement sur l'intelligence artificielle), JO L du 12.7.2024, considérant 61 et annexe III, point 8, a), https://eur-lex.europa.eu/eli/reg/2024/1689/oj (consulté le 2026-11-17). *Étaye :* les systèmes d'IA « destinés à être utilisés par une autorité judiciaire ou en son nom pour aider les autorités judiciaires à rechercher et à interpréter les faits ou la loi, et à appliquer la loi à un ensemble concret de faits » doivent être classés à haut risque ; cette classification ne s'étend pas aux « activités administratives purement accessoires ».

Les figures et les exemples sont illustratifs. « Heuris » est un nom de travail provisoire. Ne constitue pas un avis juridique.