

Mémoire organisationnelle, récits professionnels et risque de ré-attribution narrative dans les ESSMS

Vers une science de la mémoire organisationnelle assistée par IA

Patrick Danto¹ Yoann Van Messen²

¹ConfidensIA SAS — patrick.danto@confidensia.fr

²Consultant indépendant — secteur social et médico-social

Juin 2026

Statut du document

Document de travail destiné à la discussion scientifique. Les résultats d'estimation proxy sont préliminaires ; la validation humaine constitue la prochaine étape indispensable. Ce papier invite à une collaboration interdisciplinaire : psychologues cognitifs, spécialistes des sciences de l'information, juristes, professionnels ESSMS.

Abstract (English)

Social and medico-social organisations (ESSMS) face a structural paradox: the more they document, the harder it becomes to mobilise what they know. Professional narratives — case reports, incident accounts, arbitration summaries — are the living carriers of organisational memory, yet they concentrate the highest re-identification risk. This paper explores the tension between narrative utility and privacy risk, introduces the concept of *narrative re-attribution* — recognition of a situation without recovering civil identity — and proposes a formal risk model $R(d, \ell)$ decomposing pseudonymisation residual $S(d)$, statistical uniqueness $U_{\text{stat}}(d)$, narrative salience $U_{\text{sal}}(d)$, public source accessibility $A(d)$, and reader prior knowledge $K(\ell)$. A three-profile LLM proxy estimation on 668 reports (AUC = 0.712, bootstrap CI) confirms narrative salience as the dominant predictor and reader-profile effects as the largest source of variance, empirically validating the relational nature of re-identification risk. We conclude with a five-axis research programme toward an AI-assisted organisational memory science for the ESSMS sector.

Keywords: organisational memory, ESSMS, re-identification risk, narrative salience, narrative re-attribution, GDPR, LLM proxy, social work.

Résumé

Les organisations sociales et médico-sociales (ESSMS) sont confrontées à un paradoxe structurel : plus elles documentent, plus l'accès à l'intelligence réelle de l'organisation devient difficile. Les récits professionnels — rapports de situation, comptes rendus d'incidents, synthèses d'arbitrage — sont les porteurs vivants de cette mémoire, mais ils concentrent le risque de réidentification le plus élevé. Ce papier explore la tension entre utilité narrative et risque pour la vie privée, introduit le concept de *ré-attribution narrative* — la reconnaissance d'une situation sans retrouver l'identité civile — et propose un modèle formel $R(d, \ell)$ dont les composantes sont estimées par proxy LLM sur un corpus de 668 rapports. L'estimation confirme la saillance narrative comme prédicteur dominant et les effets de profil lecteur comme principale source de variance. Nous concluons par un programme de recherche en cinq axes vers une science de la mémoire organisationnelle assistée par IA.

Mots-clés : mémoire organisationnelle, ESSMS, risque de réidentification, saillance narrative, ré-attribution narrative, RGPD, proxy LLM.

Contents

1	Le paradoxe documentaire des ESSMS	2
2	Pourquoi les récits sont indispensables	2
2.1	Les organisations apprennent par les récits	2
2.2	La HAS évalue ce que les organisations peuvent mobiliser	3
2.3	Le coût cognitif de la fragmentation documentaire	4
3	La tension centrale : les récits les plus utiles sont les plus risqués	4
3.1	Une hiérarchie de l'utilité narrative	4
3.2	Ce que la généralisation détruit	4
4	Réidentification et ré-attribution narrative : une distinction cruciale	5
4.1	Ce que la jurisprudence couvre	5
4.2	Ce que la jurisprudence ne couvre pas encore	6
4.3	La mémoire événementielle professionnelle	6
4.4	La connaissance du territoire : un vecteur sous-estimé	7
5	Cadre juridique : le standard de preuve	7
6	Un cadre analytique pour mesurer le risque narratif	7
6.1	Le risque comme propriété relationnelle	7
6.2	Les composantes du modèle	8
6.3	Le modèle combiné	9
6.4	La thèse bimodale	10
7	Estimation exploratoire par proxy LLM	10
7.1	Statut et limites de l'estimation	10
7.2	Le corpus	11
7.3	Protocoles d'estimation par profil	11
7.4	Résultats de la régression proxy	12
8	Limites	13
9	Programme de recherche : vers une science de la mémoire organisationnelle assistée par IA	14
	Disponibilité	15
	Références	15

1. Le paradoxe documentaire des ESSMS

Dans beaucoup d'ESSMS, les documents existent. Les procédures existent. Les protocoles existent. Les plans d'action existent. Les évaluations existent. Et pourtant, lorsqu'un événement survient, lorsqu'une situation se tend, lorsqu'un encadrant cherche un précédent ou qu'un évaluateur HAS pose une question simple : l'organisation hésite. Pas parce qu'elle ne sait pas. Parce qu'elle ne parvient plus à mobiliser ce qu'elle sait.

Ce paradoxe est auto-produit. Depuis plusieurs années, les établissements accumulent de la documentation pour répondre à une pression croissante : exigences réglementaires, démarches qualité, obligations de traçabilité, évaluations HAS, exigences RGPD. Le mouvement était nécessaire. Mais un effet pervers est apparu : plus les organisations documentent leur fonctionnement, plus l'accès à l'intelligence réelle de l'établissement devient difficile.

Ce n'est pas un problème de volume. C'est un problème de structure. Les connaissances ne disparaissent pas — elles se fragmentent, se dispersent, s'enfouissent dans des arborescences que personne ne navigue vraiment. Chaque démarche qualité produit son lot de fichiers. Chaque changement de direction laisse ses traces dans des dossiers que personne ne retrouve. Chaque évaluation génère un plan d'action dont on perd le fil six mois plus tard.

La conséquence est visible partout : perte de temps, dépendance à quelques personnes qui savent, dilution de la mémoire collective, cadres transformés en routeurs humains d'information. Dans certains établissements, la continuité organisationnelle repose encore sur quelques professionnels capables de répondre à la phrase devenue banale : “tu sais où on trouve ça ?”

La thèse centrale de ce papier. Les récits professionnels — rapports de situation, comptes rendus d'incidents, synthèses d'arbitrage — sont les porteurs vivants de la mémoire organisationnelle. Ils sont irremplaçables. Et ils sont, de tous les documents produits par les ESSMS, ceux qui concentrent le risque de réidentification le plus élevé. Ce n'est pas un hasard : les documents les plus utiles à la mémoire organisationnelle sont aussi les plus difficiles à pseudonymiser.

2. Pourquoi les récits sont indispensables

2.1. Les organisations apprennent par les récits

La théorie de la mémoire organisationnelle [6, 7, 5] distingue plusieurs formes de connaissance : les procédures codifiées, les routines stratégiques, les bases de données, et ce que Walsh et Ungson appellent les *conserves de rétention*. Parmi ces formes, les

récits d'incidents et de situations occupent une place particulière : ils sont le médium par lequel les organisations mémorisent leurs décisions difficiles, leurs arbitrages, leurs surprises.

Weick [8] a montré que le *sensemaking* organisationnel — la construction collective du sens face à l'ambiguïté — est fondamentalement narratif. Les organisations ne décident pas par application de procédures abstraites : elles construisent des analogies, citent des précédents, racontent des histoires. “*La dernière fois qu'on a eu un cas comme ça...*” — cette phrase, prononcée lors d'une réunion de service, mobilise plus d'intelligence organisationnelle qu'un protocole en quinze points.

Nonaka et Takeuchi [9] ont théorisé la conversion des connaissances tacites en connaissances explicites : les récits sont précisément le lieu de cette conversion. Un travailleur social qui narre une situation difficile lors d'une supervision extériorise un savoir qui était implicite, contextuel, informalisé. Ce savoir-là ne se retrouve pas dans une GED. Quand la personne part, il part avec elle.

Dans le secteur ESSMS, cette dynamique est particulièrement intense. Les situations accompagnées sont complexes, singulières, souvent idiosyncrasiques. Les professionnels développent une expertise contextuelle fine : ils connaissent les familles, les réseaux locaux, les pratiques des juges du secteur, les points de tension entre structures du territoire. Cette expertise est *narrative par nature* : elle s'organise autour d'événements mémorisables, pas autour de règles générales.

2.2. La HAS évalue ce que les organisations peuvent mobiliser

La certification HAS ne porte pas uniquement sur la présence de documents. Elle porte sur leur *appropriation effective* : est-ce que les professionnels savent où trouver les procédures ? Est-ce que les réponses aux situations critiques sont harmonieuses entre équipes ? Est-ce que les enseignements des incidents passés ont été intégrés dans les pratiques ?

Les critères impératifs du référentiel HAS mesurent directement la capacité d'une organisation à faire circuler une intelligence organisationnelle fiable. Une organisation qui a ses procédures dans un logiciel qualité que personne ne consulte n'est pas en meilleure situation qu'une organisation sans procédures — elle est peut-être en moins bonne situation, parce qu'elle croit à tort être en ordre.

Or l'intelligence qui compte lors d'une évaluation HAS n'est pas dans les documents codifiés. Elle est dans la tête des professionnels, dans leurs pratiques quotidiennes, dans leur capacité à répondre à des situations imprévues en mobilisant des précédents pertinents. Cette intelligence est narrative.

2.3. Le coût cognitif de la fragmentation documentaire

Ce phénomène est largement observé dans le secteur : dans de nombreuses structures, les cadres intermédiaires sont devenus des “routeurs humains d’information”, compensant manuellement les défauts du système documentaire. Ce rôle non nommé, non financé, structurellement inévitable dès lors que la mémoire organisationnelle est fragmentée, constitue un coût caché considérable : surcharge cognitive, dépendance à des individus clés, risque de rupture lors des transitions.

La réponse technologique disponible — les systèmes RAG (*retrieval-augmented generation*) — permet d’interroger un corpus documentaire en langage naturel. Mais comme les mêmes auteurs le soulignent : “un mauvais corpus produit simplement une IA confuse plus rapide”. La qualité d’un système de mémoire assistée par IA dépend de la qualité du corpus qu’il interroge. Et les corpus qui seraient les plus précieux — ceux qui portent les récits d’incidents, les arbitrages difficiles, les situations marquantes — sont précisément ceux qui posent les problèmes de vie privée les plus aigus.

3. La tension centrale : les récits les plus utiles sont les plus risqués

3.1. Une hiérarchie de l’utilité narrative

Tous les documents produits par les ESSMS ne portent pas la même valeur mémésique. Les documents de conformité administrative (plans d’action tabulés, procédures codifiées, grilles d’évaluation) sont aisément pseudonymisables : ils contiennent peu d’entités identifiantes et leur valeur informationnelle ne dépend pas de la singularité des situations qu’ils décrivent.

Les récits de situations constituent une catégorie radicalement différente. Ils tirent leur valeur de précisément ce qui les rend difficiles à pseudonymiser :

- la **spécificité contextuelle** : un récit utile décrit une situation concrète, pas un cas générique ;
- la **densité des détails** : la combinaison de l’établissement, du territoire, de la pathologie, de la mesure judiciaire et du parcours familial crée un profil unique ;
- la **mémorisabilité** : les récits qui restent en mémoire professionnelle sont précisément ceux qui sont atypiques, marquants, chargés d’enseignement.

Autrement dit : les récits les plus utiles à la mémoire organisationnelle sont les plus singuliers. Et les plus singuliers sont les plus difficiles à anonymiser.

3.2. Ce que la généralisation détruit

Considérons les trois niveaux de traitement illustrés ci-dessous :

Table 1: Impact du niveau de traitement sur la valeur mémésique et la protection

Niveau	Exemple	Impact
Texte brut	M. Dupont, né le 12/03/1987, suivi au CHRS L’Escale depuis janvier 2023, porte un TSA avec déficience intellectuelle. Père de deux enfants placés à l’ASE de l’Oise.	Valeur mémésique maximale — données personnelles en clair.
Après LaPlume	M. <PER_1>, né le <DATE_NAISSANCE_1>, suivi au <ETAB_CHRS_1> <DATE_1>, porte un <PATHOLOGY_1> avec <PATHOLOGY_2>. Père de deux enfants placés à l’<ORG_ASE_1> <LOC_DEPARTEMENT_1>.	Valeur mémésique préservée — données protégées, sens intact.
Après généralisation	Une personne adulte est suivie dans une structure depuis un certain temps. Elle a des difficultés et une situation familiale complexe.	Valeur mémésique nulle — inutilisable pour l’apprentissage organisationnel.

La version intermédiaire (après LaPlume) est la seule viable : elle préserve la structure sémantique, les relations entre entités et l’intelligibilité métier. La généralisation totale produit un texte inerte du point de vue de la mémoire organisationnelle.

Mais même après pseudonymisation dense, un risque résiduel subsiste. C’est ce risque qui constitue l’objet de ce papier.

4. Réidentification et ré-attribution narrative : une distinction cruciale

4.1. Ce que la jurisprudence couvre

Le droit européen de la protection des données a construit une notion précise de la réidentification : rattacher un texte pseudonymisé à une personne physique identifiable. La séquence jurisprudentielle est claire : le considérant 26 du RGPD pose le test des “moyens raisonnablement susceptibles d’être utilisés” [1] ; l’arrêt Breyer [2] introduit le test de relativité ; OC c/ Commission [3] précise la notion de risque “en réalité insignifiant” ; Cegedim [4] ancre l’ensemble dans le droit administratif français en faisant peser la charge de preuve sur le responsable de traitement.

4.2. Ce que la jurisprudence ne couvre pas encore

Mais les situations les plus fréquentes dans le contexte médico-social ne sont pas des réidentifications au sens strict. Ce sont des **ré-attributions narratives** : le moment où un professionnel lit un rapport pseudonymisé et reconnaît “je sais de quelle situation il s’agit” sans nécessairement retrouver un nom ou une identité civile. Il rattache le texte à une *histoire* qu’il connaît, pas à une *personne* qu’il peut nommer.

Table 2: Réidentification vs ré-attribution narrative

	Réidentification	Ré-attribution narrative
Objet	Rattacher à une personne (identité civile)	Rattacher à une situation connue (histoire, événement)
Mécanisme	Correspondance avec registre d’identités	Reconnaissance par similarité narrative
Fréquence en ESSMS	Rare après pseudo. dense	Probablement plus fréquente
Risque RGPD	Clairement couvert	Zone grise juridique
Acteur typique	Tiers motivé avec registres	Professionnel du service ou du territoire

La question juridique ouverte est précise : la ré-attribution narrative constitue-t-elle une “identification” au sens de l’article 4(1) du RGPD et de la jurisprudence Breyer ? Breyer et OC c/ Commission raisonnent sur des données structurées ou des identifiants techniques — pas sur des récits. La jurisprudence est silencieuse sur les textes narratifs.

4.3. La mémoire événementielle professionnelle

La distinction n’est pas seulement juridique : elle est cognitive. Tulving [10] distingue la mémoire épisodique (rappel d’un épisode spécifique avec son contexte) et la mémoire sémantique (connaissance générale). Pour les professionnels de l’ESSMS, la mémoire pertinente est épisodique : “l’incendie du foyer”, “le dossier des quadruplés”, “le jeune qui a traversé l’Europe seul”.

Cette mémoire *événementielle* est qualitativement différente de la connaissance d’une personne : elle peut permettre la ré-attribution narrative d’une situation pseudonymisée sans passer par l’identification civile. Elle est activée non pas par des identifiants (noms, dates) mais par des éléments narratifs : la combinaison d’un type de structure, d’une géographie, d’une pathologie, d’un événement déclencheur.

Melega et al. [11] ont montré empiriquement que les événements à forte centralité sémantique dans un réseau narratif — ceux qui sont connectés causalement et sémantiquement aux autres événements de l’histoire — sont significativement mieux

mémorisés. Ce résultat a une implication directe pour les récits sociaux : les éléments les plus identifiants (la comorbidité rare, la mesure judiciaire atypique, le parcours migratoire singulier) sont aussi les plus mémorisables, parce qu'ils sont causalement centraux dans la narration.

4.4. La connaissance du territoire : un vecteur sous-estimé

Un professionnel du secteur ne ré-identifie pas seulement par sa mémoire personnelle. Il dispose d'un réseau : collègues de l'ASE du département, directeur du CHRS local, médecin référent, coordinateur MDPH. Un appel téléphonique informel — “tu suis un jeune avec ce profil en ce moment ?” — peut suffire à confirmer une ré-attribution narrative.

Cette *perméabilité du réseau professionnel local* varie selon les secteurs : très élevée en protection de l'enfance dans un petit département (où les professionnels se connaissent tous), plus faible pour un EHPAD en grande métropole. Elle constitue une dimension du risque que ni les standards RGPD ni les outils de pseudonymisation n'abordent actuellement.

5. Cadre juridique : le standard de preuve

Le cadre applicable repose sur une séquence de décisions convergentes. Le RGPD (considérant 26) [1] fonde le test des “moyens raisonnablement susceptibles d'être utilisés”. Breyer [2] introduit le test de relativité. OC c/ Commission [3] précise que le risque doit être “en réalité insignifiant” pour que la pseudonymisation sorte les données du périmètre RGPD. Cegedim [4] fixe le standard de preuve : la charge de démonstration pèse sur le responsable de traitement.

Le cadre LaPlume / ESSMS se distingue structurellement du contexte Cegedim : finalité professionnelle (pas commerciale), aucun stockage (traitement en mémoire), corpus limité par organisation, destinataire professionnel ayant accès au dossier. Mais la charge de preuve demeure : le responsable de traitement doit pouvoir démontrer que le risque résiduel est insignifiant.

C'est précisément ce que le modèle proposé dans la section suivante cherche à opérationnaliser.

6. Un cadre analytique pour mesurer le risque narratif

6.1. Le risque comme propriété relationnelle

La première observation est simple mais fondamentale : le risque de réidentification n'est pas une propriété intrinsèque d'un document. C'est une propriété *relationnelle* — une fonction du document, de son environnement informationnel, et du lecteur.

Un rapport parfaitement pseudonymisé (sans aucune entité résiduelle) peut rester risqué si la situation a été couverte par la presse locale. Un rapport statistiquement rare peut être très peu risqué si aucun professionnel du territoire ne l’a lu. La même situation peut être reconnaissable par un travailleur social du département et opaque pour tout autre lecteur.

$$R(d, \ell) = f(S(d), U(d), A(d), K(\ell)) \quad (1)$$

6.2. Les composantes du modèle

$S(d)$: résidu de pseudonymisation. Entités identifiantes résiduelles dans le texte après pseudonymisation, pondérées par criticité :

$$S(d) = \sum_{c \in \mathcal{C}} w_c \cdot n_c(d) \quad (2)$$

Sur 668 rapports, 198 (29,6 %) ont $S(d) = 0$; 7 (1,0 %) ont un résidu “high” (NIR partiellement masqué, institutions hors gazetteer nommées en clair). Les cas “high” correspondent aux institutions locales non couvertes par les gazetteers et aux codes postaux partiels.

$U(d)$: unicité intrinsèque. Décomposée en trois dimensions indépendantes :

$$U(d) = \alpha \cdot U_{\text{stat}}(d) + \beta_U \cdot U_{\text{sal}}(d) + \gamma \cdot U_{\text{temp}}(d) \quad (3)$$

$U_{\text{stat}}(d) = -\log \hat{P}(\mathbf{x}(d))$ mesure la rareté de la combinaison de traits taxonomiques. Un homme de 45 ans avec TSA, déficience intellectuelle et deux enfants placés est statistiquement peu fréquent dans un corpus ESSMS — mais pas nécessairement mémorable.

$U_{\text{sal}}(d) = \mathbb{E}_n[f_{\text{LLM}}(d, n)]$ mesure la saillance narrative : la capacité du récit pseudonymisé à être reconnu dans un corpus de n situations, indépendamment de la rareté statistique. Un homme qui a parcouru 800 km à vélo pour récupérer son chien, puis vécu trois mois devant un tribunal — ce récit est extraordinairement distinctif non pas parce que sa combinaison de traits est rare, mais parce qu’il est mémorable.

Ces deux dimensions *ne covarient pas nécessairement* : une situation statistiquement rare peut produire un récit ordinaire, et vice-versa. C’est leur indépendance partielle qui justifie leur distinction dans le modèle.

$A(d)$: accessibilité des sources publiques. Présence de traces publiques permettant de recouper la situation :

$$A(d) = \sum_{s \in \mathcal{S}} \lambda_s \cdot a_s(d) \quad (4)$$

Propriété clé : $A(d)$ est *indépendant* du pipeline. Un rapport parfaitement pseudonymisé ($S(d) = 0$) reste exposé via $A(d)$ si la situation a été médiatisée.

$K(\ell)$: **connaissance a priori du lecteur.**

$$K(\ell) \in [0, 1] \quad (5)$$

Table 3: Profils de lecteurs et valeurs de $K(\ell)$

Profil	$K(\ell)$	Caractéristiques
Tiers quel- conque	≈ 0	Aucune connaissance du secteur ni du ter- ritoire
Tiers motivé	$\approx 0,3$	Connaissance du secteur, accès aux sources publiques
Professionnel territoire	$\approx 0,6$	Connaissance du réseau local + K_{rseau}
Professionnel service	$\approx 0,85$	Connaissance des situations en cours
Insider	$= 1$	Hors périmètre (relevant du contrôle d'accès)

Raffinement : $K_{\text{rseau}}(\ell_2)$. Le professionnel du territoire dispose en outre d'un réseau de contacts qui peut activer une reconnaissance par recoupement informel — un appel téléphonique suffit parfois. Cette perméabilité du réseau local est une dimension du risque absente des standards actuels.

6.3. Le modèle combiné

$$R(d, \ell) = \sigma(\beta_0 + \beta_1 S(d) + \beta_2 U_{\text{stat}}(d) + \beta_3 U_{\text{sal}}(d) + \beta_4 A(d) + \beta_5 K(\ell) + \beta_6 U_{\text{stat}}(d) \cdot K(\ell) + \beta_7 U_{\text{sal}}(d) \cdot K(\ell) + \beta_8 A(d) \cdot K(\ell)) \quad (6)$$

Les termes d'interaction sont au cœur de la thèse : β_7 capture que la saillance narrative est spécialement dangereuse pour le professionnel du territoire dont la mémoire événementielle est activée par les récits marquants ; β_8 que les sources publiques ne deviennent dangereuses que pour un lecteur capable de les identifier et de les croiser.

Lecture juridique. Le seuil $R(d, \ell_0) < \varepsilon$ opérationnalise le critère “insignifiant” de OC c/ Commission (2024) et Cegedim (2026). $R(d, \ell_1) < \varepsilon$ répond au standard CEPD sur les données sensibles. Ce modèle est conçu comme un outil de démonstration du respect du standard de preuve.

6.4. La thèse bimodale

Hypothèse principale

Le risque de réidentification après pseudonymisation dense n'est pas uniformément distribué — il est **bimodal**. La grande majorité des situations ESSMS présente un risque résiduel faible, parce que la singularité situationnelle est insuffisante à l'échelle d'un service ou d'un territoire. Une minorité — exposées, singulières ou ayant connu une publicité — présente un risque élevé car $U(d) \cdot K(\ell)$ ou $A(d) \cdot K(\ell)$ crée une clé de recoupement robuste indépendamment de la qualité de la pseudonymisation.

La littérature sur la mémoire épisodique suggère que cette bimodalité n'est pas seulement statistique : elle pourrait être une propriété cognitive. Georgiou et al. [12] ont montré que la reconnaissance narrative fonctionne selon une logique de seuil : en dessous d'un niveau de saillance, le rappel ne se déclenche pas ; au dessus, il se produit avec confiance élevée.

$$R_{\text{cogn}}(d, \ell) \approx \begin{cases} \varepsilon & \text{si } U_{\text{sal}}(d) \cdot K(\ell) < \theta \\ 1 - \varepsilon' & \text{si } U_{\text{sal}}(d) \cdot K(\ell) \geq \theta \end{cases} \quad (7)$$

Si cette hypothèse est confirmée, elle renforce considérablement les recommandations pratiques du cadre.

7. Estimation exploratoire par proxy LLM

7.1. Statut et limites de l'estimation

En l'absence d'annotation humaine, nous proposons une *estimation par proxy LLM* : non pas une mesure du risque réel, mais une mesure de la *vraisemblance de reconnaissance* dans des conditions contrôlées. Le statut est celui d'une exploration méthodologique — un protocole pour étudier le phénomène avant validation humaine.

Trois biais sont documentés :

1. *Circularité partielle — périmètre limité* : GPT-4o calcule à la fois les features (S , U_{sal}) et les cibles proxy (y). Mais le corpus synthétique a été généré par cinq modèles distincts (Mistral Small 3.2, Gemma, Nemotron, Qwen, vague8). La circularité ne concerne donc que la relation features→cibles, pas la structure du corpus. Les corrélations quasi-nulles entre profils ($\text{corr}(y_{\ell_0}, y_{\ell_1}) = 0,009$) invalident l'hypothèse d'un biais systématique GPT-4o partagé.
2. *Biais de persona* : l'interprétation LLM du profil “professionnel du territoire” peut différer du comportement humain. La littérature fournit des bornes [13].

3. *Biais conservateur sur y_{ℓ_1}* : 46 % d'échecs de parsing récupérés à score 0 sous-estiment la faisabilité OSINT.

7.2. Le corpus

668 rapports sociaux issus de deux sources : 454 rapports synthétiques générés par cinq modèles distincts calibrés sur la structure de vrais rapports ; 214 vrais rapports pseudonymisés par des versions antérieures de LaPlume. La variable cible a été estimée sur le sous-corpus low+medium ($S(d) > 0$, 463 docs).

7.3. Protocoles d'estimation par profil

Profil ℓ_0 : tiers quelconque. LLM avec persona ignor-ante du secteur, sans accès externe. Résultat : 50,1 % score-0, 46,0 % score-1, 3,9 % score-2. Les 18 cas score-2 illustrent la bimodalité : “adresse précise + situation logement” ; “hôtel Ibis à Aubervilliers + AED hébergement d'urgence” ; “curatelle renforcée + La Réunion”. Ces combinaisons d'ancres géographiques et situationnelles résistent à la pseudonymisation.

Profil ℓ_1 : tiers motivé (OSINT en deux temps). Contribution originale : à notre connaissance la première opérationnalisation computationnelle du critère “moyens raisonnables” du considérant 26 RGPD. Étape 1 : génération d'un plan OSINT structuré. Étape 2 : évaluation de faisabilité avec champ `rgpd_assessment` explicite par un second LLM. Résultat : 31,5 % faisabilité négligeable, 52,7 % faible, 15,8 % moyenne. Les plans d'investigation génèrent des requêtes Serper naturellement exploitables pour A(d).

Profil ℓ_2 : professionnel du territoire (RAG + réseau). Deux composantes : matching narratif RAG ($n = 10$ candidats stratégiquement sélectionnés par territoire) ; plan d'appels informels avec probabilité de réponse p_{rp} par type d'interlocuteur. Résultat : 12,5 % de match, y_{ℓ_2} moyen = 0,223. Cas les plus élevés : La Réunion + CHRS ($y_{\ell_2} = 0,655$), Martinique + MAS ($y_{\ell_2} = 0,649$). La géographie DOM-TOM agit comme amplificateur : sur un territoire restreint, la singularité sectorielle devient quasi-identifiante.

Composante $A(d)$: presse pseudonymisée. Méthode originale : construire un corpus de situations individuelles réelles de presse, les pseudonymiser avec LaPlume, et tester le matching avec les rapports du corpus. 300 requêtes Serper / 107 snippets filtrés / 6 documents avec $A(d) > 0$ (0,9 %). Les deux documents à U_{sal} le plus élevé (0,52 et 0,54) figurent parmi les 6 matchés : saillance narrative et exposition publique se renforcent mutuellement.

7.4. Résultats de la régression proxy

Table 4: Coefficients de régression logistique (z -scores standardisés, IC 95 % bootstrap)

Feature	$\hat{\beta}$	IC 95 %	OR	Lecture
U_{sal}	+0,750	[+0,383; +1,259]	2,12	Signal dominant
$K(\ell_1)$	+1,694	[+1,225; +2,435]	5,44	Effet couche très fort
$K(\ell_2)$	+1,434	[+0,883; +2,287]	4,19	Effet couche fort
U_{stat}	+0,265	[+0,115; +0,584]	1,30	Positif, robuste
$S(d)$	+0,154	[−0,019; +0,256]	1,17	Positif, incertain
$A(d)$	−0,106	[−0,641; +0,112]	0,90	Non discriminant sur corpus synth.

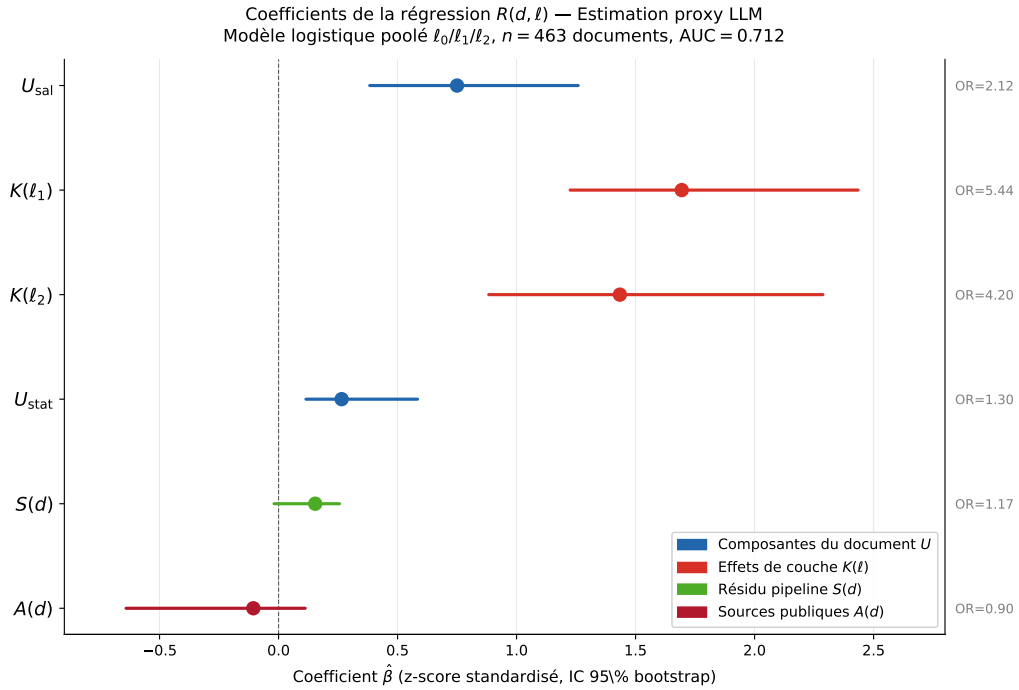


Figure 1: Forest plot des coefficients $\hat{\beta}$ (IC 95 % bootstrap). U_{sal} domine les prédicteurs document-niveau. Les effets de couche $K(\ell_1)$ et $K(\ell_2)$ dominent l'ensemble du modèle — confirmant que le risque est d'abord une propriété du lecteur, pas du document. AUC ROC = 0,712 ; 463 documents, 1 389 lignes empliées.

Enseignements clés. Les effets de couche $K(\ell)$ dominent l'ensemble du modèle ($\hat{\beta}_{K(\ell_1)} = +1,694$) : la propriété relationnelle du risque est confirmée empiriquement. Le risque n'est pas dans le document — il est dans la relation document $\times$ lecteur.

La saillance narrative U_{sal} est le prédicteur document-niveau le plus fort pour le tiers quelconque (ℓ_0). Les interactions $U_{\text{sal}} \times K(\ell_1/\ell_2)$ sont négatives : la saillance guide la reconnaissance du non-spécialiste ; les experts reconnaissent par d'autres canaux (OSINT, réseau).

$A(d)$ est quasi-nul sur corpus synthétique — attendu, sa pertinence réelle ne pourra être mesurée que sur vrais rapports avec situations médiatisées.

L’orthogonalité des trois proxys ($\text{corr}(y_{\ell_0}, y_{\ell_1}) = 0,009$; $\text{corr}(y_{\ell_0}, y_{\ell_2}) = 0,013$; $\text{corr}(y_{\ell_1}, y_{\ell_2}) = -0,036$) confirme que les trois profils mesurent des facettes véritablement distinctes du risque.

8. Limites

Les limites de ce travail sont aussi un inventaire des collaborations que nous cherchons.

1. **Corpus synthétique dominant.** La majorité des documents étant synthétique, les résultats mesurent la cohérence interne du pipeline plus que le risque réel. L’extension à de vrais rapports (sous accord éthique/DPO) est indispensable pour une validation externe. *Collaboration recherchée : établissements ES-SMS, DPO.*
2. **Variable cible construite.** En l’absence d’annotation humaine, la régression porte sur des cibles proxy — pas sur une vérité terrain. Les $\hat{\beta}$ reflètent la cohérence interne du système de proxy, pas une mesure directe du risque. *Collaboration recherchée : psychologues cognitifs, spécialistes de l’évaluation humaine.*
3. **La mémoire événementielle professionnelle n’est pas mesurée.** Nous estimons $K_{\text{seau}}(\ell_2)$ par proxy LLM. Mais la capacité réelle d’un professionnel à reconnaître une situation dépend de son histoire personnelle, de son ancienneté, de sa position dans le réseau local. Ces facteurs ne sont pas capturables par une simulation LLM. *Collaboration recherchée : sociologues des organisations, psychologues du travail.*
4. **La zone grise juridique de la ré-attribution narrative n’est pas tranchée.** La jurisprudence est silencieuse sur les textes narratifs. Notre modèle opérationnalise un risque qui n’est pas encore nommé dans le droit positif. *Collaboration recherchée : juristes spécialisés en RGPD, en droit médico-social.*
5. **Pas de validation de la bimodalité cognitive.** L’hypothèse de seuil (équation 7) est théorique. La tester empiriquement requiert des expériences de reconnaissance sur de vrais professionnels avec corpus contrôlé. *Collaboration recherchée : neuroscientifiques cognitifs, chercheurs en mémoire.*
6. **Pas de séparation train/test.** Le modèle est probablement surajusté. Les $\hat{\beta}$ sont indicatifs, pas définitifs.

9. Programme de recherche :

vers une science de la mémoire organisationnelle assistée par IA

Ce travail est moins une conclusion qu’une invitation. Le phénomène que nous décrivons — la tension entre utilité narrative et risque de ré-attribution — est à la frontière de plusieurs disciplines. Il ne peut pas être traité par une seule d’entre elles.

Nous proposons un programme de recherche en cinq axes :

Axe 1 : Mesurer la saillance narrative

Développer des protocoles de mesure de la saillance narrative dans des textes pseudonymisés, combinant : (a) les métriques LLM proposées ici, (b) des expériences comportementales sur des professionnels réels, (c) des analyses de réseaux narratifs inspirées de Melega et al. [11]. L’objectif est de calibrer le facteur correctif κ tel que $P_{\text{humain}} \approx \kappa \cdot P_{\text{LLM}}$.

Axe 2 : Étudier la mémoire événementielle professionnelle

La mémoire événementielle des professionnels ESSMS n’a jamais été étudiée systématiquement. Combien de situations un travailleur social peut-il reconnaître parmi un corpus de rapports pseudonymisés, après 1 semaine ? Après 1 mois ? Quels traits déclenchent la reconnaissance ? Cette recherche est fondamentale pour calibrer $K(\ell_2)$ empiriquement.

Axe 3 : Comprendre la ré-attribution narrative

La question juridique ouverte — la ré-attribution narrative constitue-t-elle une “identification” au sens du RGPD ? — requiert à la fois une analyse doctrinale et des données empiriques sur la fréquence du phénomène. Ce sont les données qui permettront au droit de trancher.

Axe 4 : Concevoir des techniques de préservation de l’utilité narrative

Au-delà de la pseudonymisation par masquage, explorer des approches alternatives : réécriture guidée (comme Anonpsy [15] mais adaptée aux spécificités ESSMS), généralisation sélective préservant les éléments causalement centraux, systèmes d’alerte sur les situations à risque élevé avant soumission à un LLM.

Axe 5 : Construire des infrastructures de mémoire organisationnelle pour les ESSMS

L’objectif applicatif final est de rendre les corpus de récits professionnels exploitables pour les systèmes RAG de mémoire organisationnelle — en garantissant que le

risque résiduel est documentairement maîtrisé, techniquement réduit et juridiquement défendable. Les organisations qui traiteront la mémoire organisationnelle comme une ressource stratégique — et pas comme une contrainte administrative — seront celles qui tireront le plus de valeur de l’IA. Ce programme de recherche est la condition technique et éthique de cette ambition.

Disponibilité

Pipeline LaPlume et modèle : <https://huggingface.co/jmdanto>. DOI : 10.5281/zenodo.17689395

Références

References

- [1] Parlement européen et Conseil. *Règlement (UE) 2016/679 (RGPD)*, 27 avril 2016. Considérant 26.
- [2] Cour de justice de l’Union européenne. *Patrick Breyer c/ Bundesrepublik Deutschland*, affaire C-582/14, 19 octobre 2016. ECLI:EU:C:2016:779.
- [3] Cour de justice de l’Union européenne. *OC c/ Commission européenne*, affaire C-479/22 P, 7 mars 2024. ECLI:EU:C:2024:215.
- [4] Conseil d’État. *GERS / Cegedim Santé c/ CNIL*, n° 498628, 13 février 2026.
- [5] Ermine, J.-L. (2008). *Management et ingénierie des connaissances : modèles et méthodes*. Hermès Science / Lavoisier.
- [6] Walsh, J. P. & Ungson, G. R. (1991). Organizational memory. *Academy of Management Review*, 16(1), 57–91.
- [7] Stein, E. W. (1995). Organizational memory : Review of concepts and recommendations for management. *International Journal of Information Management*, 15(1), 17–32.
- [8] Weick, K. E. (1995). *Sensemaking in Organizations*. Sage Publications.
- [9] Nonaka, I. & Takeuchi, H. (1995). *The Knowledge-Creating Company*. Oxford University Press.
- [10] Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53, 1–25.
- [11] Melega, G., Samson, K., Lee, H. & Renoult, L. (2025). The semantic structure of events consistently influences episodic memory recall over time in young and older adults. *Scientific Reports*, 15. <https://doi.org/10.1038/s41598-025-26142-6>

- [12] Georgiou, A., Can, T., Katkov, M. & Tsodyks, M. (2023). Large-scale study of human memory for meaningful narratives. *arXiv* : 2311.04742. Publié dans *Learning & Memory*, 32(2):a054043, 2025.
- [13] Bedogni, L., Khan Rumi, S. & Salim, F. D. (2020). Modelling memory for individual re-identification in decentralised mobile contact tracing applications. *arXiv* : 2010.05514.
- [14] Ford, E. et al. (2025). What is the patient re-identification risk from using de-identified clinical free text data for health research? *AI and Ethics*. <https://doi.org/10.1007/s43681-025-00681-0>
- [15] Lim, K.H. & Kim, B.-H. (2026). Anonpsy: A graph-based framework for structure-preserving de-identification of psychiatric narratives. *arXiv* : 2601.13503.
- [16] Sweeney, L. (1997). Weaving technology and policy together to maintain confidentiality. *Journal of Law, Medicine & Ethics*, 25, 98–110.
- [17] Narayanan, A. & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. *IEEE Symposium on Security and Privacy*, pp. 111–125.
- [18] Rocher, L., Hendrickx, J. M. & de Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10, 3069.
- [19] Danto, P. (2026). LaPlume v2 : modèle de pseudonymisation NER pour le secteur médico-social français. Zenodo. DOI : 10.5281/zenodo.17689395