Universidade Federal de São João del-Rei
Núcleo de Educação a Distância
Re@d - Revista de Educação a Distância
ISSN: 2675-0651

AVALIAÇÃO DA APRENDIZAGEM NA EDUCAÇÃO A DISTÂNCIA DIANTE DA INTELIGÊNCIA ARTIFICIAL GENERATIVA: VALIDADE, INTEGRIDADE ACADÊMICA E DIRETRIZES AVALIATIVAS

Felippe Miranda da Silva[1]

RESUMO: A difusão de sistemas de inteligência artificial generativa modificou as condições em que estudantes da educação a distância produzem textos, códigos e soluções de problemas. Se uma tarefa assíncrona pode ser respondida por um modelo em segundos, o produto entregue deixa de ser evidência segura da aprendizagem pretendida. Este artigo analisa os desafios da avaliação da aprendizagem na educação a distância do ensino superior diante desse acesso e propõe estratégias que reduzam a vulnerabilidade a usos indevidos ou que incorporem a ferramenta com justificativa pedagógica. Trata-se de revisão bibliográfica narrativa, sem coleta de dados empíricos, articulando estudos sobre avaliação, alinhamento construtivo, taxonomia revisada de objetivos educacionais, integridade acadêmica, orientações internacionais sobre IA generativa e a Lei Geral de Proteção de Dados Pessoais. Sustenta-se que a validade depende sobretudo do desenho das evidências, e não da detecção retrospectiva de textos. Registro de processo, justificativa de decisões e uso declarado da ferramenta, quando o resultado de aprendizagem o exigir, oferecem caminho mais consistente do que a confiança em detectores, sujeitos a falsos positivos, vieses e implicações éticas e jurídicas. A contribuição é um conjunto de diretrizes e um quadro analítico que compara tipos de avaliação quanto à resistência relativa à substituição por IA, ao esforço docente, à escalabilidade, ao potencial formativo e ao risco de exclusão. Nenhuma estratégia é apresentada como absolutamente resistente.

PALAVRAS-CHAVE: Avaliação da aprendizagem. Educação a distância. Inteligência artificial generativa. Integridade acadêmica. Alinhamento construtivo.

1 Bacharel em Engenharia de Software e Pós-Graduado em Docência para a Educação Profissional e Tecnológica. E-mail: felippemiranda@outlook.com (Lattes)

ABSTRACT: The spread of generative artificial intelligence systems has changed the conditions under which distance-education students produce texts, code, and problem solutions. If an asynchronous task can be answered by a model in seconds, the submitted product is no longer secure evidence of the intended learning. This article examines the challenges of assessing learning in higher-education distance provision when such access is available, and it proposes strategies that either reduce vulnerability to misuse or incorporate the tool on pedagogical grounds. The study is a narrative literature review, with no empirical data collection. It brings together work on assessment, constructive alignment, the revised taxonomy of educational objectives, academic integrity, international guidance on generative AI, and Brazil's General Data Protection Law. The argument is that validity depends mainly on the design of evidence, rather than on retrospective detection of generated text. Process records, justification of decisions, and declared use of the tool, when the learning outcome so requires, offer a more consistent path than reliance on detectors, which are subject to false positives, bias, and ethical and legal implications. The contribution is a set of guidelines and an analytical framework comparing assessment types by relative resistance to substitution by AI, faculty effort, scalability, formative potential, and risk of exclusion. No strategy is presented as absolutely resistant.

KEYWORDS: Learning assessment. Distance education. Generative artificial intelligence. Academic integrity. Constructive alignment.

1. INTRODUÇÃO

A educação a distância no ensino superior organiza a mediação didática por tecnologias, com diálogo e estrutura que não coincidem, necessariamente, com a copresença contínua de docente e estudante (Moore, 1993; Litto; Formiga, 2009). Nessa modalidade, parte relevante das evidências de aprendizagem circula de forma assíncrona: fóruns, relatórios, códigos, projetos e provas realizadas fora de supervisão imediata. A disponibilização pública de sistemas de inteligência artificial generativa, a partir de 2022, tornou trivial a produção de textos e soluções formalmente plausíveis para muitos desses gêneros. O problema que daí decorre não é apenas disciplinar. É de validade. Se o produto pode ser obtido sem que o estudante tenha mobilizado a competência que o curso pretende certificar, a nota deixa de sustentar a inferência que a instituição comunica à sociedade.

A integridade acadêmica, entendida como compromisso com a honestidade e com a justiça do processo de demonstração do conhecimento (International Center for Academic Integrity, 2021), é tensionada quando o uso da ferramenta não é declarado e quando a tarefa foi desenhada para desempenho autoral sem auxílio. Ao mesmo tempo, a proibição genérica, difícil de fiscalizar na EaD, não define o que deve ser aprendido nem como essa aprendizagem será evidenciada. Cotton, Cotton e Shipway (2024) argumentam que a resposta institucional não pode restringir-se à detecção de textos gerados. A UNESCO (2023) orienta um uso de IA generativa centrado na agência humana, com proteção de dados e fortalecimento da capacidade docente, e não a mera interdição como política suficiente.

A questão norteadora deste artigo é: como avaliar de modo válido a aprendizagem na educação a distância do ensino superior quando o estudante tem acesso irrestrito a ferramentas de inteligência artificial generativa? O objetivo geral é analisar criticamente os desafios dessa avaliação e propor estratégias que reduzam a vulnerabilidade a usos indevidos ou que incorporem a IA de forma pedagogicamente justificada. Os objetivos específicos são: discutir funções, validade e confiabilidade da avaliação no ensino superior a distância; articular o alinhamento construtivo e a taxonomia revisada de Bloom como critérios para decidir o que avaliar; examinar implicações da IA generativa para a integridade acadêmica, inclusive limites dos detectores e proteção de dados; propor estratégias de maior resistência relativa e estratégias de incorporação justificada, com atenção a custo, escala e equidade; e formular diretrizes para instituições e coordenações de curso.

A justificativa é dupla. De um lado, a certificação superior produz consequências públicas: progressão, diplomação e sinalização de competência. De outro, turmas numerosas, tutoria enxuta e dispersão geográfica limitam a transposição mecânica de práticas presenciais de vigilância. Propõe-se, por isso, um exame que una o referencial pedagógico a analogias conceituais da engenharia de software, usadas como dispositivos de análise e não como evidência empírica. Revisão de código, testes, controle de versão, registro de decisões e depuração ajudam a distinguir produto final de processo verificável, e a tratar controles de segurança como camadas falíveis, não como garantias absolutas.

O texto organiza-se em referencial teórico, metodologia de revisão narrativa, discussão com proposições avaliativas e quadro analítico, diretrizes institucionais e considerações finais.

2. REFERENCIAL TEÓRICO

2.1 Avaliação da aprendizagem no ensino superior e na educação a distância

A avaliação da aprendizagem cumpre funções distintas, que convém não colapsar em um único ato de atribuição de nota. A função diagnóstica busca mapear conhecimentos prévios e dificuldades no início de um percurso, ou em um ponto de retomada, para ajustar o ensino. A função formativa organiza a coleta de evidências ao longo do processo, de modo que a informação seja usada para modificar atividades de ensino e estratégias de estudo (Black; Wiliam, 1998). A função somativa sintetiza, ao término de uma unidade ou de um componente curricular, o desempenho em relação a critérios, com fins de registro, progressão ou certificação. Scriven (1967) distinguiu o uso formativo do somativo no campo da avaliação de programas; na avaliação da aprendizagem, a distinção tornou-se central porque o mesmo instrumento pode servir a usos diferentes, conforme o momento e a consequência atribuída ao resultado.

Sadler (1989) mostrou que o feedback só se completa quando o estudante apreende o padrão de qualidade, compara seu trabalho a esse padrão e age para reduzir a distância. Sem essa apropriação, o comentário docente permanece informação externa. Na EaD, em que o intervalo entre a produção do estudante e o retorno do tutor pode ser longo, a fragilidade do ciclo formativo já era um problema de desenho. A IA generativa acrescenta outro: o produto que chega ao tutor pode ter sido ajustado por um modelo até parecer conforme ao padrão, sem que o estudante tenha realizado a comparação por conta própria. O sensor, na analogia com sistemas, passa a informar um

estado que não é o do aprendiz. Decisões pedagógicas tomadas sobre essa telemetria defeituosa, como a dispensa de revisão de um conceito, apoiam-se em evidência inválida.

Validade, neste artigo, designa a justificativa da inferência que se faz a partir da evidência coletada. Um instrumento é válido para um uso determinado quando o desempenho observado sustenta a conclusão sobre o resultado de aprendizagem pretendido. Confiabilidade designa a consistência dessa observação: avaliadores diferentes, ou o mesmo avaliador em momentos diferentes, chegam a julgamentos compatíveis, e o instrumento não oscila por fatores alheios ao construto. As duas qualidades não se substituem. Uma prova objetiva corrigida por gabarito pode ser altamente confiável e, ainda assim, inválida para inferir autoria intelectual se outra pessoa, ou um modelo, respondeu aos itens. Inversamente, uma arguição oral rica pode ter validade aparente e confiabilidade baixa se os critérios não forem explícitos e se o julgamento variar com o cansaço do avaliador ou com o sotaque do estudante.

Na educação a distância, três condições agravam esse desenho. A primeira é a distância transacional: a separação não é só geográfica, mas pedagógica, função do diálogo e da estrutura do curso (Moore, 1993). A segunda é a escala. Componentes com centenas de estudantes e equipes de tutoria reduzidas tornam custosa qualquer evidência que exija leitura longa ou interação síncrona universal. A terceira é a heterogeneidade de condições de estudo: trabalho, conectividade, deficiência, língua e familiaridade com tecnologias de escrita.

No plano normativo, a Lei nº 9.394/1996, em seu art. 80, prevê o incentivo à educação a distância em todos os níveis e modalidades. Para a graduação, a oferta é hoje disciplinada pelo Decreto nº 12.456/2025, que revogou o Decreto nº 9.057/2017, e pela Portaria MEC nº 506/2025, que o regulamenta, com as alterações da Portaria MEC nº 795/2025. Nos termos do art. 23 do decreto, a portaria determina que as unidades curriculares ofertadas total ou parcialmente a distância tenham duração mínima de dez semanas e incluam, no mínimo, uma avaliação de aprendizagem obrigatoriamente presencial, cabendo ao projeto pedagógico do curso prever as atividades formativas de realização presencial obrigatória. Essa exigência é um piso, não uma solução. Uma única avaliação presencial não resolve, por si, a validade das atividades online que compõem o restante da nota. Como esse marco tem sido alterado por atos sucessivos, a coordenação deve verificar a norma aplicável à sua oferta no momento da implementação. O recorte normativo considerado aqui é o da graduação, e não se reconstitui o detalhe de cada ato.

A analogia com testes de software esclarece um equívoco frequente. Testes automatizados aumentam a confiabilidade da verificação de comportamentos especificados, desde que os casos de teste representem o que se quer garantir. Uma suíte que só exercita exemplos públicos pode ser contornada por uma solução que passa nesses exemplos e falha fora deles. Na avaliação, o gabarito e a rubrica publicados são necessários à justiça e à aprendizagem do padrão (Sadler, 1989). Se, porém, a tarefa se esgota em reproduzir um padrão já integralmente exemplificado, um modelo generativo pode satisfazer os casos visíveis sem que o estudante domine o construto. A saída não é ocultar os critérios de qualidade. É variar a instância, exigir justificativa de decisões e, quando o resultado de aprendizagem for o desempenho sem auxílio, controlar as condições de realização. Critério público e dado inédito não são incompatíveis.

2.2 Alinhamento construtivo e taxonomia revisada como critérios do que avaliar

Biggs (1996) formulou o alinhamento construtivo como coerência entre resultados pretendidos de aprendizagem, atividades de ensino e tarefas de avaliação. Biggs e Tang (2011) desenvolveram essa coerência no contexto do ensino universitário, com ênfase no que o estudante faz, e não apenas no que o docente declara. O princípio é simples e exigente: a avaliação deve pedir a evidência daquilo que se disse querer que o estudante se tornasse capaz de fazer. Quando a tarefa pode ser cumprida por um procedimento que contorna essa capacidade, o alinhamento se rompe, ainda que o enunciado pareça sofisticado.

A taxonomia revisada por Anderson e Krathwohl (2001) oferece uma linguagem para nomear o que se pretende. No eixo dos processos cognitivos, distinguem-se lembrar, entender, aplicar, analisar, avaliar e criar. No eixo do conhecimento, distinguem-se o factual, o conceitual, o procedimental e o metacognitivo. A combinação importa mais do que o verbo isolado no plano de ensino. "Analisar" um caso cujo texto já circula em bases de treino de modelos não é o mesmo processo que analisar um conjunto de restrições elaborado para aquela oferta e defender, sob pergunta, por que uma alternativa foi rejeitada. O conhecimento metacognitivo, relativo a como se sabe e a como se regula o próprio desempenho, é particularmente relevante: a avaliação que só olha o texto final não acessa esse eixo.

Seria um erro concluir que processos mais altos na taxonomia são, por natureza, resistentes à IA. Modelos generativos produzem análises, avaliações e textos criativos de aparência competente, sobretudo quando o gênero é convencional. Também seria um erro tratar o lembrar como dispensável ou como única zona de risco. Há resultados de aprendizagem em que a recuperação sem

auxílio é parte legítima da competência profissional, por exemplo em procedimentos que precisam estar disponíveis sob pressão e sem consulta. Nesses casos, o alinhamento exige uma condição de realização compatível com o construto, e não uma redação online sobre o procedimento. Em outros resultados, o que se quer é precisamente a capacidade de usar ferramentas, verificar saídas e decidir. Proibir a IA nesses casos também rompe o alinhamento, porque a evidência passa a ser de um desempenho que o curso não declarou como alvo.

O critério prático que se extrai de Biggs (1996) e de Anderson e Krathwohl (2001) pode ser formulado assim: primeiro se explicita o resultado, incluindo se o desempenho alvo é com ou sem auxílio de IA; depois se escolhe a evidência que torna esse resultado observável; por fim se verifica se as atividades de ensino dão oportunidade de praticar exatamente isso. A IA generativa força a explicitação que muitos planos de ensino deixavam implícita. "Produzir um ensaio" não especifica se a autoria intelectual sem assistência é o construto, ou se o construto é a qualidade do argumento, admitido o uso de ferramentas de redação. Enquanto a especificação permanecer ambígua, a acusação de uso indevido e a permissão informal coexistirão sem critério, com dano à justiça.

Há um segundo uso da taxonomia, menos óbvio. Ela ajuda a localizar o que a nota está, de fato, medindo. Uma dissertação fluente pode evidenciar lembrar e entender de superfície, ou pode evidenciar avaliar, conforme as exigências de justificativa, de confronto com fontes e de tratamento de objeções. Se a rubrica premia sobretudo coesão e vocabulário, o instrumento mede um construto estreito, hoje facilmente assistido por modelos. Ajustar a rubrica para verificação de fontes, distinção entre dado e inferência e adequação a restrições locais muda o construto medido. Essa mudança é uma decisão pedagógica, não um truque de segurança. Ela deve ser coerente com o que se ensinou.

2.3 Inteligência artificial generativa, integridade acadêmica e proteção de dados

O International Center for Academic Integrity (2021) descreve a integridade acadêmica como compromisso com valores como honestidade, confiança, justiça, respeito, responsabilidade e coragem. A definição não se esgota na ausência de fraude. Inclui as condições institucionais que tornam o trabalho honesto possível e a avaliação justa. Dawson (2021) distingue, de modo útil, integridade e segurança da avaliação. A primeira remete ao compromisso ético e à cultura institucional. A segunda remete às condições que dificultam a fraude e protegem a validade das inferências. O livro é anterior à popularização ampla dos modelos generativos de grande porte e trata de formas já conhecidas de fraude digital. A distinção, ainda assim, evita dois reducionismos: o

de supor que um código de honra substitui o desenho da tarefa, e o de supor que controles técnicos substituem a formação ética.

Formas de uso indevido, para os fins deste artigo, incluem a submissão de texto, código ou solução gerados por IA como se fossem produção própria, quando a tarefa exigia desempenho sem esse auxílio ou exigia declaração; a fabricação de referências e de dados, favorecida por saídas que inventam fontes com aparência bibliográfica correta; a paráfrase de saída de modelo com o propósito exclusivo de contornar detectores; o uso de um segundo dispositivo durante avaliação síncrona não supervisionada; e a combinação de terceirização humana com geração automática. Nem todo uso de IA é uso indevido. A classificação depende do resultado de aprendizagem declarado e das regras comunicadas antes da tarefa. Uso não declarado, em contexto de proibição, viola a honestidade mesmo que o conteúdo esteja correto. Uso declarado, em contexto de permissão, pode ainda ser pedagogicamente pobre se o estudante não verificar a saída. São falhas de natureza diferente e pedem respostas diferentes.

Detectores de texto gerado por IA têm sido adotados como atalho de segurança. A literatura disponível não autoriza essa confiança. Liang et al. (2023) relataram viés de detectores contra escritores não nativos de inglês, com elevação de falsos positivos sobre textos humanos. Weber-Wulff et al. (2023) examinaram ferramentas de detecção e relataram limites de acurácia e de confiabilidade. Não se transpõem esses achados, de modo automático, ao português ou às variedades linguísticas do Brasil. O mecanismo descrito, porém, justifica cautela: detectores exploram regularidades estatísticas do texto, e escritos de quem produz em língua adicional, ou em variedades pouco representadas nos dados de treino, podem afastar-se do padrão esperado para texto humano "típico". Tratar escore de detector como prova de autoria expõe estudantes a erro, e pode concentrar o erro sobre quem já escreve em condições desiguais. A corrida entre geradores, paráfrases e detectores tende a tornar o instrumento obsoleto em ciclos curtos, com custo recorrente e sem ganho estável de validade.

Há implicações éticas e jurídicas. Uma acusação de má conduta com base apenas em probabilidade algorítmica inverte o ônus de modo difícil de contestar, porque o estudante não acessa o modelo de detecção nem consegue, em geral, provar autoria negativa. A Lei nº 13.709/2018, Lei Geral de Proteção de Dados Pessoais, incide quando textos, nomes, registros acadêmicos, gravações de voz ou imagens são tratados de forma que identifique ou torne identificável uma pessoa. O art. 5º define dado pessoal e inclui dados biométricos entre os dados pessoais sensíveis. O art. 6º

estabelece princípios como finalidade, adequação, necessidade, transparência, segurança e responsabilização. O art. 7º lista bases legais, cuja escolha cabe à instituição com assessoria jurídica, e não se pretende esgotá-la aqui. O art. 20 assegura ao titular o direito de solicitar revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses. Uma reprovação fundada somente em escore de detector aproxima-se dessa hipótese e, além disso, é metodologicamente frágil.

O envio de produções estudantis a ferramentas de detecção ou de geração hospedadas por terceiros, por vezes no exterior, implica finalidade específica, informação ao titular e atenção à transferência internacional prevista no art. 33. O art. 46 exige medidas de segurança compatíveis com o risco. Gravações de arguições orais são dados pessoais; se usadas para identificação biométrica, aproximam-se de dado sensível. A conservação deve ser limitada ao propósito de registro avaliativo e de eventual recurso, não convertida em base de treino de modelos sem base legal e sem transparência. A minimização recomenda não adotar vigilância contínua por *proctoring* como padrão, quando um desenho de tarefa menos intrusivo puder proteger a inferência avaliativa. Essa posição não decorre de ingenuidade quanto à fraude. Decorre da desproporção entre o dado coletado e a garantia obtida, já que a supervisão por imagem também é contornável e produz seus próprios falsos indícios.

Freire (1996) sustenta que ensinar exige rigor metódico, pesquisa, criticidade e respeito à autonomia de quem aprende. A referência não transforma Freire em teórico da IA generativa, tema inexistente em sua obra. Serve para recusar duas saídas pobres: a avaliação como dispositivo de suspeita, e a autonomia como licença para entregar trabalho não compreendido. O rigor, no contexto atual da EaD, inclui saber o que a evidência mostra e o que ela não mostra. A UNESCO (2023) converge, em outro registro, ao enfatizar agência humana, proteção de dados e capacitação docente. Cotton, Cotton e Shipway (2024) reforçam a necessidade de redesenhar avaliações e de explicitar expectativas, em lugar de depositar na detecção a preservação da integridade.

3. METODOLOGIA

Este artigo é teórico-propositivo. O procedimento adotado é a revisão bibliográfica narrativa, adequada à articulação de campos que não compartilham o mesmo vocabulário: avaliação da aprendizagem, educação a distância, integridade acadêmica, proteção de dados e práticas da engenharia de software. Não houve coleta de dados empíricos, aplicação de instrumentos, entrevistas, observação de turmas ou análise de produções estudantis. Exemplos que aparecem na

discussão são ilustrações hipotéticas, construídas para tornar visível um *trade-off*, e não relatos de casos.

As fontes que informam a argumentação são de tipos distintos: livros e artigos de periódicos sobre avaliação e sobre integridade acadêmica; literatura de referência em engenharia de software, usada apenas para ancorar analogias; relatório de organismo internacional sobre IA generativa na educação e na pesquisa; legislação brasileira (Lei nº 9.394/1996, Decreto nº 12.456/2025, Portaria MEC nº 506/2025, Portaria MEC nº 795/2025, Lei nº 13.146/2015 e Lei nº 13.709/2018); e documento de referência sobre valores da integridade acadêmica. A seleção obedeceu a três critérios: pertinência conceitual ao problema da validade avaliativa diante do acesso a modelos generativos; autoridade da fonte, entendida como obra de referência na área, artigo em periódico com revisão por pares, documento de organismo reconhecido ou norma oficial; e verificabilidade da fonte por DOI, ISBN ou publicação oficial. Não se adotou protocolo de revisão sistemática. Não há, portanto, *strings* de busca, recorte temporal operacionalizado, contagem de registros recuperados, números de estudos incluídos ou excluídos, nem avaliação formal de risco de viés. Bases usuais de periódicos e catálogos de livros foram o horizonte de localização das fontes, sem que se pretenda descrever uma busca reprodutível.

Essa opção tem consequências. A revisão narrativa depende do julgamento de quem seleciona e está sujeita a omissão de trabalhos relevantes, em especial literatura cinzenta institucional, produção muito recente e estudos empíricos locais ainda não consolidados. A velocidade de mudança das ferramentas também limita qualquer descrição de comportamento de modelos específicos. Por isso, as proposições referem-se a propriedades relativamente estáveis do problema (acesso do estudante a um gerador capaz de produzir gêneros escolares, falibilidade da detecção, necessidade de alinhamento), e não a um produto comercial. As analogias com revisão de código, testes, controle de versão, registro de decisões e depuração são instrumentos de raciocínio. Práticas de revisão, teste, gerência de configuração e documentação de projeto são tratadas na engenharia de software (Sommerville, 2016). O uso que aqui se faz delas não constitui transferência metodológica testada, nem evidência de que produzam os mesmos efeitos na educação.

A análise procedeu por confronto entre o problema de validade e os conceitos do referencial. Para cada estratégia proposta, examinaram-se cinco dimensões, definidas de antemão: resistência relativa à substituição integral do desempenho por IA; esforço docente; escalabilidade em turmas grandes de EaD; potencial formativo; e risco de exclusão, que abrange acessibilidade e equidade. A

classificação apresentada nos Quadros 1 e 2 da seção 4.3 é analítica. Não resulta de medição, de painel de especialistas ou de experimento. Os níveis ordinais (baixa, moderada, alta) têm definição operacional no Quadro 2, organizam a comparação e devem ser lidos no contexto das ressalvas do próprio quadro.

4. DISCUSSÃO E PROPOSIÇÕES

A questão norteadora não admite uma técnica única. Admitir acesso irrestrito do estudante à ferramenta, condição realista na EaD assíncrona, implica escolher o que a evidência ainda pode mostrar. Duas famílias de estratégia decorrem do alinhamento construtivo. Na primeira, o resultado de aprendizagem é o desempenho sem auxílio, e a avaliação procura tornar custosa a substituição integral, sem prometer imunidade. Na segunda, o resultado inclui o uso crítico da ferramenta, e a avaliação passa a pedir evidências desse uso. Ambas podem coexistir no mesmo componente curricular, aplicadas a resultados diferentes. Nenhuma é absolutamente resistente à IA. Voz sintética, auxílio em tempo real, terceirização humana e geração de históricos falsos de processo contornam, com custo variável, cada controle isolado. O que se pode sustentar é defesa em camadas: cada camada cobre uma falha da anterior e todas têm custo.

4.1 Estratégias com maior resistência relativa à substituição por IA

A arguição oral síncrona curta pede que o estudante explique uma decisão já registrada por escrito: por que rejeitou uma alternativa, como um dado do enunciado altera a conclusão, o que significaria um termo que empregou. A analogia é a depuração. Não se pergunta apenas se o programa "passou". Pergunta-se pelo caminho de execução. Uma ilustração hipotética, não observada em campo: em uma disciplina de análise de requisitos, o estudante entrega um registro de decisão sobre um conjunto de atas fictícias, elaboradas pelo docente para aquela oferta. Na arguição de poucos minutos, o avaliador altera uma restrição e pede que o estudante diga o que muda. Quem compreendeu a decisão responde com hesitação produtiva. Quem apenas leu um texto gerado tende a repetir a conclusão original. O indício não é prova. Estudantes ansiosos, ou pouco fluentes na variedade linguística valorizada pela banca, também hesitam. O critério precisa incidir sobre a coerência da explicação com o material, não sobre desenvoltura.

O custo docente é alto se a arguição for universal em turmas de centenas de estudantes. A escalabilidade melhora com amostragem informada pelo risco: componentes de maior peso, entregas com inconsistência interna, ou um sorteio anunciado previamente. A amostragem não

autoriza generalizar suspeita. Autoriza verificar, com critério público, uma parte das evidências, como equipes de engenharia aprofundam a revisão em módulos críticos e não em cada linha com a mesma intensidade. A equidade é o limite principal. Conexão instável, jornada de trabalho, fuso, deficiência, necessidade de intérprete de Libras e ansiedade de desempenho podem tornar a sessão síncrona excludente. A Lei nº 13.146/2015 assegura educação em sistema inclusivo. Uma arguição oral de alto peso, sem alternativa equivalente, concentra a desvantagem. Propõe-se peso limitado, agendamento com antecedência, gravação apenas quando necessária ao registro e ao recurso, e alternativa de igual construto, como defesa em duas etapas ou prova supervisionada, para quem não puder participar em condições justas. A gravação, se adotada, submete-se à LGPD, com finalidade restrita e prazo de conservação definido.

A avaliação de processo, por portfólios, entregas parciais e versionamento, desloca a evidência do arquivo final para a trajetória. No controle de versão, o histórico mostra o que mudou e a mensagem de versão declara a intenção da mudança. Transposta à avaliação, a prática pede rascunhos datados no ambiente institucional, resposta a um comentário específico do tutor e breve nota sobre o que foi alterado e por quê. O potencial formativo é alto, porque o feedback incide sobre uma versão que ainda pode ser revisada (Black; Wiliam, 1998). A resistência à IA é apenas moderada. Um estudante pode gerar a sequência inteira de versões, simulando evolução. O controle se fortalece quando uma entrega parcial recebe uma restrição nova, impossível de satisfazer com o texto já pronto, e quando a versão seguinte precisa incorporar essa restrição de modo localizável. O esforço docente cresce com o número de marcos. Em turmas grandes, a viabilidade depende de marcos poucos, rubricas enxutas e tutoria capacitada, não de comentário livre ilimitado. Do ponto de vista da equidade, o versionamento em ferramenta pouco acessível, ou que exija banda e familiaridade técnica não ensinadas no curso, cria barreira. O ambiente virtual já usado pela instituição, com orientação clara, é preferível a um repositório externo que ainda transfere dados desnecessariamente.

Tarefas contextualizadas com dados inéditos elevam o custo da resposta genérica. O modelo é capaz de produzir texto sobre um tema público. É menos capaz de respeitar, sem alucinação, um conjunto de restrições que não circulava antes, desde que o avaliador leia a adequação a essas restrições e não apenas a fluência. A ilustração continua hipotética: um caso com números, nomes e normas internas inventados para a oferta, cuja resposta correta depende de uma incompatibilidade plantada entre dois anexos. A rubrica falha o trabalho que ignore a incompatibilidade, ainda que bem escrito. O esforço está na elaboração e na renovação dos dados a cada oferta, e na correção

orientada por esse detalhe. A escalabilidade é moderada se o caso for o mesmo para a turma e a correção for criterial, e cai se cada estudante receber um caso totalmente distinto sem apoio de tutores. A equidade melhora quando o contexto não pressupõe capital cultural não ensinado, como familiaridade com uma empresa específica ou com um dialeto profissional não trabalhado em aula. Dado inédito não significa enigmático. Significa não resolvível por recuperação de um texto já publicado.

Provas presenciais ou supervisionadas, em polo ou em condição controlada, continuam sendo o meio mais direto de avaliar desempenho sem auxílio, quando esse é o construto. Na graduação a distância, a norma vigente já exige ao menos uma avaliação de aprendizagem obrigatoriamente presencial por unidade curricular (Portaria MEC nº 506/2025). Isso oferece base normativa para essa camada, mas não define seu peso nem seu construto, que são escolhas pedagógicas. A resistência é relativa, não absoluta: cola, substituição de pessoa e falhas de supervisão persistem. O esforço e a *log*ística são altos, e a escalabilidade depende da rede de polos e da regulamentação aplicável, que a coordenação deve checar no momento da implementação. A equidade geográfica é sensível. Exigir deslocamento longo para uma prova de baixo peso pode ser desproporcional. Um arranjo mais defensável combina avaliação presencial ou supervisionada nos resultados que exigem desempenho sem auxílio com atividades formativas online de menor peso certificador, nas quais o uso de materiais e, quando declarado, de IA pode ser parte do estudo. Usar a prova presencial apenas como punição surpresa, dissociada do que foi praticado, rompe o alinhamento e o princípio de que o estudante deve conhecer o padrão (Sadler, 1989).

A análise de decisões e de *trade-offs* pede que o estudante explicite opções, critérios e perdas aceitas. Não se avalia só a recomendação final. Avalia-se se as alternativas eram plausíveis e se o critério de escolha responde ao problema. Essa estratégia se aproxima do registro de decisão usado em projetos de software, em que contexto, decisão, alternativas e consequências ficam documentados para revisão posterior. A IA redige registros plausíveis. A resistência aumenta se os critérios de escolha forem registrados no ambiente institucional antes do prazo em que a consulta à ferramenta seria permitida, e se a arguição amostral incidir sobre uma alternativa rejeitada. O esforço docente está em ler a justificativa, não em detectar estilo de máquina. A escalabilidade é moderada com rubrica focada em três ou quatro decisões, não em relatórios longos. Estudantes com menos experiência em escrita acadêmica podem demonstrar decisão em tabela curta ou em áudio, desde que o construto permaneça o raciocínio e não a performance do gênero mais valorizado.

Convém repetir o limite comum. Arguição pode ser assistida por outro dispositivo. Versionamento pode ser fabricado. Dado inédito pode ser inserido no modelo pelo próprio estudante. Supervisão presencial falha. *Trade-off* pode ser gerado. A afirmação de que existiriam avaliações à prova de IA não se sustenta. O que estas estratégias fazem, quando combinadas e alinhadas ao resultado, é tornar a substituição integral mais cara do que a aprendizagem e devolver à nota uma inferência mais honesta. Onde o custo docente ou a exclusão de estudantes ultrapassar o ganho de validade, a estratégia deve ser reduzida em peso ou substituída, não transformada em ritual.

4.2 Estratégias que incorporam a IA de modo pedagogicamente justificado

Há resultados para os quais o uso da ferramenta não é desvio, e sim objeto de aprendizagem. Formar quem saiba revisar uma saída automática é coerente com práticas profissionais em que texto, código e síntese já circulam assistidos por modelos. O alinhamento, nesse caso, exige ensinar a revisão e avaliar a revisão. Proibir na prova o que se exigirá no exercício profissional, sem que o construto seja o desempenho sem auxílio, produz uma evidência ornamental.

A analogia pertinente é a revisão de código. Um par propõe uma alteração; quem revisa procura defeito, premissa não declarada, tratamento ausente de caso de borda e desacordo com os requisitos. Não se elogia o *patch* porque compila ou porque o estilo parece maduro. Na avaliação, o estudante recebe ou gera uma saída, declara a ferramenta em termos gerais e produz um parecer: o que está correto, o que é alucinação ou generalidade inútil, o que contradiz o enunciado, o que precisou ser reescrito e por quê. A rubrica incide sobre a qualidade dessa revisão. Uma ilustração hipotética: o docente fornece duas saídas de modelo para o mesmo problema, uma com citação inexistente e outra com uma restrição ignorada. O estudante deve localizar as falhas e decidir o que aproveitar. Quem apenas "melhorar a redação" não atinge o critério. O construto medido é avaliar, no sentido de Anderson e Krathwohl (2001), aplicado a um artefato específico.

O registro transparente do uso aproxima-se do controle de versão e do registro de decisão. Pede-se que o estudante conserve *prompts* relevantes, versões aceitas e rejeitadas e uma nota curta sobre o critério de aceitação. Não se trata de vigiar a intimidade da interação nem de exigir *dump* integral de conversas, o que ampliaria dados pessoais sem necessidade, em tensão com o art. 6º da LGPD. Trata-se de um *log* mínimo, suficiente para a revisão humana. A avaliação pode amostrar esse *log*, como se inspeciona o histórico de um trecho sensível, não de todo o repositório.

Estudantes devem ser informados, antes da tarefa, do que será armazenado, por quanto tempo e com que finalidade.

Rubricas para esse tipo de tarefa ganham clareza se forem escritas como casos de teste, não como impressões de qualidade. Casos possíveis, a serem adaptados ao componente: a afirmação factual foi verificada em fonte indicada pelo docente ou localizada pelo estudante; a saída que contradiz o enunciado foi rejeitada; a escolha entre duas saídas foi justificada por um critério explícito; o texto final distingue o que veio da ferramenta e o que foi decisão do estudante. A analogia com testes tem um alerta já mencionado. Testes automáticos de múltipla escolha, aplicados sem supervisão e com alto peso, são vulneráveis precisamente porque o modelo resolve o item e o gabarito não observa o processo. A suíte de critérios proposta aqui é um instrumento de julgamento humano e de autoavaliação, no sentido de Sadler (1989), não uma recomendação de ampliar provas objetivas assíncronas como defesa contra a IA.

Acordos de uso por disciplina tornam a regra parte do contrato pedagógico, e não um anexo genérico de instituição. O acordo especifica, para cada tipo de tarefa, se a IA está vedada, permitida para planejamento, permitida para revisão linguística ou exigida como objeto de análise. Escalas desse tipo têm sido propostas na literatura. Perkins et al. (2024) apresentaram a Artificial Intelligence Assessment Scale (AIAS) como quadro para explicitar o grau de uso permitido em cada avaliação, em cinco níveis que vão da ausência de IA ao uso amplo com responsabilização de quem entrega. Uma versão refinada da escala foi publicada posteriormente (Perkins; Roe; Furze, 2025). Não se reproduz aqui a escala integral. O ponto que se aproveita é a explicitação graduada, coerente com o alinhamento: a permissão é função do resultado, não um clima geral de tolerância ou de proibição. A coordenação pode adotar essa escala ou graduação própria, desde que pública, estável ao longo da oferta e ensinada antes de ser cobrada.

A incorporação da IA também tem custo, escala e equidade. Corrigir pareceres críticos exige leitor humano capacitado; não é mais barato do que corrigir ensaio, embora possa ser mais válido para o construto desejado. Em turmas grandes, a tarefa precisa ser curta e a rubrica, compartilhada com tutores, com moderação de amostras para reduzir divergência de julgamento. A equidade de acesso à ferramenta não pode ser ignorada. Se a tarefa exige um sistema pago, a instituição precisa oferecer alternativa gratuita suficiente ou acesso institucional. Caso contrário, a nota mede assinatura, não aprendizagem. Há ainda assimetria de letramento. Estudantes com mais familiaridade obtêm saídas melhores e parecem mais competentes na revisão. O ensino do próprio

procedimento de verificação, com exemplos de falha, é condição do alinhamento, não suplemento opcional. Freire (1996) é pertinente aqui no que diz respeito ao rigor: autonomia sem método de leitura crítica não é objetivo formativo compatível com o ensino superior.

Incorporar a IA não dispensa momentos de desempenho sem auxílio, se esses momentos corresponderem a resultados reais do curso. Um componente de programação pode avaliar, com uso declarado, a capacidade de revisar código gerado e, em condição supervisionada, a capacidade de ler e modificar um trecho sem assistência. São construtos diferentes. Fundi-los em uma única nota opaca impede que o estudante e a instituição saibam o que foi certificado.

4.3 Quadro-síntese analítico

O Quadro 1 compara os tipos de avaliação usados na discussão segundo cinco dimensões, cujos níveis são definidos no Quadro 2. A classificação é analítica, elaborada para este artigo. Não resulta de medição empírica, ensaio ou consulta formal a especialistas. Resistência relativa indica dificuldade e custo de substituir integralmente o desempenho do estudante, não imunidade. Os níveis são ordinais e dependem de implementação. Uma arguição mal conduzida pode ter resistência efetiva baixa. Um portfólio com restrições novas a cada marco pode ultrapassar a classificação moderada aqui atribuída. O risco de exclusão indica a probabilidade de a estratégia impor barreiras a estudantes por razões de conectividade, deslocamento, deficiência, jornada de trabalho ou fluência, e por isso deve ser lido em sentido inverso às demais colunas: quanto maior, mais necessária a alternativa de mesmo construto.

Quadro 1 – Comparação analítica de tipos de avaliação

Tipo de avaliação	Resistência relativa à IA	Esforço docente	Escalabilidade na EaD	Potencial formativo	Risco de exclusão
Prova objetiva assíncrona, sem supervisão	Baixa	Baixo, se a correção for automática	Alta	Baixo, quando o retorno se limita ao escore	Baixo a moderado
Dissertação assíncrona, sem supervisão	Baixa a moderada	Alto	Moderada a baixa em turmas grandes	Moderado, se houver feedback qualitativo	Moderado

Tipo de avaliação	Resistência relativa à IA	Esforço docente	Escalabilidade na EaD	Potencial formativo	Risco de exclusão
Prova presencial ou supervisionada	Moderada a alta, para desempenho sem auxílio	Alto na logística	Baixa a moderada	Baixo a moderado	Alto
Arguição oral síncrona curta	Moderada a alta	Alto por estudante; menor se amostral	Baixa se universal; moderada se amostral	Alto, se a pergunta seguir a resposta	Alto, mitigável com alternativa de mesmo construto
Portfólio com versões e entregas parciais	Moderada	Alto	Baixa a moderada	Alto	Moderado
Tarefa com dados inéditos e *trade-offs*	Moderada	Alto na elaboração e na correção	Moderada	Moderado a alto	Baixo a moderado
Análise crítica de saída de IA, com uso declarado	Não se aplica ao construto "sem auxílio"; mais adequada ao construto "avaliar saídas"	Moderado a alto	Moderada	Alto	Moderado
Registro de decisões com critérios prévios	Moderada	Moderado a alto	Moderada	Alto	Baixo a moderado

Fonte: elaboração própria.

Quadro 2 – Critérios de classificação ordinal das dimensões analisadas

Dimensão	Baixa (ou baixo)	Moderada (ou moderado)	Alta (ou alto)
Resistência relativa à substituição por IA	A substituição integral é possível em minutos, sem interlocução com o estudante e sem vestígio verificável do processo.	A substituição exige coordenar geração com material inédito, versões ou explicação posterior, o que eleva o custo, mas ainda é viável com auxílio em tempo real ou terceirização.	O desempenho é produzido sob condição controlada ou observado em interlocução em tempo real; substituir exige burlar a supervisão ou a arguição.
Esforço docente	Correção automática ou rubrica curta, sem leitura individual.	Leitura individual de produtos curtos, ou de amostra.	Leitura individual extensa, interação síncrona por estudante ou logística presencial.
Escalabilidade na EaD	O custo cresce com o número de estudantes e supera as equipes de tutoria usuais.	Exige amostragem, rubrica compartilhada e moderação entre tutores.	Funciona com centenas de estudantes sem ampliar a equipe.
Potencial formativo	O retorno se limita ao escore.	Há feedback qualitativo, mas após a entrega final.	O feedback incide sobre versão ainda revisável, ou a pergunta se ajusta à resposta.
Risco de exclusão	Poucas barreiras de tempo, conexão, deslocamento ou fluência oral e escrita.	Há barreiras contornáveis com alternativa de mesmo construto.	Depende de deslocamento, sincronia ou desempenho oral; exige alternativa equivalente obrigatória.

Fonte: elaboração própria.

A leitura do quadro desaconselha duas políticas simétricas. A primeira é manter como eixo certificador a dissertação ou a prova objetiva assíncronas, de alta escalabilidade e baixa resistência, e compensar a fragilidade com detector. A segunda é converter toda a avaliação em arguição individual, ganhando resistência aparente e perdendo escala, inclusão e até validade, se o construto

passar a ser desenvoltura oral. O arranjo mais coerente com o referencial combina poucas evidências de maior resistência nos resultados sem auxílio, evidências de processo com função formativa e, quando o currículo o justificar, tarefas que avaliem o uso crítico da ferramenta. O peso certificador deve acompanhar a validade do instrumento, não a facilidade de coleta.

4.4 Diretrizes para instituições e coordenações de curso

Da discussão anterior derivam diretrizes operacionais. Elas não substituem assessoria jurídica, regulação setorial nem o projeto pedagógico do curso.

A política de uso de IA precisa existir no nível institucional e ser especificada no nível da disciplina. A norma geral define valores, consequências proporcionais, vedação de acusação baseada só em detector e dever de transparência no tratamento de dados. A norma da disciplina define, tarefa a tarefa, o uso vedado, permitido ou exigido. Comunicação posterior à entrega não recupera a justiça do processo. O plano de avaliação, divulgado no início da oferta, é o lugar dessa especificação.

Expectativas se ensinam. Um acordo escrito que os estudantes não tiveram oportunidade de praticar reproduz a lógica punitiva que se quer limitar. Atividades formativas de baixo peso, com feedback sobre o que conta como declaração adequada, verificação de fonte e justificativa de rejeição de uma saída, cumprem a função descrita por Black e Wiliam (1998): a informação serve para ajustar o desempenho antes da consequência somativa. O letramento em IA, para docentes e estudantes, inclui limites conhecidos dos modelos, como invenção de referências, e limites conhecidos dos detectores, como falso positivo e viés (Liang et al., 2023; Weber-Wulff et al., 2023). Inclui também leitura da LGPD em linguagem operacional: que dados da turma serão enviados a qual tipo de ferramenta, sob qual finalidade, e como exercer pedido de revisão.

A formação docente não se reduz a oficina de *prompts*. Envolve redesenho de resultados de aprendizagem, escrita de rubricas alinhadas, decisão sobre o que exige condição supervisionada e trabalho em equipe com tutores para consistência de julgamento. Sem esse tempo, a política institucional permanece documento. Coordenações de EaD, ao demandar inovação avaliativa, precisam considerar carga de correção. Uma diretriz que aumente validade e ignore escala será contornada na implementação, com retorno silencioso à tarefa assíncrona de sempre.

A cultura de integridade, no sentido do International Center for Academic Integrity (2021) e da distinção de Dawson (2021), não se confunde com campanha de ameaça. Procedimentos de

apuração devem ser contraditórios, proporcionar defesa, separar indício de prova e prever revisão humana quando houver tratamento automatizado, em coerência com o art. 20 da Lei nº 13.709/2018. Sanções proporcionais convivem com oportunidade de refazer evidência quando o problema for de aprendizagem do padrão, não de fraude deliberada. Freire (1996) oferece o critério ético dessa distinção, não um protocolo jurídico: rigor sem humilhação, autonomia com responsabilidade.

Por fim, acessibilidade integra a diretriz, e não um anexo. Alternativas equivalentes de construto, não de formato idêntico, devem existir quando a estratégia de maior resistência excluir estudantes por deficiência, conectividade ou geografia. Ferramentas exigidas na avaliação precisam ser acessíveis sem custo oculto. Dados de voz, imagem e texto identificável não devem ser tratados além do necessário à finalidade avaliativa. Essas condições não decoram a proposta. Determinam se a validade ganha para uns é paga com invalidez e exclusão para outros.

5. CONSIDERAÇÕES FINAIS

A questão norteadora perguntava como avaliar de modo válido a aprendizagem na EaD do ensino superior quando o estudante tem acesso irrestrito a ferramentas de IA generativa. A resposta que este artigo sustenta é a seguinte. A validade não se restaura por detecção retrospectiva do texto, nem por proibição genérica difícil de verificar, nem por uma forma avaliativa supostamente imune. Restaura-se, de modo sempre parcial, quando o resultado de aprendizagem explicita se o desempenho alvo ocorre com ou sem auxílio, e quando a evidência é desenhada para esse alvo. Se o alvo é o desempenho sem auxílio, combinações de condição supervisionada, arguição curta, dados inéditos e registro de processo aumentam o custo da substituição, com limites de escala e de equidade que impedem universalizá-las. Se o alvo inclui julgamento sobre saídas de modelos, o uso declarado, o *log* mínimo e a rubrica de verificação é que tornam a nota interpretável. O alinhamento construtivo (Biggs, 1996; Biggs; Tang, 2011) e a taxonomia revisada (Anderson; Krathwohl, 2001) são os critérios para essa escolha. As analogias com revisão, teste, versionamento, registro de decisão e depuração serviram para nomear a diferença entre produto fluente e evidência inspecionável, e para recusar o controle único.

A contribuição é conceitual e propositiva: um enquadramento que articula validade, integridade e proteção de dados; um conjunto de estratégias com seus *trade-offs*; um quadro analítico de comparação; e diretrizes para instituições e coordenações. Não se oferece protocolo pronto para adoção acrítica. A contribuição da dupla lente, pedagógica e de engenharia de software,

está na recusa de soluções que um dos campos, isolado, tenderia a aceitar. A pedagogia sem atenção a falha de controles subestima a fraude. A engenharia de controles sem atenção a construto, inclusão e finalidade do dado produz vigilância que não mede aprendizagem.

As limitações são as da revisão narrativa e do caráter não empírico das proposições. Não se testou se as estratégias melhoram inferências avaliativas, se são suportáveis por equipes de tutoria, nem se reduzem inequidades. Não se examinou o comportamento de detectores em português, nem se quantificou falso positivo sobre variedades linguísticas ou sobre escritores com trajetórias escolares desiguais. A regulação da EaD e os sistemas generativos mudam em ritmo incompatível com afirmações muito específicas sobre ferramentas.

Pesquisas futuras podem submeter as estratégias a estudo empírico de validação, com desenhos que comparem inferências avaliativas, carga docente e efeitos de equidade, sem tratar resistência à IA como único critério de sucesso. Estudos sobre detectores em português, com atenção a falsos positivos e a consequências disciplinares, são necessários antes de qualquer uso institucional desses sistemas. Também são relevantes investigações sobre o efeito de escalas de uso de IA na compreensão das regras por estudantes e tutores, e sobre arranjos de amostragem oral que preservem validade sem excluir quem estuda em condições precárias de conexão. Enquanto essas evidências não existirem, a prudência recomendável é explícita: declarar o construto, alinhar a tarefa, não decidir carreira acadêmica de estudante com base apenas em escore de detector e não chamar de à prova de IA um controle que apenas eleva o custo da substituição.

REFERÊNCIAS

ANDERSON, Lorin W.; KRATHWOHL, David R. (org.). A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York: Longman, 2001.

BIGGS, John. Enhancing teaching through constructive alignment. Higher Education, v. 32, n. 3, p. 347-364, 1996. Disponível em: https://doi.org/10.1007/BF00138871. Acesso em: 28 set. 2026.

BIGGS, John; TANG, Catherine. Teaching for quality learning at university: what the student does. 4. ed. Maidenhead: Open University Press, 2011.

BLACK, Paul; WILIAM, Dylan. Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, v. 5, n. 1, p. 7-74, 1998. Disponível em: https://doi.org/10.1080/0969595980050102. Acesso em: 28 set. 2026.

BRASIL. Lei nº 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. Diário Oficial da União: seção 1, Brasília, DF, 23 dez. 1996. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9394.htm. Acesso em: 28 set. 2026.

BRASIL. Lei nº 13.146, de 6 de julho de 2015. Institui a Lei Brasileira de Inclusão da Pessoa com Deficiência — Estatuto da Pessoa com Deficiência. Diário Oficial da União: seção 1, Brasília, DF, 7 jul. 2015. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13146.htm. Acesso em: 28 set. 2026.

BRASIL. Decreto nº 9.057, de 25 de maio de 2017. Regulamenta o art. 80 da Lei nº 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional. Diário Oficial da União: seção 1, Brasília, DF, 26 maio 2017. Revogado pelo Decreto nº 12.456, de 2025. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/decreto/d9057.htm. Acesso em: 28 set. 2026.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais — LGPD. Diário Oficial da União: seção 1, Brasília, DF, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 28 set. 2026.

BRASIL. Decreto nº 12.456, de 19 de maio de 2025. Dispõe sobre a oferta de educação a distância por instituições de educação superior em cursos de graduação e altera o Decreto nº 9.235, de 15 de dezembro de 2017. Brasília, DF, 2025. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/decreto/d12456.htm. Acesso em: 28 set. 2026.

BRASIL. Ministério da Educação. Portaria MEC nº 506, de 10 de julho de 2025. Regulamenta o Decreto nº 12.456, de 19 de maio de 2025, que trata da oferta de educação a distância por instituições de educação superior em cursos de graduação. Diário Oficial da União: seção 1, Brasília, DF, 14 jul. 2025, p. 34. Disponível em: https://www.in.gov.br/consulta. Acesso em: 28 set. 2026.

BRASIL. Ministério da Educação. Portaria MEC nº 795, de 25 de novembro de 2025. Altera a Portaria nº 381, de 20 de maio de 2025, e a Portaria MEC nº 506, de 10 de julho de 2025. Brasília, DF, 2025. Disponível em: https://www.in.gov.br/consulta. Acesso em: 28 set. 2026.

COTTON, Debby R. E.; COTTON, Peter A.; SHIPWAY, J. Reuben. Chatting and cheating: ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, v. 61, n. 2, p. 228-239, 2024. Disponível em: https://doi.org/10.1080/14703297.2023.2190148. Acesso em: 28 set. 2026.

DAWSON, Phillip. Defending assessment security in a digital world: preventing e-cheating and supporting academic integrity in higher education. London: Routledge, 2021.

FREIRE, Paulo. Pedagogia da autonomia: saberes necessários à prática educativa. São Paulo: Paz e Terra, 1996.

INTERNATIONAL CENTER FOR ACADEMIC INTEGRITY. The fundamental values of academic integrity. 3. ed. [S. l.]: ICAI, 2021. Disponível em: https://www.academicintegrity.org/the-fundamental-values-of-academic-integrity. Acesso em: 28 set. 2026.

LIANG, Weixin; YUKSEKGONUL, Mert; MAO, Yining; WU, Eric; ZOU, James. GPT detectors are biased against non-native English writers. Patterns, v. 4, n. 7, art. 100779, 2023. Disponível em: https://doi.org/10.1016/j.patter.2023.100779. Acesso em: 28 set. 2026.

LITTO, Fredric M.; FORMIGA, Marcos (org.). Educação a distância: o estado da arte. São Paulo: Pearson Education do Brasil, 2009.

MOORE, Michael G. Theory of transactional distance. In: KEEGAN, Desmond (ed.). Theoretical principles of distance education. London: Routledge, 1993. p. 22-38.

PERKINS, Mike; FURZE, Leon; ROE, Jasper; MACVAUGH, Jason. The Artificial Intelligence Assessment Scale (AIAS): a framework for ethical integration of generative AI in educational assessment. Journal of University Teaching and Learning Practice, v. 21, n. 6, 2024. Disponível em: https://doi.org/10.53761/q3azde36. Acesso em: 28 set. 2026.

PERKINS, Mike; ROE, Jasper; FURZE, Leon. Reimagining the Artificial Intelligence Assessment Scale (AIAS): a refined framework for educational assessment. Journal of University Teaching and Learning Practice, v. 22, n. 7, 2025. Disponível em: https://ro.ecu.edu.au/ajte/vol22/iss7/. Acesso em: 28 set. 2026.

SADLER, D. Royce. Formative assessment and the design of instructional systems. Instructional Science, v. 18, n. 2, p. 119-144, 1989. Disponível em: https://doi.org/10.1007/BF00144329. Acesso em: 28 set. 2026.

SCRIVEN, Michael. The methodology of evaluation. In: TYLER, Ralph W.; GAGNÉ, Robert M.; SCRIVEN, Michael. Perspectives of curriculum evaluation. Chicago: Rand McNally, 1967. p. 39-83. (AERA Monograph Series on Curriculum Evaluation, 1).

SOMMERVILLE, Ian. Software engineering. 10. ed. Boston: Pearson, 2016.

UNESCO. Guidance for generative AI in education and research. Paris: UNESCO, 2023. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000386693. Acesso em: 28 set. 2026.

WEBER-WULFF, Debora; ANOHINA-NAUMECA, Alla; BJELOBABA, Sonja; FOLTÝNEK, Tomáš; GUERRERO-DIB, Jean; POPOOLA, Olumide; ŠIGUT, Petr; WADDINGTON, Lorna. Testing of detection tools for AI-generated text. International Journal for Educational Integrity, v. 19, n. 1, art. 26, 2023. Disponível em: https://doi.org/10.1007/s40979-023-00146-z. Acesso em: 28 set. 2026.

APÊNDICE A – GLOSSÁRIO

Acordo de uso (de IA). Regra pactuada por disciplina que especifica, para cada tarefa, se a IA é vedada, permitida (planejamento, revisão linguística) ou exigida como objeto de análise.

Agência humana. Princípio, associado à UNESCO (2023), de que o uso de IA na educação deve preservar a capacidade de decisão e o controle das pessoas.

Alinhamento construtivo. Coerência entre os resultados de aprendizagem pretendidos, as atividades de ensino e as tarefas de avaliação (Biggs, 1996; Biggs; Tang, 2011).

Alucinação. Saída de um modelo generativo que soa plausível, mas contém informação inventada, como uma referência bibliográfica inexistente.

Arguição oral síncrona. Conversa curta, em tempo real, na qual o estudante explica decisões já registradas por escrito.

Assíncrono / síncrono. Atividade realizada em tempos distintos para docente e estudante (assíncrona), ou com interação em tempo real (síncrona).

Avaliação diagnóstica. Mapeia conhecimentos prévios e dificuldades para ajustar o ensino.

Avaliação formativa. Coleta evidências ao longo do processo para modificar o ensino e o estudo (Black; Wiliam, 1998).

Avaliação somativa. Sintetiza o desempenho ao fim de uma unidade para fins de registro, progressão ou certificação.

Confiabilidade. Consistência da observação: avaliadores diferentes, ou o mesmo avaliador em momentos diferentes, chegam a julgamentos compatíveis.

Construto. Capacidade ou atributo que a avaliação pretende medir, por exemplo "autoria intelectual sem auxílio" ou "avaliação crítica de saídas de IA".

Dado pessoal sensível. Categoria da LGPD que inclui dados biométricos e exige proteção reforçada.

Detector de texto gerado por IA. Ferramenta que estima, por regularidades estatísticas, a probabilidade de um texto ter sido produzido por um modelo. Está sujeita a falsos positivos e vieses.

Distância transacional. Separação não só geográfica, mas pedagógica, que resulta do diálogo e da estrutura do curso (Moore, 1993).

Educação a distância (EaD). Modalidade em que a mediação didática ocorre por tecnologias, sem depender de copresença contínua de docente e estudante.

Equidade. Condição em que a estratégia avaliativa não concentra desvantagem sobre quem tem menos conectividade, tempo, acesso a ferramentas ou fluência na variedade linguística valorizada.

Escalabilidade. Capacidade de uma estratégia funcionar em turmas grandes sem custo docente inviável.

Evidência de aprendizagem. Produto ou desempenho observável que sustenta uma inferência sobre o que o estudante aprendeu.

Falso positivo. Texto humano classificado erroneamente como gerado por IA.

Feedback. Informação sobre o desempenho que só cumpre sua função quando o estudante a usa para reduzir a distância em relação ao padrão (Sadler, 1989).

Inteligência artificial generativa. Sistemas capazes de produzir texto, código, imagem e outros conteúdos a partir de instruções do usuário.

Integridade acadêmica. Compromisso com honestidade, confiança, justiça, respeito, responsabilidade e coragem no processo de demonstrar conhecimento (International Center for Academic Integrity, 2021).

Letramento em IA. Conhecimento sobre o que os modelos e os detectores fazem, seus limites e as implicações éticas e legais do uso.

LGPD. Lei nº 13.709/2018, que regula o tratamento de dados pessoais.

Log mínimo. Registro enxuto do uso de IA (*prompts* relevantes, versões aceitas e rejeitadas, critério de aceitação), suficiente para revisão humana e sem excesso de dados pessoais.

Portfólio. Conjunto de entregas parciais, rascunhos e versões que evidencia a trajetória do estudante, e não só o produto final.

***Proctoring*.** Vigilância remota da realização de provas, por imagem, áudio ou tela.

Registro de decisão. Documento em que o estudante declara contexto, alternativas, critério de escolha e perdas aceitas. Vem da prática de projetos de software.

Resistência relativa (à IA). Dificuldade e custo de substituir integralmente o desempenho do estudante por IA. Não significa imunidade.

Revisão bibliográfica narrativa. Revisão sem protocolo sistemático, baseada no julgamento de quem seleciona as fontes.

Risco de exclusão. Probabilidade de uma estratégia avaliativa impor barreiras a estudantes por conectividade, deslocamento, deficiência, jornada de trabalho ou fluência (Quadro 2).

Rubrica. Instrumento com critérios explícitos e níveis de desempenho para julgar um trabalho.

Segurança da avaliação. Condições que dificultam a fraude e protegem a validade das inferências (Dawson, 2021). Distingue-se da integridade, que é ética e cultural.

Taxonomia revisada de Bloom. Modelo com dois eixos: processos cognitivos (lembrar, entender, aplicar, analisar, avaliar, criar) e tipos de conhecimento (factual, conceitual, procedimental, metacognitivo) (Anderson; Krathwohl, 2001).

Uso indevido. Uso de IA não declarado ou contrário às regras comunicadas antes da tarefa, ou fabricação de fontes e dados.

Validade. Justificativa da inferência feita a partir da evidência coletada. Um instrumento é válido para um uso quando o desempenho observado sustenta a conclusão sobre o resultado de aprendizagem pretendido.

Versionamento. Prática de guardar e datar versões sucessivas de um trabalho, com nota sobre o que mudou e por quê.