

Correction de copies assistée par IA : d'une expérimentation ciblée à un déploiement généralisé

Un retour de terrain au cœur des pratiques d'évaluation

Emmanuelle Deglaire est professeure en droit et fiscalité à l'EDHEC depuis quinze ans. Son travail autour de l'IA appliquée à l'évaluation s'inscrit dans une démarche de recherche et d'expérimentation ayant donné lieu à plusieurs publications. Après une phase de test, l'usage de la solution **BlueTrain** a progressivement été généralisé dans les pratiques de correction de l'EDHEC.

Nous avons souhaité l'interroger afin d'apporter un éclairage concret au panorama des solutions présenté dans cette newsletter. Son retour d'expérience permet de dépasser l'analyse fonctionnelle des outils pour comprendre, en situation réelle, les apports, les ajustements nécessaires et les questions pédagogiques que soulève l'intégration de l'IA dans les processus d'évaluation.



Une expérimentation progressive, fondée sur un besoin métier concret

À l'EDHEC, l'intérêt pour les usages pédagogiques de l'IA est apparu très tôt, dès la sortie de ChatGPT. Le PiLab, le laboratoire d'innovations pédagogiques de l'école, a engagé un travail de veille approfondie sur les cas d'usage émergents tandis qu'en parallèle, un alumni développait une solution dédiée à la correction assistée par IA, qui deviendrait BlueTrain.

De cette convergence est né un partenariat tripartite réunissant le PiLab, BlueTrain et un groupe restreint d'enseignants volontaires. La démarche s'est construite comme un « jeu de ping-pong » itératif, en partant d'un besoin clairement identifié : optimiser la correction des copies.

Pourquoi la correction de copies ? La correction est unanimement identifiée comme l'une des tâches les plus chronophages des métiers d'enseignement. C'est aussi une activité très répétitive, ce qui en fait un terrain particulièrement propice à l'automatisation, à condition d'un entraînement rigoureux des modèles.

1. Modalité de l'expérimentation

Concrètement pour conduire cette expérimentation et entraîner l'algorithme, Emmanuelle Deglaire fournissait à BlueTrain :

- Le sujet d'examen,
- Une copie dite "parfaite",
- un barème détaillé.

L'algorithme est ensuite entraîné par itérations successives, avec des ajustements fréquents du barème. Plusieurs constats structurants ont émergé rapidement :

- Le passage à des examens sur ordinateur s'est imposé, la **transcription de copies manuscrites** entraînant trop de pertes et d'erreurs,
- Lorsque les copies ne respectent pas le format attendu, la correction est **immédiatement réattribuée à l'enseignant**, afin d'éviter une perte de temps,
- La **double correction** s'est révélée nécessaire, à la fois pour des raisons fonctionnelles et réglementaires. Une correction uniquement automatisée posait notamment problème au regard de l'**AI Act** [1], qui impose un regard humain systématique.

L'expérimentation a débuté sur d'anciennes copies déjà corrigées afin d'objectiver les écarts entre notation humaine et notation assistée par IA. Le déploiement s'est ensuite fait de manière progressive :

- En 2024, sur de vraies copies issues de petites promotions,
- En 2025, sur des promotions de grande taille.

Aujourd'hui, l'outil est effectivement déployé auprès de plusieurs enseignants, dans différentes disciplines (marketing, finance, etc.).

2. Gains observés et conditions de réussite

Les bénéfices identifiés du recours à une solution dédiée, du marché, sont multiples :

- Un **gain de temps significatif** pour les enseignants,
- Une **interface personnalisée**, intégrée au LMS de l'école, permettant de reprendre directement les commentaires et propositions de l'IA en mode éditeur,
- Une **amélioration de la qualité et de la sécurité de l'évaluation**, la double correction permettant de renforcer la cohérence et de limiter les erreurs.

L'IA comme révélateur du travail invisible de la notation

1. Déconstruire l'acte de corriger

L'un des apports majeurs de cette expérimentation a moins résidé dans l'IA elle-même que dans le travail de réflexion qu'elle a imposé sur l'acte de notation. Emmanuelle Deglaire témoigne ainsi que l'acte d'évaluation ne consiste pas simplement en l'attribution d'une note : c'est concevoir un examen, formaliser un barème, expliciter des critères, produire un feedback.

Emmanuelle Deglaire associe ce découpage aux travaux sur les six niveaux d'automatisation de l'apprentissage personnalisé de Molenaar [2], qui invitent à situer précisément ce qui relève de l'automatisation et ce qui demeure du ressort humain. Dans le cas présent, l'IA intervient sur la correction et le feedback, mais reste entièrement **dépendante de choix pédagogiques humains** en amont.

[1] AI Act : le règlement européen sur l'intelligence artificielle encadre les usages à haut risque, comme la correction de copies. Il impose une supervision humaine, la traçabilité des décisions et la transparence du système, garantissant que l'enseignant reste responsable des notes.

[2] Molenaar Inge, « Personnalisation de l'apprentissage : Vers une forme hybride des technologies d'apprentissage combinant l'humain et l'IA », in book: Perspectives de l'OCDE sur l'éducation numérique 2021, décembre 2022, DOI:10.1787/0943da96-fr.

Correction de copies assistée par IA : d'une expérimentation ciblée à un déploiement généralisé

Un retour de terrain au cœur des pratiques d'évaluation

2. Biais, subjectivité et illusion de l'objectivité parfaite

La crainte du biais est souvent mise en avant lorsqu'il est question d'IA. Pourtant, le **besoin d'extérioriser et objectiver des critères n'est pas propre à l'arrivée de l'IA**. En France par exemple, dans un cours où interviennent plusieurs correcteurs, qu'ils soient professeurs permanents ou encore professeurs vacataires, la multiplicité des acteurs imposait déjà de détailler soigneusement les attentes envers les travaux des étudiants pour standardiser leur évaluation.

Dans d'autres contextes, comme aux États-Unis, la correction est souvent assurée par des teaching assistants. Au Royaume-Uni, la double correction est une pratique ancienne et obligatoire. L'IA remet ainsi au centre une question clé : comment formaliser sa propre vision de l'évaluation pour pouvoir, le cas échéant, la « sous-traiter » ?

L'expérimentation montre aussi les limites de toute tentative de standardisation parfaite. Une note n'est pas une donnée certaine : un écart entre 16 et 17 n'a pas nécessairement de portée significative. Toute notation comporte ainsi une part non négligeable de subjectivité, et le recours à l'IA n'y fait pas exception.

3. Encadrement, traçabilité et sécurité

Dans ce contexte, l'enjeu principal devient celui du **cadre réglementaire et éthique**. À l'EDHEC, un protocole précis a été défini : traçabilité des copies, usage d'une IA entraînée et agréée, conservation des prompts. Le véritable risque, selon Emmanuelle Deglaire, serait de **ne pas proposer de solutions encadrées aux enseignants**, niant ainsi les usages individuels qui se développent autour d'outils gratuits, sans garanties en matière de données ou de conformité.

Un gain souvent sous-estimé de l'IA apparaît alors : le **contrôle de cohérence du processus d'évaluation**. La double correction ne renforce pas seulement la fiabilité des notes, elle améliore aussi la cohérence globale du dispositif.

Ces enseignements ont été formalisés par Emmanuelle Deglaire et Peter Daly [3], au travers du « 8C Model of AI-Enabled Correction », qui interroge la faisabilité de la mise en œuvre d'une correction de copies par IA à travers :

- Le « why » : à savoir, un esprit de curiosité dans un climat de confiance,
- Le « how » : quel choix d'examen, de prompts,
- Le « who » : responsabilités et contrôles,
- Le « how much » : coûts, avec un investissement en temps et en argent concentré sur la phase de conception.

Usages et appropriation par les enseignants et les étudiants

Au-delà des aspects techniques, l'expérience met en lumière la diversité des manières dont les enseignants se saisissent de l'outil. Certains apprécient l'efficacité et la cohérence que l'IA apporte dans le processus de correction, tandis que d'autres l'utilisent de manière plus ponctuelle.

Cette diversité nous témoigne que l'IA a pu agir comme un levier de réflexion sur les pratiques pédagogiques en invitant à formaliser ses critères d'évaluation, à expliciter le barème et à organiser le feedback de manière plus transparente. **L'appropriation passe donc autant par la dimension technique que par un travail de réflexion sur l'acte de noter.**

Pour accompagner cette adoption, l'EDHEC a mis en place des mesures simples mais structurantes : intégration de l'outil dans le LMS, questionnaires réguliers sur la perception des utilisateurs. Ces retours nourrissent l'évolution de la solution et permettent de l'adapter aux besoins réels des enseignants et des étudiants.

Cette expérience nous montre que l'introduction de l'IA en éducation ne se limite pas à la technologie elle-même. Elle ouvre aussi des pistes pour repenser l'organisation des examens, la conception des évaluations et la manière dont l'enseignement supérieur accompagne ses enseignants dans la transformation des pratiques pédagogiques.