



Intégration de l'intelligence artificielle dans le diagnostic d'erreurs en mathématiques : étude de cas sur la complémentarité IA–enseignant au secondaire I

Analyse réflexive d'un dispositif hybride basé sur GPT-4o dans une classe de 9–10-11e HarmoS

MSMET 11 : Avant-projet de mémoire

Haute école pédagogique de Lausanne

Mots-clés : interaction IA-enseignant, diagnostic d'erreurs, mathématiques, différenciation pédagogique

Nathan Paillou - P61429

1. Introduction thématique

L'essor des technologies d'intelligence artificielle (IA) dans l'éducation interroge en profondeur les pratiques pédagogiques, notamment en ce qui concerne l'analyse des erreurs et la remédiation. En mathématiques, discipline où l'erreur constitue un levier fondamental d'apprentissage, ces outils – et en particulier les modèles de langage comme GPT-4o – permettent aujourd'hui de proposer des diagnostics automatisés, un feedback personnalisé et des suggestions de remédiation en temps réel. Ce tournant technologique offre des opportunités nouvelles, tout en soulevant des questions pédagogiques, éthiques et professionnelles majeures.

Le diagnostic d'erreurs est une activité complexe qui requiert une lecture fine des productions des élèves, une interprétation des cheminements sous-jacents et une capacité à proposer des remédiations adaptées. Dans un contexte de forte hétérogénéité scolaire, le recours à un outil capable d'assister l'enseignant dans cette tâche présente un intérêt manifeste. Encore faut-il que cette assistance s'inscrive dans une logique de complémentarité, et non de substitution, et qu'elle respecte les principes d'une éducation contextualisée, humaine et réflexive.

L'étude présentée dans cet avant-projet de mémoire s'inscrit dans cette perspective. Elle propose d'examiner l'intégration d'un outil technologique développé pour l'occasion – une application web compatible iPad – permettant à des élèves de résoudre des tâches mathématiques, d'être accompagnés par l'IA dans la compréhension de leurs erreurs, et de bénéficier d'un suivi différencié fondé sur l'analyse conjointe des traces numériques et du jugement professionnel de l'enseignant. L'objectif est d'évaluer dans quelle mesure ce type d'outil permet de soutenir, affiner ou enrichir le diagnostic pédagogique, et comment il influence la posture de l'enseignant vis-à-vis de l'erreur.

Au-delà des dimensions techniques, cette recherche s'inscrit dans un cadre socio-constructiviste où l'erreur est conçue comme un indicateur des conceptions en cours d'élaboration, et non comme une simple faute à corriger. Elle interroge le rôle que peuvent jouer les technologies intelligentes dans ce processus, en tenant compte des enjeux de confiance, de transparence, et de réflexivité professionnelle. À travers une expérimentation de terrain, elle cherche ainsi à contribuer à une réflexion plus large sur la manière dont l'intelligence artificielle peut s'intégrer de manière éthique et pédagogique dans le quotidien

des classes, en valorisant l'erreur comme point de départ de l'apprentissage plutôt que comme obstacle.

2. Cadre théorique

2.1 Socio-constructivisme et statut de l'erreur

Dans les approches pédagogiques contemporaines, l'erreur est progressivement réhabilitée comme objet d'étude à part entière et levier de développement cognitif. Loin d'être une simple faute à corriger, elle est désormais reconnue comme une manifestation utile des représentations mentales de l'élève (Astolfi, 1997). Cette conception trouve ses racines dans les travaux de Vygotsky (1978), pour qui l'apprentissage prend place dans la zone proximale de développement, c'est-à-dire là où l'élève peut progresser grâce à un accompagnement adapté. Dans ce cadre, l'erreur devient un indice précieux du décalage entre ce que l'élève sait faire seul et ce qu'il peut accomplir avec aide.

La didactique des mathématiques a largement contribué à théoriser la fonction constructive de l'erreur. Brousseau (1998) a mis en évidence le rôle du contrat didactique dans la production d'erreurs typiques, souvent liées aux attentes implicites de l'enseignant ou aux régularités du système d'enseignement.

Plusieurs typologies d'erreurs sont proposées dans la littérature. La plus classique est celle de Radatz (1979), qui distingue notamment :

- les erreurs de traitement, liées à une mauvaise interprétation de l'énoncé ;
- les erreurs de transformation, qui apparaissent lors de la mise en œuvre d'une procédure ;
- les erreurs de conceptualisation, résultant d'une compréhension incomplète ou erronée des notions en jeu.

D'autres auteurs, comme Fuchs et al. (1994), proposent une approche plus issue de la psychologie cognitive. Leur typologie distingue :

- les erreurs attentionnelles, liées à l'inattention ou à la distraction ;

- les erreurs procédurales, dues à une mauvaise application d'une méthode apprise ;
- les erreurs conceptuelles, similaires à celles de Radatz, mais vues sous l'angle de la représentation mentale défaillante.

Ces classifications, bien que différentes, ne s'excluent pas mutuellement : elles peuvent être mobilisées de façon complémentaire pour mieux comprendre la diversité des erreurs rencontrées. Le choix définitif de la typologie utilisée dans cette recherche sera précisé ultérieurement, suite au début de l'étude de la didactique des mathématiques.

Un climat de classe sécurisant et propice à l'expression de l'erreur est fondamental. Des recherches récentes (Soncini et al., 2022) montrent qu'un environnement qui valorise l'erreur favorise le développement de stratégies d'apprentissage plus adaptatives, augmente l'engagement des élèves et améliore leur persévérance face à la difficulté. La posture de l'enseignant est donc centrale : il ne s'agit pas seulement de tolérer l'erreur, mais d'en faire un support de différenciation pédagogique et un vecteur d'apprentissage actif (Bucheton & Soulé, 2009).

2.2 Intelligence artificielle et diagnostic pédagogique

L'introduction de l'IA dans les environnements éducatifs permet aujourd'hui d'automatiser certaines tâches d'analyse pédagogique, notamment l'identification des erreurs. Les modèles de traitement du langage naturel, tels que GPT-4, sont capables de générer des feedbacks cohérents à partir de productions écrites ou manuscrites d'élèves, en s'appuyant sur des corpus d'apprentissage massifs (King et al., 2024 ; McNichols et al., 2023).

Les systèmes tutoriels intelligents (Corbett et al., 1997) et les dispositifs de learning analytics permettent de repérer des schèmes d'erreurs récurrents et de proposer des remédiations adaptées. Toutefois, plusieurs limites ont été soulignées : l'IA est généralement plus efficace pour identifier des erreurs factuelles ou procédurales explicites que pour analyser des erreurs conceptuelles plus implicites ou liées à des raisonnements atypiques (Zhang et al., 2025). Par ailleurs, des biais algorithmiques peuvent nuire à l'équité du diagnostic, en survalorisant certains types de réponses ou en reproduisant des stéréotypes (Li et al., 2024 ; Williamson & Eynon, 2020).

Dans ce contexte, l'enseignant reste un acteur indispensable, garant d'une interprétation contextualisée et nuancée des productions d'élèves. La technologie ne peut pas se substituer au jugement professionnel ; elle peut en revanche devenir un outil d'appui, de vérification ou de stimulation réflexive.

2.3 Complémentarité IA–enseignant et posture réflexive

Les travaux de Holstein et al. (2019) ont mis en lumière les bénéfices d'une co-construction pédagogique entre enseignant et système d'IA. Lorsqu'elle est pensée comme une aide à l'orchestration en temps réel de la classe, l'IA peut contribuer à améliorer la réactivité de l'enseignant face aux besoins des élèves. Encore faut-il que l'enseignant reste en position de pilotage, et non d'exécution. La supervision humaine reste donc essentielle pour valider, ajuster ou refuser les suggestions de l'IA.

Ce positionnement rejoint les écrits de Perrenoud (2001), qui considère la réflexivité comme une compétence-clé du métier d'enseignant. La confrontation à un diagnostic automatisé peut être un puissant déclencheur de réflexivité : elle oblige l'enseignant à expliciter ses critères d'évaluation, à interroger ses biais, à comparer ses intuitions à celles de l'outil. Elle l'invite à endosser une posture d'analyste de sa propre pratique, apte à construire une réponse pédagogique à la fois argumentée, différenciée et contextualisée.

Des auteurs comme Giraudon et al. (2020) insistent sur la nécessité de former les enseignants à un usage critique et éclairé des technologies éducatives, afin qu'ils ne deviennent pas de simples exécutants d'outils algorithmiques, mais qu'ils puissent les intégrer à leur logique professionnelle propre. Dans cette perspective, l'IA est conçue non comme un dispositif de remplacement, mais comme un support au développement professionnel.

2.4 Enjeux éthiques et réglementaires

L'intégration d'outils d'IA dans les pratiques pédagogiques ne saurait se faire sans une réflexion approfondie sur les enjeux éthiques et juridiques qu'elle soulève. La nLPD (nouvelle Loi fédérale sur la protection des données) impose des obligations strictes en matière de recueil, de traitement et de conservation des données personnelles. En contexte scolaire, ces obligations sont d'autant plus sensibles qu'elles concernent des mineurs.

L'usage de modèles d'IA hébergés sur des serveurs tiers pose notamment des questions de souveraineté, de traçabilité des données, et de consentement éclairé. La CNIL (2023) recommande une anonymisation rigoureuse, une information claire à destination des usagers, ainsi que la mise en place de procédures explicites de gestion des droits numériques. Par ailleurs, les biais algorithmiques – qu'ils soient liés à l'entraînement du modèle ou à son mode d'interprétation – doivent être anticipés pour éviter la reproduction d'inégalités ou l'exclusion involontaire de certains profils d'élèves (UNESCO, 2021 ; Eduscol, 2023).

Enfin, l'enseignant ne peut faire l'économie d'une réflexion déontologique sur la place de l'automatisation dans son métier. Il lui revient de veiller à ce que les outils utilisés soient au service de son intention pédagogique, et non l'inverse. Une éthique de la responsabilité numérique s'impose, fondée sur les principes de transparence, d'équité et d'humanité dans les interactions éducatives.

3. Problématique et questions de recherche

Dans l'enseignement des mathématiques, l'analyse d'erreurs constitue un levier pédagogique fondamental. Elle permet d'accéder aux raisonnements implicites des élèves, de repérer les obstacles à l'apprentissage et de mettre en place des stratégies de remédiation ciblées. Toutefois, face à l'hétérogénéité croissante des classes et à la complexité des productions d'élèves, cette tâche peut s'avérer chronophage, partielle et soumise à des biais d'interprétation. C'est dans ce contexte que les technologies d'intelligence artificielle, et notamment les modèles de langage comme GPT-4, suscitent un intérêt croissant : ils promettent une capacité d'analyse rapide, exhaustive et personnalisée, susceptible de soutenir l'enseignant dans ses décisions pédagogiques.

L'introduction d'une IA dans les processus de diagnostic soulève néanmoins de nombreuses interrogations. Les modèles actuels sont-ils capables de détecter des erreurs complexes, conceptuelles ou implicites ? Sont-ils en mesure de proposer des pistes de remédiation didactiquement pertinentes et adaptées au contexte spécifique de la classe ? Plus encore, comment l'enseignant peut-il rester acteur de la remédiation dans un environnement où l'IA produit elle-même des jugements pédagogiques ?

Pour répondre à ces questions, un dispositif expérimental a été conçu : une application web, développée spécifiquement pour ce mémoire, permet aux élèves de résoudre des tâches mathématiques sur un tableau blanc numérique. L'IA intégrée analyse les productions manuscrites en temps réel, identifie les erreurs, propose des indications ou des conseils, et archive l'ensemble des échanges. L'enseignant, à partir de ces données, est ensuite amené à formuler son propre diagnostic, à comparer son analyse à celle de l'IA et à concevoir des remédiations différencierées. Ce processus permet d'étudier en profondeur la dynamique d'interaction entre intelligence humaine et artificielle dans un cadre pédagogique réel.

Ce mémoire ne vise donc pas une simple évaluation de performance entre humain et machine, mais une réflexion critique sur les conditions d'une collaboration fructueuse entre les deux. Il s'inscrit dans une posture de co-construction pédagogique, attentive aux apports, aux limites, mais aussi aux effets de cette interaction sur la posture professionnelle de l'enseignant et sur les apprentissages des élèves. Il interroge enfin les dimensions éthiques, techniques et organisationnelles de l'intégration de l'IA dans l'analyse formative.

Problématique : Comment une intelligence artificielle peut-elle enrichir, sans supplanter, le travail de diagnostic d'erreurs et de remédiation de l'enseignant en mathématiques, et dans quelles conditions cette collaboration favorise-t-elle une posture réflexive, une différenciation pédagogique pertinente, et une appropriation positive de l'erreur par les élèves ?

Questions de recherche :

1. Quelles différences, convergences ou complémentarités observe-t-on entre les diagnostics d'erreurs formulés par l'IA et ceux de l'enseignant ?
2. Comment l'enseignant mobilise-t-il les analyses et suggestions de l'IA dans ses choix de remédiation et dans sa réflexion pédagogique ?
3. En quoi l'usage d'un tel dispositif modifie-t-il la posture réflexive de l'enseignant face à l'erreur ?
4. Quelle perception les élèves ont-ils de l'IA comme partenaire d'apprentissage, et comment jugent-ils l'utilité ou la pertinence des retours automatisés ?
5. Quels sont les enjeux et limites – techniques, éthiques, pédagogiques – identifiés lors de l'expérimentation, et quelles conditions favorisent une intégration réussie de l'IA dans un environnement scolaire réel ?

4. Méthodologie

4.1 Contexte et dispositif expérimental

Cette recherche s'appuiera sur une méthodologie d'étude de cas (Yin, 2014) conduite au sein d'une classe de 9e, 10e ou 11e HarmoS. Le dispositif repose sur une application web développée pour ce projet, fonctionnant sur les iPad de la classe. L'application intègre un tableau blanc numérique où les élèves rédigent leurs réponses à l'aide d'un Apple Pencil.

Un agent d'intelligence artificielle (GPT-4o), agissant comme tuteur personnalisé, est intégré à l'application. Il analyse en temps réel les productions manuscrites des élèves pour détecter les erreurs et proposer une aide sous forme d'indices progressifs, de reformulations ou de conseils, sans jamais fournir la solution. L'ensemble des interactions (réponses, modifications, feedback de l'IA) est automatiquement enregistré.

4.2 Déroulement de l'expérimentation

L'expérimentation se déroulera en deux temps, sur un thème mathématique choisi en amont pour sa pertinence didactique et sa faisabilité technique. La première séance sera consacrée au diagnostic. Durant celle-ci, les élèves résoudront des tâches mathématiques sur l'application et interagiront avec l'IA pour obtenir des retours sur leurs erreurs, sans intervention de l'enseignant afin de recueillir des interactions spontanées. Suite à cette séance, l'enseignant procèdera à une co-analyse : il établira d'abord son propre diagnostic à partir des productions brutes des élèves, puis le confrontera à celui de l'IA pour en identifier les convergences, les divergences et les complémentarités.

La seconde séance portera sur la remédiation. En s'appuyant sur les résultats de cette double analyse, l'enseignant concevra et mettra en œuvre un accompagnement différencié, pouvant prendre la forme d'exercices ciblés, de rétroactions individualisées ou de travail en petits groupes. L'objectif sera de mobiliser les forces des deux diagnostics tout en tenant compte de leurs limites respectives.

4.3 Données collectées et instruments

Pour analyser ce dispositif, plusieurs sources de données seront recueillies à l'aide d'instruments complémentaires :

- Productions des élèves et interactions avec l'IA : Les réponses manuscrites initiales, leurs modifications successives et l'historique des dialogues avec l'IA seront archivés par l'application.
- Diagnostics croisés : Une grille d'analyse croisée sera utilisée pour consigner et comparer systématiquement les diagnostics de l'enseignant et de l'IA pour chaque erreur (type d'erreur, convergence/divergence, pertinence du retour).
- Posture et décisions de l'enseignant : Une grille réflexive (ou journal de bord) guidera l'auto-analyse de l'enseignant sur ses choix de remédiation, sa confiance envers l'IA et l'évolution de sa posture.
- Perceptions des élèves : Leurs ressentis seront collectés via :
 - Un questionnaire post-expérimentation avec des échelles de Likert et des questions ouvertes (utilité perçue de l'IA, clarté, préférence de feedback).
 - Des entretiens semi-directifs avec un échantillon d'élèves pour explorer en profondeur leur vécu (rapport à l'erreur, confiance dans l'outil).

4.4 Analyse des données

Les données seront traitées par des méthodes croisées. Les productions d'élèves et les diagnostics feront l'objet d'un codage qualitatif s'appuyant sur des typologies d'erreurs issues de la littérature (ex: Radatz, 1979 ; Fuchs et al., 1994), dont le choix sera affiné avant le début de l'expérience. Les écarts entre le diagnostic humain et celui de l'IA seront analysés quantitativement (fréquences, taux d'accord) et qualitativement (analyse des cas critiques).

Les données issues des grilles réflexives et des entretiens seront soumises à une analyse thématique inductive (Braun & Clarke, 2006) pour faire émerger les catégories relatives aux postures professionnelles et aux perceptions des élèves. La validité de l'interprétation sera renforcée par la triangulation de l'ensemble des sources.

4.5 Faisabilité

La faisabilité de ce projet de recherche est jugée réaliste, car elle s'ancre dans le cadre réglementé et supervisé d'un stage de la Haute école pédagogique, garantissant un accès au terrain. La sécurité et la confidentialité sont au cœur du dispositif : l'application, conçue par le chercheur, stocke les données localement sans transmettre d'informations personnelles, et le

recueil du consentement éclairé de la direction, de l'enseignant, des parents et des élèves constituera une étape préliminaire importante. Un point de réflexivité supplémentaire concernera la gestion de ma double casquette d'enseignant et de chercheur ; le journal de bord réflexif sera utilisé non seulement pour analyser la situation pédagogique, mais aussi pour interroger ma propre posture et la subjectivité qu'elle peut induire dans le recueil et l'interprétation des données.

L'anticipation des défis humains est une priorité. Pour parer au risque principal, un refus d'autorisation de la part de l'établissement ou de l'enseignant, une communication transparente sera menée très en amont, s'appuyant sur un dossier de recherche qui présente les bénéfices pédagogiques et les garanties éthiques du projet. De même, pour favoriser l'adhésion des élèves et des familles, un soin particulier sera apporté à la rédaction des formulaires de consentement : ceux-ci expliqueront en des termes simples et clairs ce que l'IA fera (analyser les productions mathématiques pour fournir des indices), ce qu'elle ne fera pas (noter ou juger l'élève) et comment les données seront rigoureusement anonymisées et protégées. Une séance d'introduction visera également à expliquer l'outil directement aux élèves pour garantir leur compréhension et leur adhésion.

Les risques pratiques, d'ordre technique ou didactique, sont également pris en compte. Un protocole de validation matériel complet, incluant des tests sur le site de l'école, permettra de mitiger les défaillances potentielles du réseau Wi-Fi ou de l'application. Une connexion mobile de secours est d'ailleurs prévue pour garantir la continuité du service. Sur le plan pédagogique, la flexibilité est la solution retenue : plusieurs thématiques mathématiques seront préparées, et le choix final sera arrêté en étroite concertation avec l'enseignant-formateur pour s'assurer qu'il corresponde parfaitement à la progression de la classe au moment de l'expérimentation.

Enfin, l'ensemble de la démarche sera soutenu par une posture réflexive. La tenue d'un journal de bord permettra de documenter tous les imprévus et les ajustements effectués, transformant les obstacles potentiels en données précieuses sur les conditions réelles d'intégration de l'intelligence artificielle en classe.

5. Plan du mémoire et calendrier

Le plan provisoire du mémoire qui s'appuie sur les informations de la directive 05_08 de la HEP Vaud est le suivant :

- Page de couverture (programme d'études, titre, auteur, directeur, membre du jury, logo, date)
- Abstract (résumé de la recherche env. 300 mots)
- Table des matières
- Introduction
- Cadre théorique (revue de la littérature et concepts liés)
- Problématique et question de recherche (résumé du problème, hypothèses, dimensions analysées)
- Méthodologie (contexte de la recherche et classes sujettes à l'étude, description et justification des outils de recueil des données, description de l'analyse des données qualitatives)
- Présentation du déroulement de la première séance
- Présentation du déroulement de la seconde séance
- Résultats (de la première et de la seconde analyses)
- Discussion (interprétation des résultats, difficultés rencontrées, apports et limites de la recherche, intérêt de la recherche pour la pratique enseignante)
- Conclusion
- Bibliographie (normes APA)
- Annexes

La réalisation du mémoire s'échelonnera sur l'ensemble de la deuxième année du master, entre septembre 2025 et juin 2026. Elle s'organise en deux grandes phases : une phase préparatoire au semestre 3 (effectuée à distance), et une phase de mise en œuvre sur le terrain au semestre 4, une fois de retour en Suisse.

Le semestre 3 (septembre à décembre 2025) sera consacré à la consolidation du protocole, à la finalisation de l'application web (amélioration de l'interface, intégration des fonctions d'analyse GPT-4o, enregistrement automatique des interactions), ainsi qu'à la conception des instruments de recherche : questionnaires élèves, grilles d'analyse croisée des erreurs, guides d'entretien et grille réflexive enseignant. Cette période permettra également un affinement du cadre théorique et une vérification de la cohérence globale de la démarche.

Durant janvier et février 2026, des tests internes de l'application seront menés (en autonomie ou avec des élèves volontaires) afin d'en valider le fonctionnement technique et la

compatibilité avec les objectifs pédagogiques. D'éventuels ajustements du protocole ou des outils seront réalisés avant le lancement de l'expérimentation.

Les deux séances d'expérimentation auront lieu en mars ou avril 2026. Le mois de mai 2026 sera consacré à l'analyse des données, croisant approches quantitatives et qualitatives. Enfin, le mois de juin 2026 sera réservé à la rédaction finale, aux relectures et au dépôt du mémoire.

Tableau récapitulatif des échéances:

Période	Activités principales
Sept.-déc. 2025	Échange académique : finalisation de l'application, conception des outils, consolidation du protocole
Janv.-févr. 2026	Tests de l'application, ajustements techniques, relecture et validation du protocole
Mars 2026	Séance 1 – Diagnostic des erreurs via l'application (collecte des productions et des retours IA)
Avril 2026	Séance 2 – Remédiation différenciée (intervention de l'enseignant), entretiens élèves, journal réflexif
Mai 2026	Analyse des données quantitatives et qualitatives, triangulation, synthèse interprétative

Juin 2026	Rédaction du mémoire, relectures finales, dépôt
-----------	-------------------------------------------------

6. Encadrement

L'encadrement scientifique de ce projet sera assuré par un directeur de mémoire, Micha Hersch, professeur HEP associé en didactique de l'informatique. Son expertise dans les domaines de l'intelligence artificielle éducative, des interactions homme–machine ainsi que sa solide expérience en didactique des mathématiques, constitueront un appui méthodologique et didactique essentiel tout au long du projet.

6. Bibliographie

- Astolfi, J.-P. (1997). *L'erreur, un outil pour enseigner*. ESF éditeur.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brousseau, G. (1998). *Théorie des situations didactiques*. La Pensée Sauvage.
- Bucheton, D., & Soulé, Y. (2009). *Les gestes professionnels de l'enseignant: Crises et variations*. ESF.
- CNIL. (2023). *Intelligence artificielle et données personnelles : recommandations pour un usage responsable en milieu éducatif*. <https://www.cnil.fr>
- Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. In M. G. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (2nd ed., pp. 849–874). Elsevier.
- Eduscol. (2023). *Usages de l'IA en milieu scolaire : enjeux éthiques et pratiques*. Ministère de l'Éducation nationale.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1994). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal*, 31(3), 619–646. <https://doi.org/10.3102/00028312031003619>
- Giraudon, P., Flandin, S., & Ria, L. (2020). Former à l'usage critique des technologies éducatives : quels enjeux ? *Revue française de pédagogie*, (202), 87–96. <https://doi.org/10.4000/rfp.10296>
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3290605.3300831>

King, S. D., Ferguson, R., & Luckin, R. (2024). Generative AI in education: Evaluating GPT-4 for diagnostic feedback. *Journal of Learning Analytics*, 11(1), 43–65. <https://doi.org/10.18608/jla.2024.7311>

Li, J., Zhang, X., & Xu, C. (2024). Bias in AI-powered educational tools: Risks and remedies. *AI & Society*. Advance online publication. <https://doi.org/10.1007/s00146-024-01734-9>

McNichols, T., Chen, M., & Davis, D. (2023). Natural language processing for formative feedback in mathematics. *Computers & Education: Artificial Intelligence*, 4, 100116. <https://doi.org/10.1016/j.caeari.2023.100116>

Perrenoud, P. (2001). *La compétence professionnelle de l'enseignant : entre savoirs, schèmes d'action et pouvoir d'agir*. ESF.

Radatz, H. (1979). Error analysis in mathematics education. *Journal for Research in Mathematics Education*, 10(3), 163–172. <https://doi.org/10.2307/748804>

Soncini, A., Tugnoli, S., & Pastorelli, M. (2022). Promoting error-friendly climates in math classrooms: Effects on student motivation. *European Journal of Psychology of Education*, 37, 865–883. <https://doi.org/10.1007/s10212-021-00547-4>

UNESCO. (2021). *Recommandation sur l'éthique de l'intelligence artificielle*. <https://unesdoc.unesco.org/ark:/48223/pf0000380455>

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223–235. <https://doi.org/10.1080/17439884.2020.1798995>

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Sage.

Zhang, M., Heffernan, N., & Singh, A. (2025). Understanding the limitations of AI in detecting conceptual errors in student math responses. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-025-00309-2>