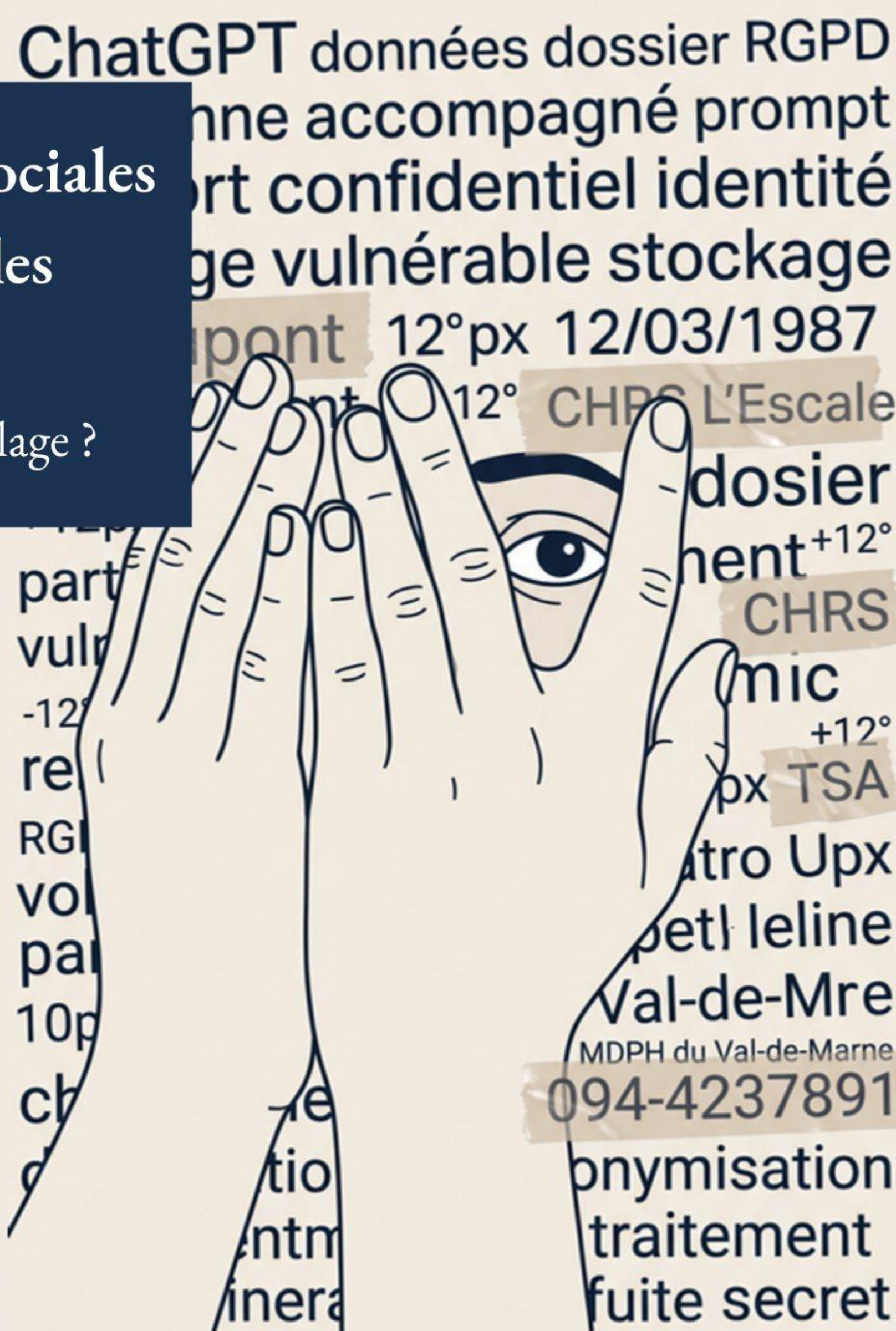
IA & données sociales et médico-sociales

Faut-il choisir entre immobilisme et bricolage ?

Anonymisation
Pseudonymisation
Réidentification



BMSE

Soyons honnêtes : les professionnels utilisent déjà massivement des outils d'IA avec des données personnelles.

Alprs quelles sont les dimensions à prendre en compte pour réguler ?

→ Les données sont-elles stockées quelque part ?
→ Servent-elles à entraîner un modèle ?
→ Qui peut y accéder et les réidentifier ?
→ Le contrat avec votre fournisseur vous protège-t-il réellement ?



Tous les usages ne se valent pas, voici comment les évaluer.

Tous les usages ne se valent pas

Usage	Données stockées ?	Entraîne-ment ?	Contrat DPA ?	Risque
ChatGPT gratuit	Oui	Oui	Non	🔴
ChatGPT/Claude Pro	Selon paramètres	Non	Oui	🟠
Agent IA via API	Non en transit	Non	Oui	🟠
Cloud privé souverain	Non	Non	Oui	🟡
Modèle local	Non	Non	N/A	🟢
Pseudonymisation + API	Données pseudo.	Non	Oui	🟢

Ce tableau dit où vont les données. Pas si le risque résiduel est acceptable.

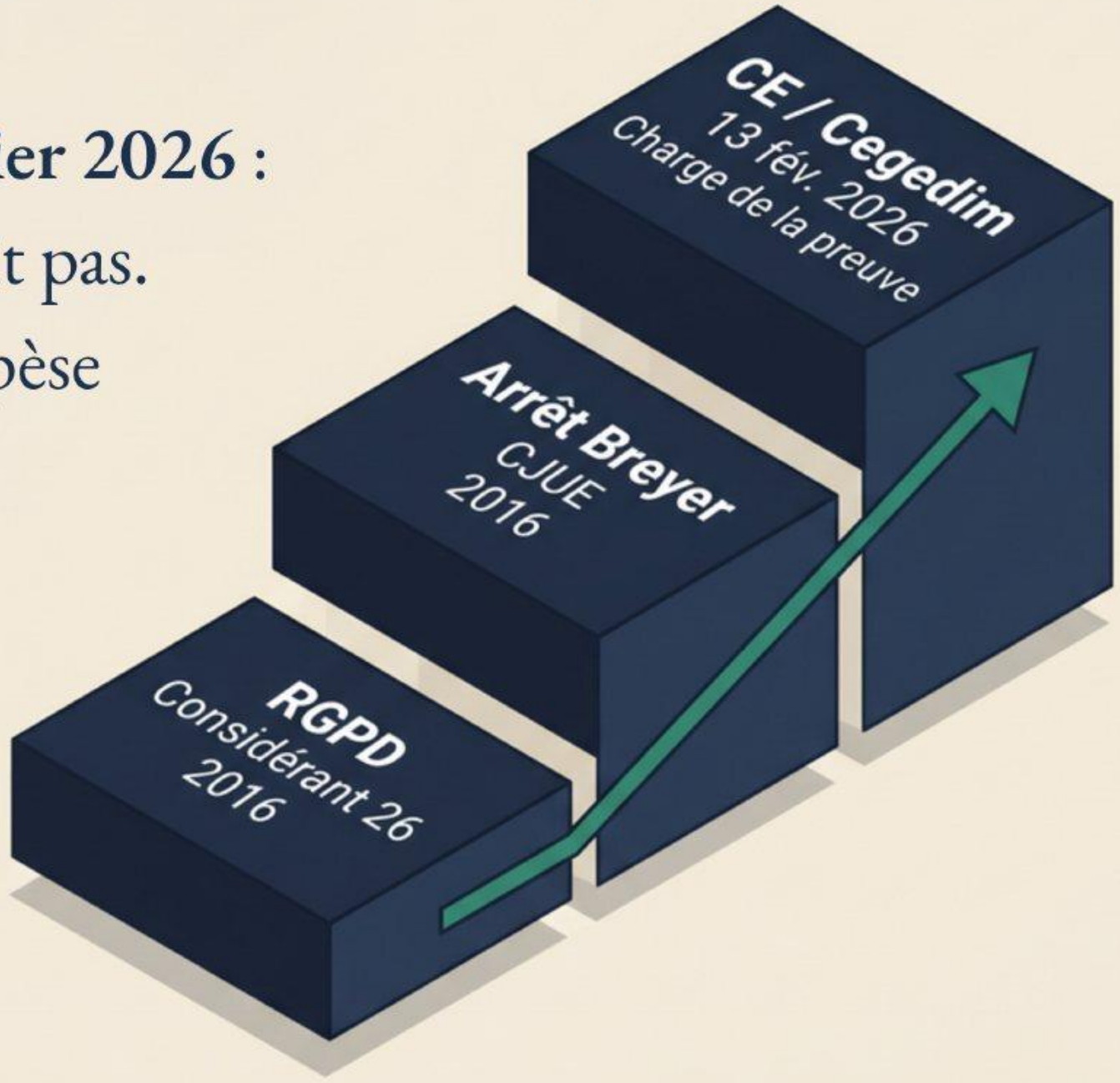
Trois textes à connaître

RGPD, considérant 26 (2016) : le risque de réidentification doit être "en réalité insignifiant" pour sortir du champ

RGPD.Arrêt Breyer, CJUE 2016 : le risque s'évalue du point de vue du destinataire, pas seulement du responsable de traitement.

CE, Cegedim, 13 février 2026 : pseudonymiser ne suffit pas. La charge de la preuve pèse sur le responsable de traitement.

Concrètement :
il faut démontrer que le risque résiduel est maîtrisé.



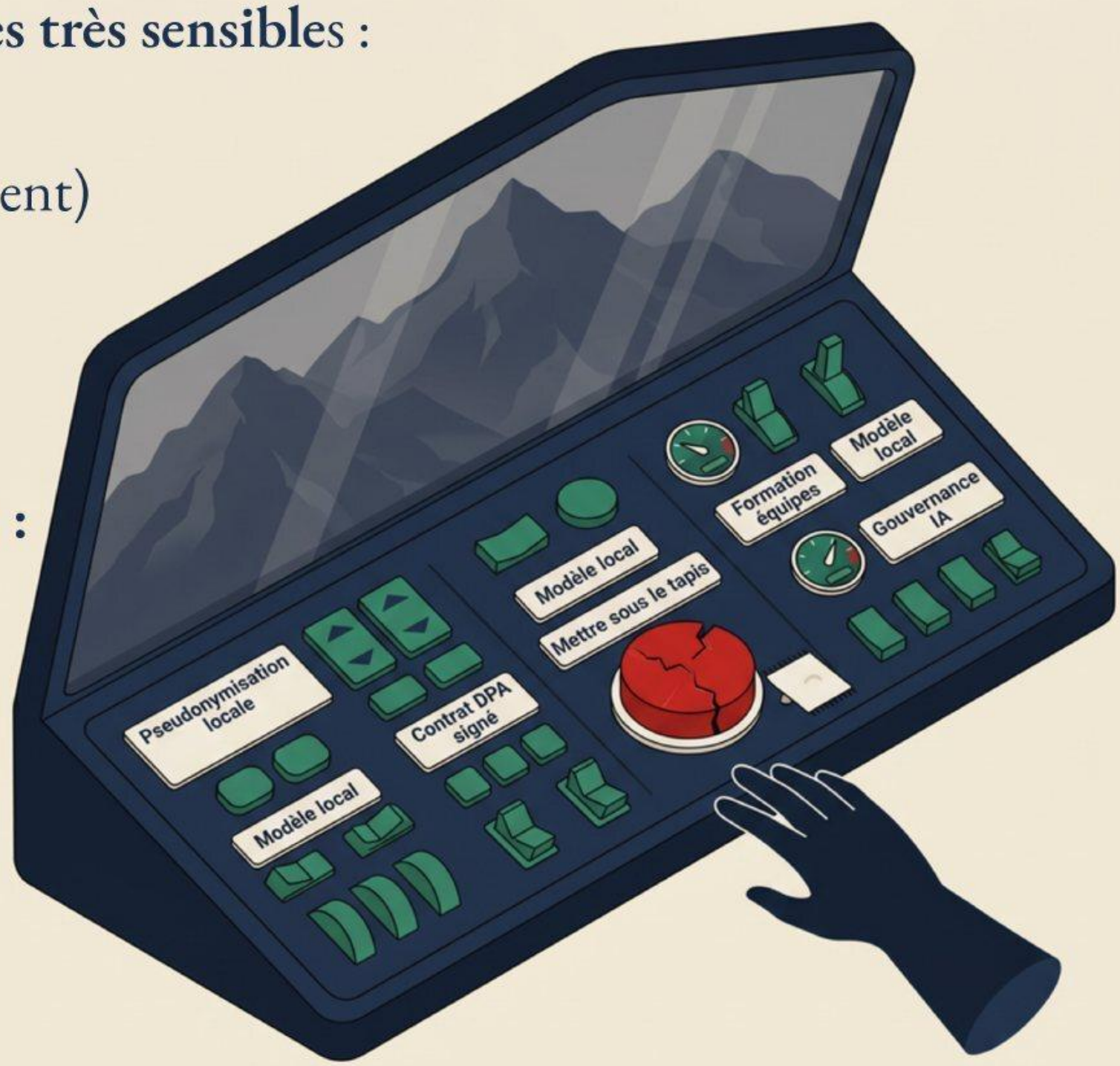
BMSE

Nos conseils en fonction des usages

Tâches bureautiques du quotidien et données peu sensibles :
Copilot avec DPA Microsoft (données dans le tenant M365) peut suffire.

Tâches simples et données très sensibles :
IA locale
(avec un peu d'investissement)

Usages complexes et mémoire organisationnelle (RAG) :
pseudonymisation +
API avec mesure du
risque.
(et c'est là que nos travaux entre en jeu)



BMSE

Nos travaux : pseudonymiser mieux

Les outils génériques détectent 3 à 8 catégories. LaPlume en détecte 101, spécifiques au secteur ESSMS.

Nous avons développé un outils propre au secteur ESSMS, qui a été testé avec un protocole rigoureux (annotation indépendante, double aveugle, arbitrage)

Performance (F1 pondéré) :
96,7 %

Entités à risque critique :
98,4 %

Entités à risque élevé :
97,3%



BMSE

La nouvelle frontière : la mémoire organisationnelle

Pseudonymiser est la condition nécessaire. Pas l'objectif.
L'objectif : rendre exploitables les savoirs que les organisations accumulent sans pouvoir les mobiliser — retours d'expérience, arbitrages difficiles, situations marquantes.

Les systèmes RAG le permettent
Mais les **documents les plus précieux pour la mémoire organisationnelle sont ceux qui posent les problèmes de protection des données les plus aigus.**
C'est la tension centrale que nos travaux cherchent à résoudre.



BMSE

L’intuition du modèle de risque

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

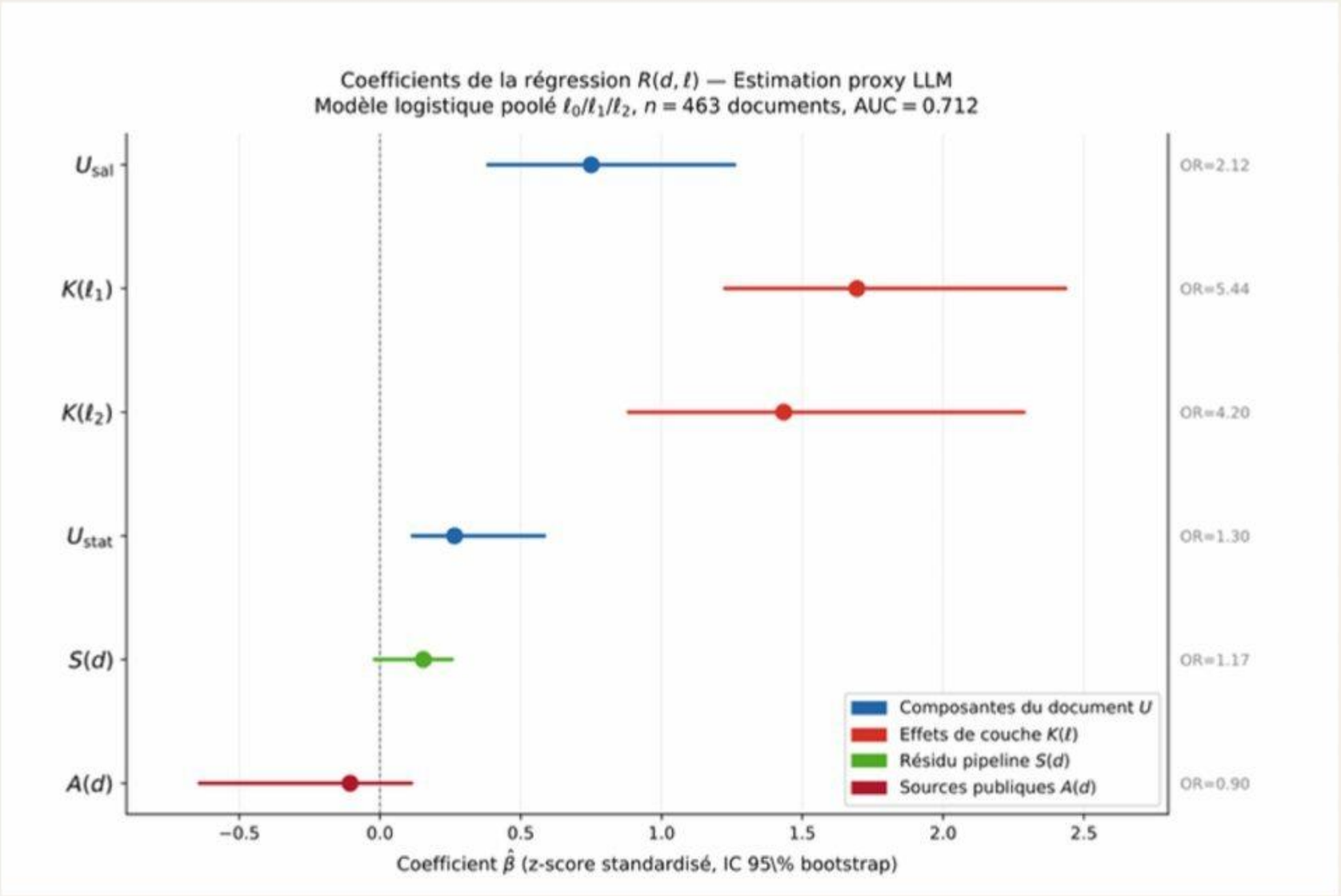
Un document parfaitement pseudonymisé peut rester identifiable.
La réidentification R dépend à la fois du document (d) et du lecteur (l)
Elle est une fonction de :

- S(d), le résidu de pseudonymisation,
- U(d) l’unicité intrinsèque du récit (rareté des combinaisons, saillance narrative)
- A(d) l’accès à des sources publics,
- K(l) la connaissance a priori du lecteur (un quidam, un professionnel du territoire, un tiers motivé).



BMSE

Et on a cherché à mesurer ce risque de réidentification



Sur 463 documents sociaux et médico-sociaux variés (corpus synthétique publié en open-source), on en déduit que

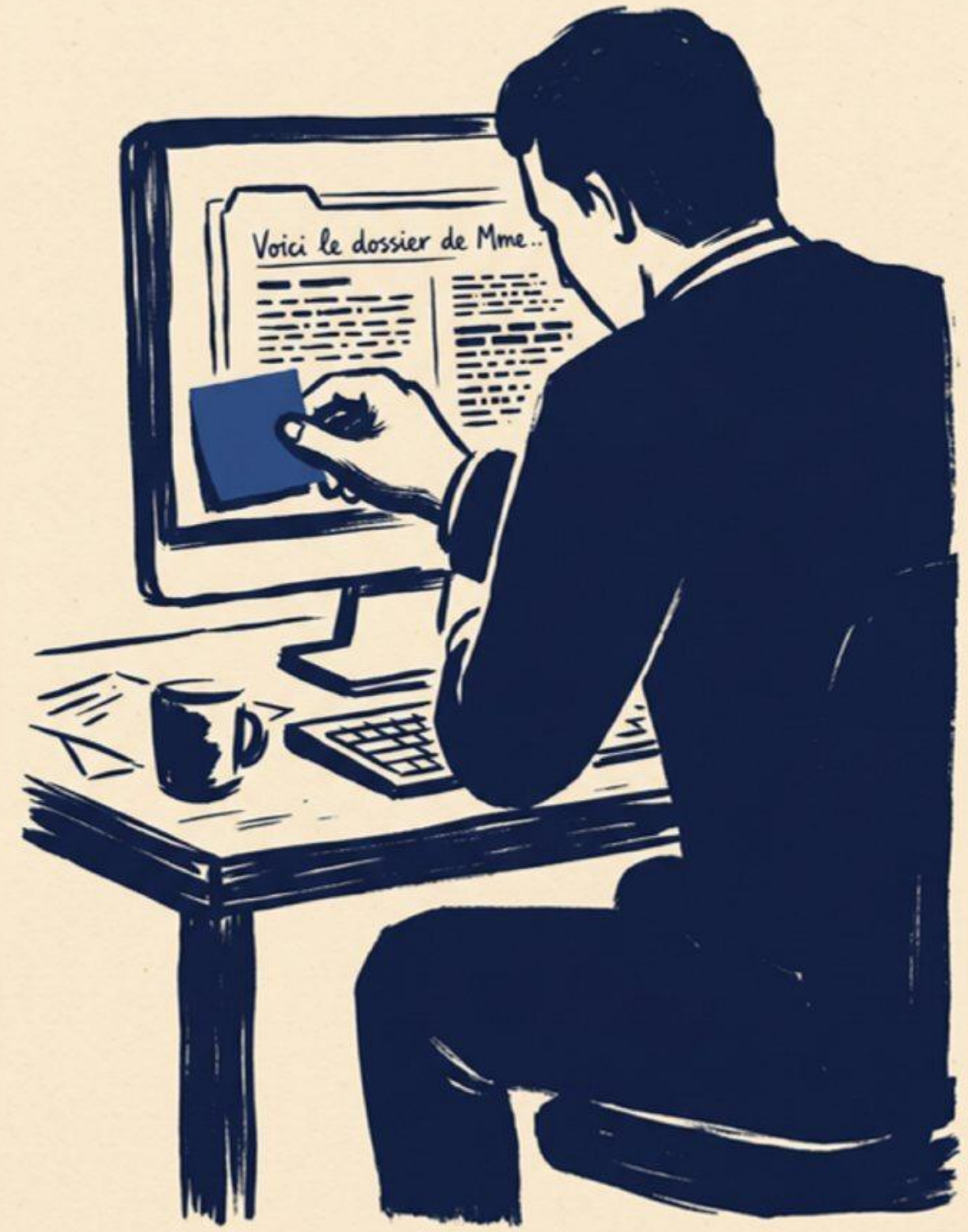
- Ce qui permet la réidentificaiton, c'est surtout la connaissance du lecteur. Un professionnel du territoire a 5x plus de chance de reconnaître une situation
- Les récits mémorables sont 2x plus reconnaissables, ce n'est pas la rareté qui crée le risque, c'est la mémorabilité
- La grande majorité des situations présentent un risque faible. Une minorité concentre l'essentiel du risque

Prochaines étapes

Ces travaux sont exploratoires. Ce qu'il reste à faire est au moins aussi intéressant.

Nous cherchons :
→ Des établissements ESSMS partenaires (sous accord DPO)
→ Des chercheurs en psychologie cognitive, droit RGPD, sciences de l'information
→ Des DPO et directeurs pour partager leur expérience terrain

Preprints sur Zenodo. Modèle et corpus sur Hugging Face et GitHub.



💬 **Comment gérez-vous aujourd'hui la tension entre mémoire organisationnelle et protection des données ?**