Tester des réponses IA sur des cas limites

Mesurer les réussites, échecs et variations avec des jeux d’essai définis.

Huit demandes fictives, incluant ambiguïté, donnée absente et consigne contradictoire. Jeu de tests annoté et synthèse des résultats.

Cas et données fictifs. Supports modifiables, sans programmation.

Une méthode en quatre étapes

1

Préparer

2

Évaluer

3

Synthétiser

4

Étendre

Justifier chaque note ; ne pas généraliser huit cas à tous les usages possibles.

Choisir des limites qui éprouvent les règles

Un cas limite révèle ce qui se passe lorsqu’une information manque, qu’une consigne se contredit ou qu’une convention est ambiguë. Le jeu de huit demandes est entièrement fictif. Il porte sur dates, calculs, confidentialité, longueur, taux et sources. Définissez ce qu’une réponse recevable devrait faire avant d’attribuer une note. L’objectif n’est pas de piéger un outil par une formulation sans usage, mais de vérifier des comportements qui comptent pour le travail.

Expliquer la règle, puis la vérifier sur les pièces du cas.

Conserver l’incertitude lorsqu’elle est nécessaire

Une invitation sans lieu ne doit pas recevoir une adresse inventée. Une date écrite 03/04 peut demander de préciser la convention. Une consigne imposant au maximum cent mots et exactement deux cents ne peut pas être satisfaite telle quelle. Une réponse utile peut donc demander une clarification ou annoncer la contradiction. Évaluez ce comportement par rapport au besoin, plutôt que de récompenser systématiquement une réponse qui produit du texte malgré tout.

Expliquer la règle, puis la vérifier sur les pièces du cas.

Noter chaque cas avec une justification

La grille du dossier attribue 2, 1 ou 0 selon la fidélité et le besoin de correction. Les notes de référence sont 2, 0, 1, 2, 0, 2, 1, 0. La somme vaut huit points sur seize. Une réponse prudente mais trop affirmative sur une donnée absente peut recevoir une note intermédiaire. Documentez les variantes acceptables afin qu’un autre évaluateur puisse comprendre une différence de jugement sans modifier les critères après coup.

Expliquer la règle, puis la vérifier sur les pièces du cas.

Ne pas extrapoler le petit échantillon

Huit réponses préparées ne représentent ni tous les usages ni la variabilité d’un service réel. Le rapport distingue observations du jeu, erreurs critiques et hypothèses de tests futurs. Pour une mesure réelle, il faudrait conserver versions, conditions et répétitions pertinentes. Cette séance reste réalisable hors connexion : les productions fournies sont des simulations rédigées pour l’exercice. Aucune note de performance commerciale ne doit en être déduite.

Expliquer la règle, puis la vérifier sur les pièces du cas.

Un autre exemple pour raisonner

Répondre « le lieu n’est pas fourni » peut être plus utile qu’une invitation complète comportant une adresse inventée. Le critère porte sur la fidélité et l’usage, pas seulement sur la quantité de texte produite.

Une note intermédiaire doit être expliquée. N’annoncez pas une mesure réelle à partir de productions préparées. Gardez les erreurs critiques visibles même si la somme semble acceptable.

Travail pratique 1 et 2

Préparer
Lire les huit demandes et définir la conduite attendue sans regarder la note du corrigé.

Évaluer
Attribuer une note et une justification à chaque réponse fournie. Distinguer fait faux, omission et incertitude mal traitée.

Conserver les sources, noter les preuves et garder les inconnues.

Travail pratique 3 et 4

Synthétiser
Calculer la somme et présenter les familles de défauts. Identifier les erreurs critiques sans les masquer dans le total.

Étendre
Proposer deux nouveaux cas et une méthode de répétition, sans déclarer ces essais réalisés.

Conserver les sources, noter les preuves et garder les inconnues.

Critères de réussite

1. Huit cas annotés.
2. Notes justifiées et somme exacte.
3. Incertitudes et contradictions traitées.
4. Aucune généralisation injustifiée.

Justifier chaque note ; ne pas généraliser huit cas à tous les usages possibles.

Vérifier sa compréhension

1. Que faire d’un lieu absent ?

2. Quel est le total de 100 et 80 ?

3. Cent mots maximum et exactement deux cents : quel constat ?

Le questionnaire complet est dans les supports. Correction avec le formateur.

Sources et portée

CNIL — IA générative et vie privée
https://www.cnil.fr/sites/default/files/2026-05/affiche_ia_generative_et_vie_privee_fr.pdf

Microsoft Support — Fonctions Excel
https://support.microsoft.com/fr-fr/excel/excel-functions-by-category

Consultation du 12 novembre 2026. Les essais humains et Microsoft Office restent à relever.