

***Explainable Artificial Intelligence (XAI) e Human-in-the-Loop (HITL):
Tecnologia e Mediação Docente no Feedback Formativo***

***Explainable Artificial Intelligence (XAI) and Human-in-the-Loop (HITL):
Technology and Teacher Mediation in Formative Feedback***

Márcio Silveira Nascimento¹; Aline Santos Mossette²; Sidney Pires Martins³; Éber José dos Santos⁴

Resumo

A inteligência artificial (IA) tem sido incorporada à educação como recurso capaz de otimizar processos e personalizar aprendizagens, mas seu uso suscita desafios relacionados à explicabilidade, à justiça e à mediação pedagógica. Este estudo teve como objetivo analisar a integração entre a *Explainable Artificial Intelligence* (XAI) e o modelo *Human-in-the-Loop* (HITL) na oferta de *feedback* formativo em ambientes digitais. Para isso, adotou-se uma abordagem teórico-reflexiva baseada em revisão bibliográfica interpretativa, a fim de problematizar produções acadêmicas relevantes e sistematizar contribuições sobre o tema. Os resultados indicam que, isoladamente, a IA apresenta limitações como a reprodução de vieses e a falta de contextualização pedagógica; entretanto, quando combinada à mediação docente em protocolos híbridos, possibilita a produção de devolutivas claras, personalizadas e inclusivas. Conclui-se que a articulação entre XAI e HITL representa um caminho promissor para fortalecer a aprendizagem flexível e garantir práticas avaliativas mais éticas, transparentes e equitativas, nas quais a tecnologia complementa, mas não substitui, a ação pedagógica.

Palavras-chave: mediação docente; *feedback* formativo; inteligência artificial na educação; personalização da aprendizagem.

Abstract

Artificial intelligence (AI) has been incorporated into education as a resource capable of optimizing processes and personalizing learning, but its use raises challenges related to explainability, fairness, and pedagogical mediation. This study aimed to analyze the integration of Explainable Artificial Intelligence (XAI) and the Human-in-the-Loop (HITL) model in providing formative feedback in digital environments. To this end, we adopted a theoretical-reflective approach based on an interpretative literature review, aiming to problematize relevant academic productions and systematize contributions on the topic. The results indicate that, in isolation, AI presents limitations such as the reproduction of biases and the lack of pedagogical contextualization; however, when combined with teacher mediation in hybrid protocols, it enables the production of clear, personalized, and inclusive feedback. We conclude that the articulation between XAI and HITL represents a promising path to strengthening flexible learning and ensuring more ethical,

¹ GEPPFOR-UFV; Instituto Federal do Amazonas (IFAM); marciosn.geo@gmail.com; <https://orcid.org/0000-0002-8963-3140>.

² GEPPFOR-UFV; Instituto Federal da Bahia (IFBA); alinemossette1@gmail.com; <https://orcid.org/0000-0003-0271-2522>.

³ GEPPFOR-UFV; Universidade Federal de Viçosa (UFV); sidney.martins@ufv.br; <https://orcid.org/0000-0002-4890-9307>.

⁴ GEPPFOR-UFV; Universidade Católica de São Paulo (PUC-SP); ejsantos2010@gmail.com; <https://orcid.org/0000-0003-2021-2062>.

transparent, and equitable assessment practices, in which technology complements, but does not replace, pedagogical action.

Keywords: teaching mediation; formative feedback; artificial intelligence in education; personalization of learning.

Introdução

A aplicação da IA na educação tem se mostrado relevante por possibilitar a criação de soluções que enriquecem o processo de ensino-aprendizagem, funcionando como recurso de apoio tanto às práticas docentes quanto ao estudo dos alunos. Nesse sentido, constitui uma alternativa metodológica pertinente à era da informação, contribuindo para otimizar o tempo e favorecer a consolidação de aprendizagens (Parreira, Lehmann & Oliveira, 2021).

Apesar desses avanços, também emergem preocupações relacionadas à qualidade pedagógica, à explicabilidade, entendida como a clareza e a transparência das decisões tomadas pelos sistemas de IA, e à justiça, compreendida como a garantia de que tais tecnologias não reforcem desigualdades ou preconceitos. Esses aspectos demandam atenção em sua implementação. De acordo com Du Boulay (2023), questões éticas vinculadas à privacidade e à segurança dos dados dos estudantes configuram-se como centrais, uma vez que a coleta e a análise massiva de informações educacionais levantam dilemas complexos sobre consentimento, uso responsável dos dados e proteção da privacidade individual.

Diante desse cenário, ganha relevância a *Explainable Artificial Intelligence* (XAI), que corresponde a um conjunto de métodos e técnicas destinados a tornar compreensíveis para os usuários humanos os processos, resultados e decisões gerados por algoritmos de *Machine Learning* (ML), uma área da inteligência artificial em que sistemas aprendem automaticamente a partir de dados, identificando padrões e realizando previsões sem necessidade de programação explícita para cada tarefa, promovendo maior transparência e confiança em seu uso (Miller, 2019; Thampi, 2022). Idealmente, a XAI deve explicar competências do sistema, justificar ações passadas e em andamento, além de explicitar as informações utilizadas para orientar decisões (Gunning et al., 2019).

Em paralelo, o modelo *Human-in-the-Loop* (HITL) oferece outra perspectiva importante, ao integrar a expertise humana ao processo de ML, no qual sistemas artificiais aprendem a partir de exemplos com base em métodos estatísticos (Arão, 2024). Nesse arranjo, humanos orientam o treinamento algorítmico, selecionando dados relevantes e ajustando decisões, o que possibilita maior precisão e eficiência (Kumar et al., 2024). A combinação entre XAI e HITL, portanto, desponta como estratégia promissora para equilibrar inovação tecnológica e mediação pedagógica.

Quando observados sob a ótica da aprendizagem flexível, esses recursos tecnológicos tornam-se ainda mais significativos, pois podem apoiar processos autorregulatórios por

meio da interação entre estudantes, docentes e sistemas digitais (Da Silva & Alliprandini, 2025). Nesse processo, destaca-se o papel do *feedback* formativo, reconhecido como fator de grande impacto no desempenho acadêmico (Hattie & Timperley, 2007). Tal prática funciona como estratégia de motivação e regulação da aprendizagem, favorecendo a autonomia dos estudantes e o desenvolvimento de competências (Santos & Kroeff, 2018).

Para analisar criticamente esse cenário, o presente estudo adotou uma abordagem teórico-reflexiva, estruturada a partir de uma revisão bibliográfica de caráter interpretativo, entendida como um processo de análise crítica das produções acadêmicas mais relevantes sobre o tema. Como destacam Lima e Mito (2007), a revisão bibliográfica deve ultrapassar a simples descrição de conteúdos, assumindo uma dimensão problematizadora que permita questionar, confrontar e ampliar as perspectivas já existentes. Nessa direção, a análise não se restringe a compilar informações, mas busca estabelecer articulações conceituais que subsidiem a construção de novas interpretações.

De forma complementar, Sousa, Oliveira e Alves (2021) afirmam que a revisão bibliográfica interpretativa tem como objetivo identificar, examinar e criticar documentos publicados, favorecendo tanto a atualização quanto o aprimoramento do conhecimento disponível. Esse caráter analítico-crítico torna a metodologia especialmente aderente à presente pesquisa, uma vez que esta se propõe a refletir sobre a integração entre XAI e modelos HITL no contexto do *feedback* formativo. A natureza emergente e ainda pouco consolidada desse campo exige justamente um exercício de interpretação crítica da literatura, capaz de sistematizar as contribuições existentes, evidenciar lacunas e apontar caminhos para práticas avaliativas mais éticas, inclusivas e inovadoras.

Ainda assim, é importante destacar que, embora a IA já esteja incorporada em plataformas educacionais para fornecer devolutivas imediatas e personalizadas (Holmes, Bialik & Fadel, 2019), permanecem desafios significativos. Entre eles, destacam-se a presença de vieses algorítmicos, a opacidade dos modelos e as limitações na adaptação às necessidades individuais dos alunos (Noble, 2018; Mehrabi et al., 2021). Diante desse cenário, este artigo propõe uma reflexão crítica sobre a integração entre XAI e HITL na oferta de *feedback* formativo, analisando seus impactos e implicações éticas e pedagógicas, bem como discutindo desafios e oportunidades para sua implementação.

Feedback formativo e aprendizagem flexível

O *feedback* formativo constitui-se como um processo avaliativo contínuo que visa não apenas aferir resultados, mas sobretudo orientar e potencializar a aprendizagem. Segundo Hattie e Timperley (2007), trata-se de uma das estratégias pedagógicas de maior impacto sobre o desempenho acadêmico, desde que seja claro, específico e ofereça direcionamentos que auxiliem o estudante a compreender seus progressos e dificuldades. Hoffmann (2014) destaca que a avaliação formativa deve ser concebida

como um processo dialógico e investigativo, no qual o professor assume o papel de mediador, ajudando o aluno a refletir sobre sua trajetória e a desenvolver maior autonomia. Essa perspectiva desloca a avaliação de uma função meramente classificatória para uma dimensão pedagógica e emancipatória.

Em convergência com essa visão, o conceito de aprendizagem flexível ganha relevância ao propor práticas educacionais que se ajustem às necessidades, ritmos e estilos de aprendizagem dos estudantes. Para Collis e Moonen (2012), a flexibilidade se expressa em múltiplas dimensões temporais, espaciais, metodológicas e tecnológicas, permitindo que o estudante tenha maior controle sobre seu processo formativo. Morán (2015) argumenta que a aprendizagem flexível implica a integração de metodologias ativas e tecnologias digitais, criando ambientes que favorecem a personalização, a autonomia e a participação do aluno como protagonista.

Nesse sentido, o *feedback* formativo desempenha papel central na aprendizagem autorregulada, pois permite que os alunos avaliem seu próprio trabalho e gerem devolutivas que os auxiliem a identificar lacunas, corrigir erros e planejar novas estratégias (Nicol & Macfarlane-Dick, 2006). Esse processo evidencia que a aprendizagem não é apenas receptiva, mas ativa, com o estudante participando diretamente da construção do seu conhecimento.

Assim, a devolutiva de qualidade se torna um elemento fundamental para o desenvolvimento da aprendizagem. Hattie e Timperley (2007) afirmam que, para ser eficaz, o *feedback* deve ser claro, orientador e contextualizado, promovendo não apenas o engajamento, mas também a autonomia dos alunos. A relevância desses elementos é reforçada por práticas pedagógicas que priorizam a clareza das informações, a orientação precisa sobre os próximos passos e a contextualização do conteúdo, garantindo que a devolutiva contribua efetivamente para a consolidação da aprendizagem e para a autorregulação do estudante.

Complementarmente, a qualidade do *feedback* também está ligada à sua especificidade e à atenção às necessidades de cada estudante e do docente. Lordelo e Dazzani (2009) destacam que o *feedback* deve ser detalhado, oferecido em momento oportuno e capaz de promover uma comunicação eficiente, incentivando a reflexão crítica do aluno sobre seu desempenho. Essa perspectiva reforça a ideia de que a devolutiva deve fornecer informações precisas sobre as fortalezas e áreas que necessitam de desenvolvimento, consolidando o processo de aprendizagem.

Ao transpor essas práticas para a educação digital, observa-se que as ferramentas baseadas em IA apresentam vantagens significativas, como rapidez na geração de *feedback* e personalização das devolutivas. Entretanto, Selwyn (2019) alerta que a adoção acrítica dessas tecnologias pode reforçar desigualdades existentes e reduzir a dimensão humana do ensino. Por isso, a explicabilidade e a transparência tornam-se requisitos fundamentais para a integração responsável da IA (Floridi et al., 2018).

Essa discussão leva à necessidade de uma reflexão crítica sobre os impactos da IA na educação. Selwyn (2019) aponta que o uso desregulado da tecnologia pode acentuar a

chamada "divisão digital", em que nem todos os estudantes têm igual acesso a recursos tecnológicos, ampliando disparidades e comprometendo a equidade educacional. Além disso, a utilização de IA sem explicabilidade adequada pode enfraquecer a mediação humana, essencial para interpretar, contextualizar e aprofundar os conteúdos educacionais. Floridi et al. (2018) enfatizam que a transparência nos processos decisórios das ferramentas de IA é crucial para que educadores e alunos compreendam e possam questionar as decisões automatizadas, promovendo confiança e responsabilidade no uso da tecnologia. Nesse sentido, práticas pedagógicas que incentivem o pensamento crítico, como o debate, a análise comparativa de informações e a problematização de resultados gerados por algoritmos, tornam-se essenciais para que os estudantes desenvolvam uma postura ativa e reflexiva diante das mediações tecnológicas (Nascimento & Santos, 2025).

Dessa forma, a integração de IA na educação deve ser orientada por práticas que assegurem equidade de acesso e transparência, garantindo que a tecnologia complemente, e não substitua, o trabalho pedagógico. Nesse cenário, o *feedback* formativo permanece como um componente central da aprendizagem flexível, pois sua clareza, orientação e contextualização, quando combinadas com o uso responsável da IA, fortalecem a autonomia do estudante, promovem engajamento e contribuem para a construção de experiências educativas mais significativas e inclusivas.

Por fim, a articulação entre *feedback* formativo e aprendizagem flexível mostra-se essencial para a educação contemporânea, uma vez que a flexibilidade demanda processos avaliativos constantes e orientadores que apoiem o estudante em suas escolhas e percursos individuais. Nesse sentido, Luckesi (2011) defende que a avaliação deve ser compreendida como um ato de acompanhamento, oferecendo devolutivas que auxiliem no desenvolvimento integral do sujeito, em contraste com práticas excludentes e punitivas. Assim, ao associar a adaptabilidade da aprendizagem flexível às orientações pedagógicas do *feedback* formativo, constrói-se um modelo educativo mais inclusivo, participativo e emancipador, capaz de responder aos desafios da sociedade em rede e às demandas do século XXI.

Explainable Artificial Intelligence (XAI) e Avaliação Híbrida mediada por modelos Human-In-The-Loop (HITL)

A XAI refere-se a um conjunto de métodos e técnicas voltados a tornar os processos, resultados e decisões produzidos por algoritmos de *machine learning* (ML), ou aprendizado de máquina, mais compreensíveis para os usuários humanos (Miller, 2019; Gunning et al., 2019). Diferentemente dos chamados modelos de "caixa-preta", que apresentam conclusões sem permitir a compreensão de sua lógica interna, a XAI busca promover clareza, justificando as decisões tomadas e permitindo que professores e estudantes compreendam os critérios que fundamentam determinadas devolutivas (Alves & Andrade, 2022; Adadi & Berrada, 2018). Nesse sentido, enquanto a "caixa-preta" representa um desafio ético e pedagógico por dificultar a interpretação e a

confiabilidade dos resultados, a XAI aproxima-se da metáfora da “caixa de vidro”, na qual os processos são visíveis, rastreáveis e passíveis de análise crítica.

No campo educacional, essa característica é crucial para que a tecnologia não apenas automatize tarefas, mas também fortaleça a confiança, a transparência e a participação crítica dos sujeitos envolvidos no processo de ensino-aprendizagem (Floridi et al., 2018). Ao oferecer justificativas claras para suas recomendações, a XAI possibilita que as decisões automatizadas sejam analisadas e validadas criticamente antes de sua aplicação, ampliando tanto o controle docente quanto a autonomia discente (Arrieta et al., 2020; Zawacki-Richter et al., 2019).

Complementarmente, a avaliação híbrida mediada por modelos HITL introduz a expertise humana no ciclo de funcionamento da IA, garantindo que os processos algorítmicos sejam supervisionados, ajustados e enriquecidos pela mediação docente (Kumar et al., 2024). Em vez de delegar integralmente a avaliação às máquinas, o HITL preserva a centralidade do professor, que orienta o treinamento dos algoritmos, seleciona dados relevantes e interpreta os resultados. Essa configuração assegura que a IA funcione como ferramenta de apoio, e não como substituto da ação pedagógica.

A avaliação híbrida, nesse contexto, configura-se como uma abordagem que integra o potencial analítico da inteligência artificial à sensibilidade pedagógica humana. Nesse modelo, a IA oferece rapidez, precisão estatística e capacidade de processar grandes volumes de dados, enquanto o professor assume um papel interpretativo, ético e contextual. A máquina pode fornecer devolutivas preliminares ou indicadores de desempenho, mas cabe ao docente validar, complementar ou ajustar tais informações, contemplando dimensões qualitativas do aprendizado que os algoritmos não conseguem captar integralmente, como a criatividade, a motivação, a criticidade e a singularidade das trajetórias formativas. Dessa forma, a avaliação híbrida mediada por modelos HITL promove um equilíbrio: a tecnologia potencializa o monitoramento e a personalização, ao passo que a mediação docente assegura a profundidade reflexiva e a relevância pedagógica do processo avaliativo.

Nesse arranjo, tanto professores quanto estudantes se beneficiam. Os docentes passam a compreender os fundamentos por trás das devolutivas automatizadas, o que lhes permite ajustá-las, complementá-las ou até mesmo contestá-las quando necessário, prevenindo a reprodução de vieses ou interpretações equivocadas (Noble, 2018). Para os alunos, a clareza nos critérios que orientam seu desempenho favorece maior confiança e engajamento no processo avaliativo. Assim, a explicabilidade da IA aliada à supervisão humana amplia a qualidade do *feedback*, tornando-o contextualizado, ético e ajustado às necessidades individuais.

Além disso, ao reduzir vieses e ampliar a aplicabilidade dos modelos em situações complexas, o HITL fortalece a participação ativa dos docentes nos processos avaliativos mediados por IA, assegurando que as devolutivas estejam alinhadas aos objetivos pedagógicos e não apenas a métricas automatizadas (Mosqueira-Rey et al., 2023). Esse equilíbrio entre eficiência tecnológica e sensibilidade humana é central para o avanço de práticas avaliativas inclusivas e emancipadoras.

A articulação entre XAI e HITL, portanto, mostra-se estratégica para a educação contemporânea. Enquanto a XAI assegura transparência e confiabilidade, o HITL garante a indispensável mediação docente, compondo um modelo híbrido em que a tecnologia potencializa, mas não substitui, o trabalho pedagógico. Essa integração fortalece práticas de avaliação formativa, possibilitando *feedbacks* claros, personalizados e contextualizados, ao mesmo tempo em que mantém o olhar crítico do professor. Em um cenário marcado pela crescente digitalização do ensino, a combinação dessas abordagens contribui para a construção de uma educação mais justa, inclusiva e eticamente orientada, em que a tecnologia atua como aliada da ação educativa e do protagonismo humano.

Limites do *feedback* automatizado, percepção docente e proposta de protocolo híbrido

A análise dos resultados revela três aspectos centrais que orientam a reflexão sobre a integração da IA nos processos avaliativos educacionais. O primeiro ponto refere-se aos limites do *feedback* automatizado. Embora sistemas baseados em IA ofereçam rapidez, padronização e escalabilidade, frequentemente carecem de contextualização pedagógica, o que pode comprometer a qualidade do processo de aprendizagem. Segundo Noble (2018), algoritmos podem reproduzir vieses presentes nos dados com os quais são treinados, resultando em *feedbacks* que não consideram as especificidades de cada estudante. Essa limitação é corroborada por estudos que apontam que, sem intervenção humana, o *feedback* automatizado tende a ser superficial e descontextualizado, não atendendo às necessidades individuais dos alunos (Morán, 2015; Collis & Moonen, 2012).

O segundo aspecto refere-se à percepção docente sobre o uso da IA na educação. Pesquisas indicam que, embora os professores reconheçam os benefícios da tecnologia, como a agilidade na correção de atividades e a personalização do ensino, há uma ênfase na necessidade de supervisão humana. Martins et al. (2024) destacam que os docentes valorizam a capacidade da IA de processar grandes volumes de dados, mas enfatizam que a mediação pedagógica é essencial para garantir a qualidade e a ética do processo avaliativo. Outros estudos reforçam essa percepção, evidenciando que a presença do professor é crucial para interpretar e contextualizar os *feedbacks* automatizados (Morán, 2015; Vicari, 2021).

Diante desses desafios, surge a proposta de um protocolo híbrido, fundamentado no modelo HITL. Nesse modelo, a IA atua como ferramenta para fornecer devolutivas iniciais, mas o professor mantém a responsabilidade de revisar, interpretar e contextualizar essas informações, garantindo que o *feedback* seja claro, explicável e pedagogicamente relevante. Kumar et al. (2024) argumentam que a integração da inteligência humana aos sistemas artificiais é essencial para supervisão contínua, ajustes e prevenção de vieses, promovendo uma avaliação justa e inclusiva. Estudos adicionais reforçam que a tecnologia deve servir como apoio à prática docente, e não

como substituto, assegurando relevância pedagógica e alinhamento com os objetivos educacionais (Kenski, 2014; Morán, 2015).

O modelo HITL emerge, portanto, como estratégia eficaz para que o feedback automatizado seja não apenas rápido e escalável, mas também contextualizado, relevante e alinhado aos objetivos de aprendizagem individuais. A seguir, o Quadro 1 sintetiza o protocolo, detalhando as etapas, as ações da IA, a mediação do professor e os resultados esperados.

Quadro 1
Protocolo Human-in-the-Loop para feedback formativo

Etapas	Ação da IA	Ação do Professor	Resultado Esperado
1. Coleta de dados	Analisar produções dos estudantes (textos, exercícios, quizzes) e identificar padrões de desempenho.	Validar se os dados analisados são representativos e condizem com o objetivo pedagógico da atividade.	Base de dados confiável para gerar <i>feedback</i> inicial.
2. Geração de <i>feedback</i> inicial	Produzir devolutivas automáticas (ex.: apontar erros recorrentes, sugerir recursos complementares).	Revisar a pertinência das devolutivas, filtrando vieses e inadequações.	<i>Feedback</i> preliminar rápido, mas supervisionado.
3. Mediação pedagógica	Disponibilizar explicações estruturadas e exemplos adaptados.	Contextualizar os resultados, acrescentar orientações específicas, relacionar com os objetivos da disciplina.	<i>Feedback</i> enriquecido e conectado ao currículo.
4. Ajuste para diversidade	Oferecer alternativas personalizadas (ritmo, dificuldade, estilo de recurso).	Garantir que as sugestões respeitem as singularidades cognitivas, culturais e socioeconômicas dos alunos.	<i>Feedback</i> inclusivo e equitativo.
5. Devolutiva ao estudante	Entregar recomendações adaptadas em tempo real.	Complementar com observações motivacionais, reforçar aspectos positivos e indicar próximos passos.	Estudante recebe <i>feedback</i> imediato, claro e humanizado.
6. Monitoramento contínuo	Gerar relatórios de progresso e padrões de evolução.	Utilizar os relatórios para replanejar atividades, intervir em dificuldades e ajustar o ensino.	Ciclo contínuo de avaliação e aprendizagem flexível.

Fonte: Elaborado pelos autores (2025), com base em Hattie & Timperley (2007), Nicol & Macfarlane-Dick (2006), Floridi et al. (2018), Zawacki-Richter et al. (2019), Banihashem et al. (2025).

A análise detalhada do Quadro 1 evidencia como o protocolo HITL articula tecnologia e mediação docente para fortalecer o feedback formativo de forma contextualizada e inclusiva. Esse equilíbrio é construído gradualmente em cada etapa do processo.

Na primeira etapa, a coleta de dados automatizada permite que a IA identifique padrões de desempenho, economizando tempo e garantindo rapidez no processamento das informações. Contudo, essa etapa só adquire valor pedagógico quando o professor valida se os dados analisados são representativos e condizentes com os objetivos da

atividade, assegurando sua relevância no processo avaliativo (Morán, 2015; Martins et al., 2024)

Seguindo para a segunda etapa, a geração de *feedback* inicial automatizado oferece devolutivas rápidas, sinalizando erros recorrentes e sugerindo recursos de estudo. A escalabilidade dessa fase é inegável, mas a revisão docente torna-se indispensável para filtrar vieses e ajustar as recomendações, de modo a alinhá-las às necessidades específicas de cada aluno (Lordelo & Dazzani, 2009; Vicari, 2021).

Na terceira etapa, de mediação pedagógica, evidencia-se o papel central do professor como interpretador e contextualizador do *feedback* gerado pela IA. Enquanto a tecnologia fornece informações estruturadas, é o docente quem as transforma em orientações personalizadas, conectadas ao currículo e à trajetória individual do estudante, estimulando sua autonomia e capacidade crítica (Hoffmann, 2014).

Esse movimento se aprofunda na quarta etapa, dedicada ao ajuste para diversidade, em que a IA apresenta alternativas personalizadas em termos de ritmo, dificuldade e estilo de recurso. Nesse ponto, a ação docente garante que tais sugestões respeitem as singularidades cognitivas, culturais e socioeconômicas dos estudantes, assegurando equidade e inclusão (Kenski, 2014; Perrenoud, 2015).

A quinta etapa, que trata da devolutiva ao estudante, representa a materialização do *feedback* humanizado. Aqui, a agilidade da IA é combinada à mediação pedagógica, resultando em devolutivas que não apenas informam sobre o desempenho, mas também motivam, reforçam conquistas e orientam próximos passos de forma clara e acessível (Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006).

Por fim, a sexta etapa, de monitoramento contínuo, encerra o ciclo ao transformar o *feedback* em um processo dinâmico e flexível. Relatórios de progresso gerados pela IA oferecem indicadores relevantes para o replanejamento de atividades e intervenções pedagógicas, permitindo que o professor acompanhe a evolução dos estudantes e favoreça sua autorregulação, em consonância com os princípios de aprendizagem contínua defendidos por Floridi et al. (2018).

Em síntese, o protocolo HITL mostra que a integração de IA na educação não deve ser entendida como substituição da prática docente, mas como um processo de colaboração em que a eficiência tecnológica se alia à sensibilidade humana. Essa combinação possibilita *feedbacks* mais rápidos, claros e personalizados, sem abrir mão da dimensão pedagógica, ética e inclusiva. Assim, o modelo contribui para fortalecer a aprendizagem flexível, fomentar a autonomia estudantil e garantir maior equidade nos processos avaliativos, consolidando um caminho promissor para uma educação mais justa e orientada pelo diálogo entre humanos e tecnologias (Perrenoud, 2015; Vicari et al., 2023).

Considerações finais

Este estudo partiu do problema da integração entre inteligência artificial e avaliação educacional, com foco nos desafios da explicabilidade, da justiça e da mediação pedagógica. Teve como objetivo analisar estratégias de combinação entre a XAI e o modelo HITL na oferta de *feedback* formativo em ambientes digitais.

A análise desenvolvida confirmou a hipótese de que a IA, quando utilizada de forma isolada, apresenta limites significativos, como a reprodução de vieses e a falta de contextualização pedagógica. Por outro lado, os resultados demonstraram que a integração da mediação docente em protocolos híbridos possibilita superar essas limitações, articulando eficiência tecnológica e sensibilidade pedagógica. Assim, evidencia-se que a IA não deve substituir, mas complementar o trabalho docente, promovendo *feedbacks* claros, explicáveis, personalizados e inclusivos.

Como contribuição, a pesquisa propôs um protocolo HITL para *feedback* formativo flexível, que detalha etapas de interação entre IA e professores, assegurando clareza, pertinência pedagógica e respeito à diversidade dos estudantes. Essa proposta reforça a centralidade da mediação humana e aponta caminhos para práticas avaliativas mais justas e equitativas no contexto da educação digital.

Entretanto, o estudo apresenta limitações por se tratar de uma investigação teórico-reflexiva, baseada em revisão bibliográfica interpretativa. Isso significa que seus achados precisam ser testados em contextos empíricos, com diferentes níveis de ensino e modalidades, a fim de validar a aplicabilidade e os efeitos do protocolo sugerido.

Para pesquisas futuras, recomenda-se a realização de estudos empíricos que avaliem a efetividade do modelo HITL em situações reais de ensino-aprendizagem, considerando variáveis como engajamento, motivação, desempenho acadêmico e equidade no acesso às tecnologias. Além disso, seria relevante explorar como políticas educacionais e formações docentes podem apoiar a implementação responsável de modelos híbridos de avaliação mediados por IA.

Em síntese, a investigação reforça que a integração de XAI e HITL no *feedback* formativo constitui um caminho promissor para equilibrar inovação tecnológica e princípios de justiça educacional. Ao proporcionar devolutivas que aliam clareza, explicabilidade e contextualização pedagógica, esse modelo contribui para o fortalecimento da aprendizagem flexível, permitindo que estudantes avancem em seus próprios ritmos sem perder a qualidade da mediação docente. Além disso, garante que a tecnologia não seja apenas uma ferramenta de automatização, mas um recurso de apoio à ação crítica e ética do professor, ampliando as possibilidades de personalização e inclusão no processo educativo. Nesse sentido, a combinação entre eficiência algorítmica e sensibilidade humana se apresenta como estratégia fundamental para superar os limites do *feedback* puramente automatizado e evitar a reprodução de desigualdades estruturais no acesso ao conhecimento.

Conclui-se, portanto, que a adoção de modelos avaliativos baseados na integração entre XAI e protocolos HITL pode representar um marco na construção de uma educação digital mais justa e orientada pelo bem comum. Tais modelos permitem que a inovação tecnológica se converta em um instrumento de emancipação e não de exclusão, fortalecendo a autonomia estudantil, valorizando o papel do professor e consolidando práticas avaliativas que respeitam a diversidade dos contextos de aprendizagem. Assim, a educação mediada por IA, quando guiada por princípios éticos e pedagógicos, pode se tornar não apenas mais eficiente, mas também mais humana, inclusiva e emancipadora.

Referências

- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, 6, 52138-52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- Alves, M. A. S., & Andrade, O. M. de. (2022). Da “caixa-preta” à “caixa de vidro”: o uso da explainable artificial intelligence (XAI) para reduzir a opacidade e enfrentar o enviesamento em modelos algorítmicos. *Direito Público*, 18(100). <https://doi.org/10.11117/rdp.v18i100.5973>
- Arão, C. (2024). Por trás da inteligência artificial: uma análise das bases epistemológicas do aprendizado de máquina. *Trans/formação*, 47, e02400163. <https://doi.org/10.1590/0101-3173.2024.v47.n3.e02400163>
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, 58, 82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Banihashem, S. K., Noroozi, O., Khosravi, H., Schunn, C. D., & Drachsler, H. (2025). Pedagogical framework for hybrid intelligent feedback. *Innovations in Education and Teaching International*, 1-17. <https://doi.org/10.1080/14703297.2025.2499174>
- Collis, B., & Moonen, J. (2012). *Flexible learning in a digital world: Experiences and expectations*. Routledge. <https://doi.org/10.4324/9780203046098>
- Da Silva, M. A. R., & Alliprandini, P. M. Z. (2025). Aprender a Aprender com Sistemas Aprimorados por Inteligência Artificial: Potencialidades para a correção da aprendizagem. *Revista Docência e Cibercultura*, 9(1), 1-20. <https://doi.org/10.12957/redoc.2024.83150>
- Du Boulay, B. (2023). Inteligência Artificial na Educação e Ética. *RE@D – Revista de Educação a Distância e eLearning*, 6(1), 1-17. <https://doi.org/10.34627/redvol6iss1e202303>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People-An ethical framework for a good AI society:

- Opportunities, risks, principles, and recommendations. *Minds and machines*, 28(4), 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science robotics*, 4(37), eaay7120. <https://www.science.org/doi/10.1126/scirobotics.aay7120>
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hoffmann, J. (2014). *Avaliação: Mito & desafio – Uma perspectiva construtivista* (41ª ed.). Mediação.
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Kenski, V. M. (2014). *Tecnologias e tempo docente*. Papirus Editora.
- Kumar, S., Datta, S., Singh, V., Datta, D., Singh, S. K., & Sharma, R. (2024). Applications, challenges, and future directions of human-in-the-loop learning. *IEEE Access*, 12, 75735-75760. <https://doi.org/10.1109/ACCESS.2024.3401547>
- Lima, T. C. S., & Mito, R. C. T. (2007). Procedimentos metodológicos na construção do conhecimento científico: A pesquisa bibliográfica. *Revista Katálisis*, 10, 37-45. <https://doi.org/10.1590/S1414-49802007000300004>
- Lordelo, J. A. C., & Dazzani, M. V. (2009). *Avaliação educacional: desatando e reatando nós*. EDUFBA.
- Luckesi, C. C. (2011). *Avaliação da aprendizagem: Componente do ato pedagógico*. São Paulo: Cortez.
- Martins, R. H., Viana, H. B., Barbosa, R., & Tavares, C. Z. (2024). O uso da inteligência artificial na educação: análise e percepção de professores. *Revista Intersaberes*, e24do3002. <https://doi.org/10.22169/revint.v19.e24do3002>
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6), Article 115. <https://doi.org/10.1145/3457607>
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267, 1-38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Morán, J. (2015). Mudando a educação com metodologias ativas. In C. A. de Souza & O. E. T. Morales (Orgs.), *Coleção mídias contemporâneas. Convergências midiáticas, educação e cidadania: Aproximações jovens* (Vol. 2, pp. 15-33). Ponta Grossa: Foca Foto–PROEX/UEPG.
- Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2023). Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review*, 56(4), 3005-3054. <https://doi.org/10.1007/s10462-022-10246-w>

- Nascimento, M. S., & Santos, Éber J. dos. (2025). Cidadania Digital e Inteligência Artificial: Estratégias Pedagógicas para o Desenvolvimento do Pensamento Crítico. *Revista de Estudos de Cultura*, 11(27), 163-179. <https://doi.org/10.32748/revec.v11i27.22710>
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2), 199-218. <https://doi.org/10.1080/03075070600572090>
- Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. In *Algorithms of oppression*. New York university press. <https://doi.org/10.18574/nyu/9781479833641.001.0001>
- Parreira, A., Lehmann, L., & Oliveira, M.. (2021). O desafio das tecnologias de inteligência artificial na Educação: percepção e avaliação dos professores. *Ensaio: Avaliação e Políticas Públicas em Educação*, 29(113), 975-999. <https://doi.org/10.1590/S0104-40362020002803115>
- Perrenoud, P. (2015). *Dez novas competências para ensinar*. Artmed editora.
- Santos, J., & Kroeff, M. (2018). A contribuição do feedback no processo de avaliação formativa. *EDUCA – Revista Multidisciplinar em Educação*, 5(11), 20-39. <https://doi.org/10.26568/2359-2087.2018.2776>
- Selwyn, N. (2019). *Should robots replace teachers?: AI and the future of education*. John Wiley & Sons.
- Sousa, A. S., Oliveira, G. S., & Alves, L. H. (2021). A pesquisa bibliográfica: Princípios e fundamentos. *Cadernos da FUCAMP*, 20(43), 64-83. <https://revistas.fucamp.edu.br/index.php/cadernos/article/view/2336>
- Thampi, A. (2022). *Interpretable AI: Building explainable machine learning systems*. Simon and Schuster.
- Vicari, R. M. (2021). Influências das Tecnologias da Inteligência Artificial no ensino. *Estudos Avançados*, 35, 73-84. <https://doi.org/10.1590/s0103-4014.2021.35101.006>
- Vicari, R. M., Brackmann, C., Mizusaki, L., & Galafassi, C. (2023). *Inteligência artificial na educação básica*. Novatec Editora.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators?. *International journal of educational technology in higher education*, 16(1), Article 39. <https://doi.org/10.1186/s41239-019-0171-0>

Recebido 07/09/2025
Aceite 01/10/2025
Publicado 11/10/2025

Este artigo está disponível segundo uma licença Creative Commons Attribution-NonCommercial 4.0 International License
