Évaluer à l'ère de l'IA générative

D'après l'épisode de « Rendez-vous en terre digitale » enregistré avec Rochane Kherbouche, responsable de la cellule numérique de Bruxelles Formation, animé par Clément Cahagne et Olivier Legrand. Rochane publie en 2026 un ouvrage sur l'évaluation à l'ère de l'IA aux éditions La Chronique Sociale, préfacé et validé scientifiquement par la professeure Christelle Lison (Université de Sherbrooke).



Clément Cahagne — pauuse
Ingénierie pédagogique & IA appliquée à la formation
linkedin.com/in/clementcahagne clement.cahagne@pauuse.fr

Écouter l'épisode : Les voix du digital learning — Rochane : l'évaluation à l'heure de l'IA générative

Pourquoi ce document

L'évaluation est le parent pauvre de la pédagogie. On parle volontiers de méthodes, d'outils, d'animation — beaucoup moins de la façon dont on vérifie que quelque chose a été appris. Or c'est le point que l'IA générative a le plus violemment déstabilisé.

Ce document sert deux usages : le lire, et le donner à une IA. En base de connaissance, il lui fournit un cadre d'analyse à la place de ses généralités sur « comment évaluer avec l'IA ». C'est ce qui distingue un partenaire de réflexion d'un générateur de quiz.

1 L'état des lieux

Rochane a interrogé plus de mille professionnels via un questionnaire de 21 questions diffusé sur sa communauté IA et sur LinkedIn : technopédagogues, formateurs pour adultes, ingénieurs pédagogiques, enseignants du supérieur, responsables et directeurs de formation, en France, Belgique, Suisse, Québec, Afrique du Nord et Afrique francophone.

Deux résultats structurent tout le reste.

- **Deux répondants sur trois** estiment leurs évaluations vulnérables à un usage non détecté de l'IA. Pas « améliorables ». Vulnérables.
- **Une proportion comparable** désigne l'absence de formation à l'IA comme l'obstacle principal. Pas l'outil, pas la technologie : la formation. Les professionnels ont détecté le problème, mais ils n'ont pas de démarche pour le résoudre.

L'écart le plus frappant se situe entre ce que les gens disent faire et ce qu'ils font réellement : beaucoup décrivent des dispositifs qu'ils envisagent, très peu ont une démarche structurée.

Le vrai problème n'est pas la triche

Trois formats dominent l'évaluation : la dissertation à domicile, le QCM standard, l'étude de cas générique. Ils partagent un défaut commun — ils peuvent être entièrement délégués à une IA générative en trente secondes, avec un résultat parfois indiscernable d'un travail humain sérieux.

Ce n'est pas une question de triche, c'est une question d'alignement. Nos évaluations ont été conçues pour un monde où produire du texte demandait un effort cognitif. Ce monde a changé. L'effort cognitif n'est plus là où on le mesurait.

La question d'audit

Une seule question, à poser sur chacune de vos évaluations :

> **L'IA peut-elle faire ce travail à la place de l'apprenant sans que cela se voie ?**

Si la réponse est oui, la consigne est vulnérable. C'est le point de départ de toute la démarche.

Pourquoi les détecteurs sont une impasse

Entre mars et décembre 2023, la question posée partout a été : « quel logiciel pour détecter ChatGPT ? » Le verdict est sans appel. Des textes écrits par des humains sont classés comme artificiels — la Déclaration d'indépendance des États-Unis a été détectée à 100 % comme générée par IA. Et des productions entièrement générées passent sous le radar après une simple reformulation.

Ces outils promettent une certitude qu'ils ne peuvent pas tenir. La réaction, quand on l'explique, est souvent l'incrédulité ou l'agacement : les gens cherchent un rempart, ils veulent qu'on leur dise que tout ira bien. Mais le problème n'a jamais été ChatGPT. C'est notre rapport à l'évaluation elle-même.

Les trois postures

Rochane propose une métaphore : l'IA est une rivière qui monte. Trois réactions sont possibles.

Posture résistante — construire une digue

Interdire, surveiller, détecter. C'est le premier réflexe, et il est parfois légitime.

- **Quand elle a du sens.** Quand le produit lui-même est la compétence évaluée. Résumer un texte est une compétence : si c'est elle qu'on évalue, l'IA n'a pas sa place, point. Quand un jugement ne peut pas être délégué : un soignant ne délègue pas son jugement clinique, il y a des vies en jeu. Et pour l'évaluation des savoirs, sur laquelle nous reviendrons.
- **Sa limite.** La rivière finit toujours par contourner ou déborder. Et pendant qu'on construit la digue, on ne réfléchit pas à habiter autrement au bord de l'eau.
- **Sa conséquence pratique.** Une évaluation en posture résistante suppose souvent un retour au présentiel ou à l'oral. On ne peut pas interdire l'IA à distance sans faire une confiance totale.

Posture collaborative — nager avec le courant

L'IA est autorisée, et c'est la façon dont l'apprenant navigue avec elle qui est évaluée : ce qu'il accepte, ce qu'il rejette, ce qu'il reformule. On évalue le discernement, pas la production.

- **L'outil central : le processus documenté.** Les traces de la pensée plutôt que le produit final. L'apprenant explique ses choix — pourquoi cette suggestion acceptée, pourquoi celle-là rejetée, qu'est-ce qui a été modifié et pour quelle raison. C'est là que la compétence disciplinaire se révèle.
- **Sa contrainte.** Concevoir un portfolio documenté représente 15 à 20 heures d'investissement initial, et la première correction prend plus de temps qu'un devoir classique. Le retour sur investissement s'installe progressivement : la deuxième année est déjà moins lourde, la troisième franchement rentable.

Posture formative — l'IA comme sparring partner

Ni adversaire à neutraliser, ni outil à tolérer : un partenaire intellectuel qui fait progresser.

L'exemple développé par Rochane est une application de dialogue évaluatif socratique, construite sur Google AI Studio, où l'IA pose des questions à l'apprenant sans jamais lui donner la réponse. C'est inconfortable — et c'est précisément comme cela que l'apprentissage se produit.

> **Elles ne s'excluent pas.** Les trois postures cohabitent dans un même dispositif. Un oral en posture résistante pour la validation finale, un portfolio collaboratif pour suivre la démarche, un dialogue socratique en posture formative pour la préparation. Tout dépend de ce qu'on veut mesurer.

Savoirs, savoir-faire, savoir-être

Un point discuté longuement dans l'épisode, parce qu'il ne fait pas consensus.

- **Ce sur quoi tout le monde s'accorde.** Les évaluations de savoir-faire et de savoir-être sont augmentées par l'IA : elle ouvre l'accès au discernement et à la réflexivité, difficiles à observer auparavant. Les évaluations de savoirs, elles, basculent souvent vers la posture résistante et vers le présentiel.
- **La règle de fond.** On peut autoriser l'IA, on peut l'interdire, on ne peut pas l'imposer. Une évaluation mesure une compétence, pas l'usage d'un outil. Obliger à utiliser l'IA pour produire un rendu, c'est risquer d'évaluer la maîtrise de l'outil au lieu de la compétence disciplinaire.
- **Le continuum.** Entre l'interdiction totale et l'usage libre non documenté, il existe tout un spectre. Pour certaines compétences, interdire a du sens. Pour d'autres, l'usage de l'IA fait partie du métier réel — un gestionnaire de projet produit ses rapports avec, autant évaluer sa capacité à l'orchestrer.
- **La ligne rouge.** Le copier-coller sans réflexion : l'abandon du jugement, le cerveau sorti du processus. C'est là que commence la triche, et nulle part avant.

Le désaccord assumé de l'épisode

Rochane défend une métaphore : le savoir est le carburant, l'évaluation doit mesurer le moteur, pas le réservoir. Ce qu'on évalue désormais, c'est ce qu'on fait de ce savoir dans une situation réelle, parfois imprévue, absente des données du modèle — dans la taxonomie de Bloom, mémoriser redevient le point de départ minimal.

Clément objecte qu'une voiture n'avance pas sans carburant : sans culture incorporée, certaines conversations et certains raisonnements deviennent impossibles, et on ne peut pas interrompre un échange pour aller chercher un contexte manquant.

Rochane précise alors sa position : le savoir, c'est comprendre le *pourquoi*, pas seulement le *quoi*. Le *quoi*, l'IA le produit en trente secondes ; l'expérience clinique, le vécu de chantier, la lecture d'une salle de formation qui part en vrille, non.

> **Gardez les deux bouts.** C'est le débat, pas sa résolution, qui est utile en conception.

Ce qui conditionne la réussite

Trois niveaux à travailler ensemble

1. **L'institution** — ce qui est autorisé, ce qui est interdit, écrit noir sur blanc. Une charte absente coûte cher : une étudiante poursuivie pour un mémoire jugé rédigé par IA a obtenu gain de cause devant le tribunal, faute de règle écrite dans son établissement.
2. **L'usage critique** — utiliser l'IA s'apprend. Quand lui faire confiance, quand la contredire, comment vérifier ce qu'elle produit : rien de tout cela n'est inné. Le formateur doit le montrer concrètement avant d'évaluer là-dessus.
3. **Le discernement** — l'évaluer suppose de savoir lire un processus. Or les enseignants ont été formés à lire un texte et dire s'il est bon, pas à lire une démarche et dire si elle révèle une pensée compétente. Ce n'est pas un déficit de bonne volonté : c'est une question d'outillage et de temps.

La fracture culturelle

Après la fracture numérique vient la fracture culturelle. D'un côté, ceux qui délèguent leur travail à un algorithme et gobent ce qu'il produit : ils ne monteront pas en compétence. De l'autre, ceux qui comprennent que l'IA peut fonctionner comme un exosquelette, permettant de porter des charges plus lourdes. La différence ne se joue pas sur l'outil, mais sur l'accompagnement.

Le fil rouge

On ne demande pas à l'apprenant d'éviter l'IA. On lui demande de rester le **pilote**.

Les outils cités

OUTIL	USAGE EN ÉVALUATION
NotebookLM	RAG sur ses propres sources : retrouver une étude, une anecdote, un passage précis dans un corpus. Rochane y a chargé 300 sources pour écrire son livre. Précieux aussi pour les apprenants en révision.
Google AI Studio	Gratuit, et largement ignoré. C'est là qu'a été construit le dialogue évaluatif socratique, de A à Z.
Notion	Support de portfolios réflexifs : dépôt multimodal, partage simple, sans avoir à créer un site dédié.

> **Un signal repris dans l'épisode.** Une école belge pionnière sur le sujet a mis en place un codebook d'usage de l'IA et formé ses enseignants. Lors d'un échange d'une demi-heure sur leur dispositif, aucun nom d'outil n'a été prononcé — uniquement des usages. C'est le meilleur indicateur de maturité d'une institution.

Cinq questions à poser sur une évaluation existante

1. L'IA peut-elle produire ce livrable à la place de l'apprenant sans que cela se voie ?
2. Qu'est-ce que cette évaluation rend visible qu'une IA ne peut pas produire à la place de l'apprenant ?
3. La preuve demandée est-elle réellement observable, ou reste-t-elle déclarative ?
4. Quelle posture est en jeu — résistante, collaborative, formative — et est-ce la bonne au regard de ce que je veux mesurer ?
5. Les conditions de passation tiennent-elles dans la vraie vie : effectifs, temps de correction, distanciel ?

> **Le test décisif.** Soumettez votre évaluation actuelle à une IA. Si elle la réussit seule, elle n'évaluait déjà pas grand-chose.

Pour aller plus loin

- Rochane Kherbouche, ouvrage sur l'évaluation à l'ère de l'IA, éditions La Chronique Sociale, préface de Christelle Lison.
- L'épisode complet de « Rendez-vous en terre digitale », 45 minutes : lien tout en haut de ce document !



Clément Cahagne — pauuse
Ingénierie pédagogique & IA appliquée à la formation
linkedin.com/in/clementcahagne clement.cahagne@pauuse.fr