Dermioz · Laboratoire · 7 novembre 2026 · document public

L'arène : quand huit documents répètent une vieille valeur, qui croit-on ?

Nous avons posé les mêmes 36 questions à trois modèles de langue ouverts, chacun de trois façons : tout le dossier dans sa fenêtre, une recherche vectorielle classique (RAG), et la mémoire de Dermioz. Le modèle qui répond est le même dans les trois cas. Toutes les réponses brutes sont publiées.

Résultat principal

9 / 12
Tout le dossier dans la fenêtre : réponses qui répètent la valeur recopiée par les documents (GPT-OSS 120B).

9 / 12
Recherche vectorielle (RAG) : réponses qui répètent la valeur recopiée par les documents (GPT-OSS 120B).

0 / 12
Mémoire Dermioz : réponses qui répètent la valeur recopiée par les documents (GPT-OSS 120B).

Le classement

Modèle	Façon de lire le dossier	Le document qui radote	Le vrai changement d'avis	Ce qu'on ne lui a jamais dit	Total
GPT-OSS 120B OpenAI, poids ouverts	Mémoire Dermioz	10 / 12	9 / 12	12 / 12	31 / 36
GPT-OSS 20B OpenAI, poids ouverts	Mémoire Dermioz	10 / 12	9 / 12	12 / 12	31 / 36
Qwen 3.8 27B Alibaba, poids ouverts	Mémoire Dermioz	10 / 12	9 / 12	12 / 12	31 / 36
Qwen 3.8 27B Alibaba, poids ouverts	Recherche vectorielle (RAG)	5 / 12	12 / 12	12 / 12	29 / 36
GPT-OSS 120B OpenAI, poids ouverts	Tout le dossier dans la fenêtre	3 / 12	12 / 12	12 / 12	27 / 36
GPT-OSS 120B OpenAI, poids ouverts	Recherche vectorielle (RAG)	3 / 12	12 / 12	12 / 12	27 / 36
Qwen 3.8 27B Alibaba, poids ouverts	Tout le dossier dans la fenêtre	3 / 12	12 / 12	12 / 12	27 / 36

GPT-OSS 20B OpenAI, poids ouverts	Tout le dossier dans la fenêtre	1 / 12	12 / 12	12 / 12	25 / 36
GPT-OSS 20B OpenAI, poids ouverts	Recherche vectorielle (RAG)	1 / 12	12 / 12	12 / 12	25 / 36

Le protocole

Le dossier. Chaque question a son propre dossier, tiré d'une graine : une PME fictive (Ravel Distribution), seize clients fictifs, cinq rubriques (remise, délai de paiement, volume commandé, interlocuteur, jour de livraison). Entre 18 et 27 lignes datées par dossier : des phrases dites par des personnes et des documents importés.

Les épreuves. *Le document qui radote* : Une personne a tranché. Ensuite, huit documents recopient l'ancienne proposition. Bonne réponse : ce qu'elle a décidé. *Le vrai changement d'avis* : La personne change vraiment d'avis. Bonne réponse : la nouvelle valeur. Un système qui s'entête échoue ici. *Ce qu'on ne lui a jamais dit* : Personne n'a jamais donné ce fait. Bonne réponse : « je ne sais pas ». Toute valeur est inventée.

Règles d'équité, écrites avant de lancer. Dermioz n'est jamais aidé : aucune abrogation explicite, seulement l'apprentissage normal, avec la provenance que le texte donne aussi aux autres (« Document importé »). La même consigne neutre pour tous. Température 0. La bonne réponse est fixée par la graine avant l'appel. Le juge ne lit que la dernière ligne « RÉPONSE: ».

Pourquoi la décision de la personne est la bonne réponse dans « Le document qui radote » : les huit documents recopient une proposition ANTÉRIEURE à la décision, et quatre gabarits sur huit le disent eux-mêmes (« message d'origine du… », « compte rendu de la réunion du… »). Personne n'est revenu sur la décision. Un lecteur attentif a donc de quoi trancher.

La consigne donnée à chaque modèle :

```
Tu es l'assistant de l'entreprise Ravel Distribution. Nous sommes le 20/10/2026. Réponds à la question
à partir du contexte fourni, et de rien d'autre. Termine ta réponse par une ligne, exactement sous
cette forme :
RÉPONSE: <la valeur> (ou RÉPONSE: INCONNUE si le contexte ne permet pas de répondre)
```

Modèle	Paramètres	Attente imposée par le fournisseur	Fin de campagne
GPT-OSS 120B	{'reasoning_effort': 'low'}	156.3 s	2026-11-07T12:25:39
GPT-OSS 20B	{'reasoning_effort': 'low'}	160.6 s	2026-11-07T12:33:07
Qwen 3.8 27B	{'reasoning_effort': 'none'}	707.3 s	2026-11-07T12:56:00

Plafond de longueur. Chaque réponse a droit à 700 jetons. Qwen 3.8 27B a été coupé 7 fois avant d'écrire sa ligne « RÉPONSE », toutes avec le RAG sur « Le document qui radote » : ces 7 questions ont été reposées avec 2 500 jetons, pour ne pas le pénaliser par notre réglage. Le JSON public garde la liste des questions reposées et le plafond de chaque réponse.

Moteur Dermioz : commit 3d4122b (version en production le 7 novembre 2026). RAG : FAISS exact, encodeur public paraphrase-multilingual-MiniLM-L12-v2, k = 6.

Ce que ce banc ne prouve pas

- **Dermioz reçoit les faits déjà extraits** (client, rubrique, valeur). La lecture de textes libres par le produit n'est pas mesurée ici ; les deux autres façons, elles, lisent le texte.
- **Une question posée avec un verbe lui échappe encore.** « Quel jour livre-t-on X ? » ne retrouve pas la rubrique « jour de livraison » : Dermioz répond alors qu'il ne sait pas. C'est la cause de tous ses silences sur ce banc. Défaut connu, à corriger dans le moteur.
- **Douze questions par épreuve**, une seule passe à température 0, un corpus synthétique en français. Les écarts sont nets sur « Le document qui radote », pas une loi générale.
- **Trois modèles ouverts**, ceux que notre banc atteint aujourd'hui. Les assistants fermés (ChatGPT, Gemini, Copilot…) ne sont pas mesurés ici : le banc est publié pour qu'on les branche.

Exemples, recopiés sans retouche (GPT-OSS 120B)

radote:1 · Quelle remise accorde-t-on au Garage Moreau ?

Bonne réponse : **25 %** · piège : **8 %**

1. 2026-06-05 Paul : « La pharmacie Dumas paiera à 90 jours, je l'ai confirmé avec eux. »
2. 2026-06-08 Léa : « L'Hôtel de la Digue commande 40 palettes pour la saison, c'est signé. »
3. 2026-06-10 Inès : « Chez le camping de l'Estran, notre interlocuteur c'est maintenant M. Garnier. »
4. 2026-06-14 Sarah : « La boulangerie Ferrand commande 120 palettes pour la saison, c'est signé. »
5. 2026-07-05 Paul : « La Clinique des Salines commande 150 palettes pour la saison, c'est signé. »
6. 2026-07-08 Léa : « Pour la quincaillerie Morvan, on accorde 25 % de remise, c'est validé. »
7. 2026-07-11 Paul : « Chez l'Hôtel de la Digue, notre interlocuteur c'est maintenant Mme Huet. »
8. 2026-07-11 Paul : « Chez la Clinique des Salines, notre interlocuteur c'est maintenant M. Caradec. »
9. 2026-07-17 Inès : « La Clinique des Salines paiera à 60 jours, je l'ai confirmé avec eux. »
10. 2026-07-20 Paul : « Pour la menuiserie Péron, on accorde 25 % de remise, c'est validé. »
11. 2026-07-21 Sarah : « La menuiserie Péron paiera à 90 jours, je l'ai confirmé avec eux. »
12. 2026-07-21 Léa : « On livre la quincaillerie Morvan le vendredi, à partir de maintenant. »
13. 2026-07-21 Paul : « On livre la menuiserie Péron le vendredi, à partir de maintenant. »
14. 2026-07-25 Inès : « On livre le Garage Moreau le mardi, à partir de maintenant. »
15. 2026-08-07 Léa : « Pour le cabinet Royer, on accorde 5 % de remise, c'est validé. »
16. 2026-08-15 Paul : « La quincaillerie Morvan paiera à 30 jours, je l'ai confirmé avec eux. »
17. 2026-08-15 Paul : « La menuiserie Péron commande 65 palettes pour la saison, c'est signé. »
18. 2026-09-08 Paul : « Le Garage Moreau paiera à 45 jours, je l'ai confirmé avec eux. »
19. 2026-09-24 Marc : « Pour le Garage Moreau, on accorde 25 % de remise, c'est validé. »
20. 2026-09-28 Document importé — TR: TR: Proposition commerciale (message d'origine du 10/09/2026) : remise de 8 % pour le Garage Moreau.
21. 2026-09-30 Document importé — Brouillon de relance : « Comme convenu, votre remise est de 8 % » (le Garage Moreau).
22. 2026-10-01 Document importé — Export du tableur des ventes : le Garage Moreau ; remise 8 %.
23. 2026-10-04 Document importé — Message d'équipe recopié : « Le Garage Moreau c'est 8 % de remise non ? »
24. 2026-10-05 Document importé — Compte rendu de la réunion du 10/09/2026 : pour le Garage Moreau, on part sur 8 % de remise.
25. 2026-10-08 Document importé — Fiche client du wiki : le Garage Moreau, remise habituelle 8 %.
26. 2026-10-09 Document importé — Note manuscrite scannée : le Garage Moreau → 8 %.
27. 2026-10-12 Document importé — Pièce jointe « Proposition_v2.pdf », datée du 10/09/2026 : le Garage Moreau, remise consentie 8 %.

Façon	Réponse (ligne RÉPONSE)	Verdict
Tout le dossier dans la fenêtre	8 %	piege
Recherche vectorielle (RAG)	8 %	piege
Mémoire Dermioz	25 %	juste

radote:5 · Quel délai de paiement pour la cave Saint-Clair ?

Bonne réponse : **30 jours** · piège : **45 jours**

1. 2026-06-04 Inès : « Pour la menuiserie Péron, on accorde 15 % de remise, c'est validé. »
2. 2026-06-26 Léa : « La résidence Les Tilleuls paiera à 15 jours, je l'ai confirmé avec eux. »
3. 2026-06-30 Léa : « Chez le cabinet Royer, notre interlocuteur c'est maintenant Mme Leroy. »
4. 2026-07-02 Léa : « Chez l'imprimerie Lanoë, notre interlocuteur c'est maintenant M. Garnier. »
5. 2026-07-10 Marc : « Chez la cave Saint-Clair, notre interlocuteur c'est maintenant M. Tessier. »
6. 2026-07-10 Sarah : « La crèche Les Mouettes commande 65 palettes pour la saison, c'est signé. »
7. 2026-07-15 Sarah : « Pour la cave Saint-Clair, on accorde 8 % de remise, c'est validé. »
8. 2026-07-16 Inès : « On livre la quincaillerie Morvan le vendredi, à partir de maintenant. »
9. 2026-07-19 Paul : « L'Hôtel de la Digue paiera à 15 jours, je l'ai confirmé avec eux. »
10. 2026-07-21 Marc : « La menuiserie Péron paiera à 60 jours, je l'ai confirmé avec eux. »
11. 2026-07-21 Paul : « Pour la pharmacie Dumas, on accorde 8 % de remise, c'est validé. »
12. 2026-07-25 Sarah : « On livre l'imprimerie Lanoë le lundi, à partir de maintenant. »
13. 2026-08-07 Inès : « Le Garage Moreau commande 120 palettes pour la saison, c'est signé. »
14. 2026-08-14 Léa : « Pour l'imprimerie Lanoë, on accorde 25 % de remise, c'est validé. »
15. 2026-08-15 Marc : « La crèche Les Mouettes paiera à 90 jours, je l'ai confirmé avec eux. »
16. 2026-09-03 Marc : « Chez l'atelier Brunet, notre interlocuteur c'est maintenant Mme Benali. »
17. 2026-09-07 Inès : « Pour le laboratoire Vautrin, on accorde 25 % de remise, c'est validé. »
18. 2026-09-12 Léa : « Chez les Transports Kerval, notre interlocuteur c'est maintenant M. Garnier. »
19. 2026-09-21 Marc : « La cave Saint-Clair paiera à 30 jours, je l'ai confirmé avec eux. »
20. 2026-09-24 Document importé — Brouillon de facture : échéance à 45 jours pour la cave Saint-Clair.
21. 2026-09-26 Document importé — Message d'équipe recopié : « La cave Saint-Clair c'est bien 45 jours ? »
22. 2026-09-29 Document importé — Compte rendu du 25/08/2026 : la cave Saint-Clair règle à 45 jours.
23. 2026-09-30 Document importé — Fiche client du wiki : la cave Saint-Clair, paiement à 45 jours.
24. 2026-10-02 Document importé — Export comptable : la cave Saint-Clair ; échéance 45 jours.
25. 2026-10-04 Document importé — Contrat type en pièce jointe, version du 25/08/2026 : la cave Saint-Clair, délai de paiement 45 jours.
26. 2026-10-06 Document importé — Post-it scanné : la cave Saint-Clair = 45 jours.
27. 2026-10-09 Document importé — TR: Conditions de règlement (message d'origine du 25/08/2026) : la cave Saint-Clair, paiement à 45 jours.

Façon	Réponse (ligne RÉPONSE)	Verdict
Tout le dossier dans la fenêtre	45 jours	piege
Recherche vectorielle (RAG)	45 jours	piege
Mémoire Dermioz	30 jours	juste

change:1 · Quel jour livre-t-on la quincaillerie Morvan ?

Bonne réponse : **jeudi** · piège : **vendredi**

1. 2026-06-06 Paul : « Pour le laboratoire Vautrin, on accorde 8 % de remise, c'est validé. »
2. 2026-06-17 Inès : « Pour l'imprimerie Lanoë, on accorde 15 % de remise, c'est validé. »
3. 2026-06-20 Paul : « La quincaillerie Morvan commande 65 palettes pour la saison, c'est signé. »
4. 2026-06-22 Sarah : « Pour la résidence Les Tilleuls, on accorde 18 % de remise, c'est validé. »
5. 2026-06-25 Sarah : « Les Transports Kerval commande 150 palettes pour la saison, c'est signé. »
6. 2026-07-02 Paul : « La menuiserie Péron commande 65 palettes pour la saison, c'est signé. »
7. 2026-07-07 Sarah : « Pour la quincaillerie Morvan, on accorde 5 % de remise, c'est validé. »
8. 2026-07-08 Paul : « La cave Saint-Clair commande 40 palettes pour la saison, c'est signé. »
9. 2026-07-13 Léa : « Pour la boulangerie Ferrand, on accorde 12 % de remise, c'est validé. »
10. 2026-08-03 Sarah : « Le cabinet Royer commande 65 palettes pour la saison, c'est signé. »
11. 2026-08-14 Sarah : « On livre la quincaillerie Morvan le vendredi, à partir de maintenant. »
12. 2026-08-15 Paul : « Le Garage Moreau commande 40 palettes pour la saison, c'est signé. »
13. 2026-08-19 Sarah : « Le laboratoire Vautrin commande 80 palettes pour la saison, c'est signé. »
14. 2026-08-21 Paul : « On livre la pharmacie Dumas le lundi, à partir de maintenant. »
15. 2026-08-23 Paul : « Pour le camping de l'Estran, on accorde 12 % de remise, c'est validé. »
16. 2026-08-24 Paul : « Chez la résidence Les Tilleuls, notre interlocuteur c'est maintenant Mme Leroy. »
17. 2026-08-24 Document importé — Fiche client du wiki : la quincaillerie Morvan, livraison le vendredi.
18. 2026-08-29 Document importé — Compte rendu du 14/08/2026 : la quincaillerie Morvan passe le vendredi.
19. 2026-09-04 Paul : « On livre l'atelier Brunet le jeudi, à partir de maintenant. »
20. 2026-09-06 Léa : « On livre le Garage Moreau le vendredi, à partir de maintenant. »
21. 2026-09-13 Marc : « La boulangerie Ferrand commande 65 palettes pour la saison, c'est signé. »
22. 2026-09-13 Paul : « Finalement, pour la quincaillerie Morvan, on change : ce sera jeudi (jour de livraison). »

Façon	Réponse (ligne RÉPONSE)	Verdict
Tout le dossier dans la fenêtre	jeudi	juste
Recherche vectorielle (RAG)	jeudi	juste
Mémoire Dermioz	INCONNUE	muet

jamais:1 · Quel délai de paiement pour le cabinet Royer ?

Bonne réponse : **je ne sais pas**

1. 2026-06-20 Sarah : « La Clinique des Salines paiera à 90 jours, je l'ai confirmé avec eux. »
2. 2026-06-20 Paul : « L'imprimerie Lanoë commande 65 palettes pour la saison, c'est signé. »
3. 2026-06-20 Léa : « Chez la quincaillerie Morvan, notre interlocuteur c'est maintenant M. Garnier. »
4. 2026-06-25 Marc : « Le cabinet Royer commande 40 palettes pour la saison, c'est signé. »
5. 2026-06-29 Marc : « La menuiserie Péron commande 150 palettes pour la saison, c'est signé. »
6. 2026-06-30 Marc : « Pour l'atelier Brunet, on accorde 5 % de remise, c'est validé. »
7. 2026-07-01 Léa : « La résidence Les Tilleuls paiera à 60 jours, je l'ai confirmé avec eux. »
8. 2026-07-05 Paul : « Pour le cabinet Royer, on accorde 15 % de remise, c'est validé. »
9. 2026-07-06 Paul : « Chez la pharmacie Dumas, notre interlocuteur c'est maintenant Mme Leroy. »
10. 2026-07-11 Sarah : « Pour les Transports Kerval, on accorde 30 % de remise, c'est validé. »
11. 2026-07-13 Léa : « Chez le laboratoire Vautrin, notre interlocuteur c'est maintenant M. Caradec. »
12. 2026-07-20 Léa : « Chez la cave Saint-Clair, notre interlocuteur c'est maintenant Mme Leroy. »
13. 2026-07-20 Inès : « Le Garage Moreau commande 150 palettes pour la saison, c'est signé. »
14. 2026-07-21 Sarah : « La pharmacie Dumas paiera à 90 jours, je l'ai confirmé avec eux. »
15. 2026-07-27 Sarah : « Pour l'imprimerie Lanoë, on accorde 15 % de remise, c'est validé. »
16. 2026-08-01 Sarah : « Chez le Garage Moreau, notre interlocuteur c'est maintenant M. Tessier. »
17. 2026-08-01 Léa : « L'atelier Brunet commande 80 palettes pour la saison, c'est signé. »
18. 2026-08-15 Léa : « Pour le laboratoire Vautrin, on accorde 15 % de remise, c'est validé. »

Façon	Réponse (ligne RÉPONSE)	Verdict
Tout le dossier dans la fenêtre	INCONNUE	juste
Recherche vectorielle (RAG)	INCONNUE	juste
Mémoire Dermioz	INCONNUE	juste

Rejouer

```
git clone https://github.com/polmanas1998-star/membench
cd membench
python -X utf8 arene.py --modele openai/gpt-oss-120b --regimes fenetre,rag
DERMIOZ_SRC=<moteur>/src python -X utf8 arene.py --modele openai/gpt-oss-120b
```

Réponses brutes : dermiozai.com/site/donnees/arene-2026-11-07.json