
Dossier 07 — Laboratoires et universités

Texte maître : holding-seb-docs@a7df2793

[docs/dossiers-publics/07_LABORATOIRES_UNIVERSITES.md](#)

[elysea.app](#) · contact@elysea.app

Public : équipes de recherche en évaluation des IA, en psychiatrie et santé publique (prévention du suicide), en informatique, en droit du numérique ; institutions suisses (utiles aussi pour Venture Kick et le chèque d'innovation : voir Fo4, Fo5).

Identité : infrastructure cognitive éthique embarquable par SDK. ELYSÉA n'est ni compagnon, ni chatbot, ni coach, ni thérapeute, ni produit de bien-être, ni dispositif médical.

1. Leur problème

- La plupart des bancs ne rapportent pas la significativité : « most benchmarks do not report statistical significance of their results » [S-MES-12].
 - La validité de construit est faible : revue de 445 bancs, motifs qui « undermine the validity of the resulting claims » [S-MES-13].
 - Les juges LLM biaisent les taux s'ils ne sont pas corrigés par sensibilité et spécificité [S-MES-09].
 - Les bancs cliniques récents ne sont pas réannotables : K-Bench garde prompts, exemples du juge et transcriptions non publiés [S-MES-06] ; VERA-MH est ouvert mais limité à l'idéation suicidaire dans son premier volet [S-MES-07] [S-MES-08].
 - Alluminate est mono-tour, centré sur les réponses dangereuses, avec un jeu privé et une notation relative [S-MES-03] [S-MES-04].
-

2. Ce que la campagne ELYSÉA met à leur disposition (ou mettra)

Élément	État	Pièce
Protocole préenregistré, 4 bras (A nu · B garde naïve · C ELYSÉA · D prompt tiers)	Scellé le 28/08, amendé ; rescèlement à venir	elysea-tests-e2e:audits/PROTOCOLE_CAMPAGNE_ - 4BRAS_3TIIRS_SCELLE_2026-08-28.md:3@c9898dd3 · holding-seb-docs:dossiers/PROTOCOLE_TEL_QU_ - EXECUTE.md:28@8560b37
Sur-refus mesuré (bancs fp70, neg50)	Défini	elysea-tests-e2e:audits/PROTOCOLE_CAMPAGNE_ - 4BRAS_3TIIRS_SCELLE_2026-08-28.md:323@ - c9898dd3
3 juges de 3 familles, aucun Anthropic ni OpenAI	Configurés	elysea-tests-e2e:scripts/run-campagne- - 4bras-2026-09-04.mjs:232-234@c9898dd3
Validation des juges	Juges mécaniques et panel validés sur des jeux tenus à l'écart (R-70, proposé, branche de e2e #517) ; validation juge/humain D23 assurée par l'annotation de Sébastien après le pilote	elysea-tests-e2e:audits/AMENDEMENT_R70_ - SENS_DECOUPE_STRUCTURELLE_2026-10-01.md:46@ - 67316397
Réconciliation des changements post-scellement	Projet ; 9 changements sans pièce de décision	holding-seb-docs:dossiers/RECONCILIATION_ - PIECES_CAMPAGNE_2026-09-30.md:6@8560b37
Résultats	à mesurer — campagne en cours	holding-seb-docs:depot-zenodo-2026-09-05/ - RESULTATS_CAMPAGNE.md:2@8560b37
Dépôt Zenodo à DOI	Procédure écrite ; aucun DOI	holding-seb-docs:depot-zenodo-2026-09-05/ - GESTES_ZENODO_SEBASTIEN.md:37@8560b37
Banc public rejouable, annotations publiées, réannotation ouverte	Décidé par Sébastien le 01/10 ; aucune pièce	—

3. Ce que la campagne doit prouver pour être recevable (état de l'art)
1. Cible de mesure et définitions écrites avant la mesure [S-MES-01] [S-MES-13].
 2. Jeu caché ; règle public/caché documentée [S-MES-01] [S-MES-03].
 3. Accord juge/humain comparé à l'accord humain/humain [S-MES-08] ; juge figé [S-MES-06].
 4. Taux corrigés du biais du juge, intervalle incluant la calibration [S-MES-09].
 5. Erreurs-types regroupées, différences appariées entre bras, puissance préenregistrée [S-MES-10].
 6. Sur-refus et latence à côté de la sécurité [S-MES-11].
 7. Code, échantillon stratifié de transcriptions, annotations publiés [S-MES-01].
 8. Auto-évaluation BetterBench publiée [S-MES-12].

4. Objections fréquentes et réponses

Objection	Réponse
<i>Vous êtes juge et partie.</i>	Oui, c'est pourquoi les annotations seront publiées et la réannotation ouverte (décision du 01/10). Un laboratoire peut réannoter et publier un désaccord.
<i>Vos juges sont-ils validés ?</i>	Les juges mécaniques et le panel de 3 juges sont validés sur des jeux tenus à l'écart (amendement R-70, statut proposé, branche de c2e #517 : eLysea-tests-e2e:audits/AMENDEMENT_R70_SENS_DECOUPE_STRUCTURELLE_2026-10-01.md:46@67316397) ; validation juge/humain D23 assurée par l'annotation de Sébastien après le pilote. Le protocole refuse le tir sans cette validation.
<i>Le jeu sera contaminé.</i>	Le jeu caché ne sera publié qu'après la mesure ; ce que publie le banc public reste à définir avec les labos.
<i>Clinique : qui a validé la frontière orientation / écoute ?</i>	Aucun clinicien externe aujourd'hui. C'est l'objet d'une collaboration (section 5).

5. Collaborations proposées (aucune n'existe)

Collaboration	Apport du laboratoire	Apport d'ELYSÉA
Réannotation indépendante d'un échantillon	Annotateurs qualifiés (cliniciens pour la crise)	Accès aux transcriptions anonymisées et à la grille
Audit statistique du protocole	Revue Miller / Lee / NIST	Protocole, code, données
Co-publication	Analyse, rédaction	Données, banc rejouable
Projet financé	Partenaire de recherche (chèque Innosuisse Fo4, Venture Kick Fo5)	Porteur

6. Conformité

Données de conversation à risque (santé mentale) : RGPD art. 35, AIPD faite et non signée (holding-seb-docs:docs/-juridique/AIPD_ELYSEA_2026-09-30.md:20@8560b37). **Décision** : seules des transcriptions de bancs (cas synthétiques) sont partagées ; jamais de données d'utilisateurs réels.

7. Offre

Gratuite. Aucun paiement entre ELYSÉA et un laboratoire.

8. Prochain pas concret

Après le scellement : proposer par écrit à une équipe suisse d'évaluation ou de prévention du suicide une réannotation d'échantillon. Aucun contact par cette mission.

Sources du dossier

S-MES-01 · S-MES-03 · S-MES-04 · S-MES-06 · S-MES-07 · S-MES-08 · S-MES-09 · S-MES-10 · S-MES-11 · S-MES-12 · S-MES-13