

“Sorria e diga pH”: uso de Inteligência Artificial generativa para a avaliação de uma sequência didática com enfoque no Ensino Médio

Murilo Nicolás Mombelli^{1*} (IC), Felipe Cristiano Belluci¹ (IC), Guilherme Gonçalves Costa¹ (PG), Wilton J. D. Nascimento Júnior¹ (PG), Gildo Giroto Júnior¹ (PQ)

*murilonmombelli@gmail.com

¹Instituto de Química, Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brasil

Palavras-Chave: Inteligência Artificial, Planejamento didático-pedagógico, Formação de professores.

Resumo: O uso de Inteligência Artificial (IA) com finalidades educacionais tem constituído um importante debate na literatura. Tendo isso em vista, o objetivo deste trabalho foi o de analisar a viabilidade do uso de IA generativa para a avaliação de Sequências Didáticas (SD) com enfoque no Ensino Médio, levantando algumas possibilidades e limitações nesse processo. Na primeira etapa, elaboramos uma SD com temática ácido-base. Em seguida, utilizamos a IA generativa de texto ChatGPT (GPT-3.5) para avaliar cinco aspectos da SD proposta: (1) contexto; (2) objetivos de aprendizagem; (3) adequação teórico-metodológica; (4) estratégias didáticas; e (5) estratégias avaliativas. As respostas geradas foram submetidas à Análise Textual Discursiva (ATD). A análise evidenciou um possível uso da IA para a avaliação da SD, requerendo uma comunicação clara e bem-informada para seu refinamento. O estudo sugere um uso crítico-reflexivo, cuidadoso e com viés formativo das IA para a obtenção de feedback sobre o planejamento didático-pedagógico.

INTRODUÇÃO

A ubiquidade das tecnologias digitais na sociedade contemporânea tem desempenhado um papel que nos impacta a todo momento, e dentro do campo educacional, isso não é diferente (Leite, 2023). As Tecnologias Digitais de Informação e Comunicação (TDIC) aplicadas à educação vêm se destacando de inúmeras formas, dentro e fora de sala de aula. Segundo Leite,

Professores e estudantes utilizam aplicativos, *softwares*, vídeos, jogos digitais, internet etc. Contudo, a inserção dessas tecnologias no ensino, muitas vezes, é (ou foi) questionada por seus envolvidos (professores, diretores, pais, estudantes) (Leite, 2023, p. 915).

Nesse contexto, as Inteligências Artificiais (IA) têm gerado um intenso debate sobre aspectos como o currículo e o papel dos professores, além, é claro, de questões éticas associadas ao seu uso (Duque *et al.*, 2023).

Uma primeira definição de IA foi proposta em 1955 por McCarthy *et al.* (2006), considerando-as processos que imitam nossa cognição, também tomando decisões de modo similar aos seres humanos. Entre as IA desenvolvidas desde então, os *chatbots* conquistaram um espaço considerável no âmbito educacional. Estes são programas de computador que interagem com usuários através de serviços de IA baseados em nuvem, permitindo a simulação de diálogos (Okonkwo; Ade-Ibijola, 2021).

Um *chatbot* que ganhou notável atenção desde seu lançamento em 30 de novembro de 2022 é o *Chat Generative Pre-Trained Transformer* — ChatGPT. Trata-se de um grande modelo de linguagem (LLM, do inglês *Large Language Model*) de IA generativa, ou seja, que é capaz de gerar conteúdo como textos e/ou imagens com base em padrões ou características aprendidas durante o treinamento com extensos

conjuntos de dados existentes (Baidoo-Anu; Owusu Ansah, 2023). O ChatGPT foi desenvolvido pela OpenAI, e está atualmente disponível nas versões gratuita (GPT-3.5) e paga (GPT-4) (OpenAI, 2024).

No que diz respeito ao uso desta ferramenta com finalidades educacionais, ainda persistem questões diversas sendo investigadas na literatura: Até que ponto o conteúdo gerado pelas interações com o ChatGPT é confiável em relação à exatidão dos conteúdos gerados? Há ressalvas para sua utilização no campo educacional? Como é possível sua utilização? Qual é o impacto do uso desse tipo de ferramenta na avaliação e no currículo de Química e Ciências? Como os professores podem se ajustar a essa mudança com a vinda de estudantes que, muito possivelmente, se utilizarão dela?

Voltando-nos especificamente para a faceta do planejamento didático-pedagógico, neste estudo propomos a investigação de alguns dos elementos do ChatGPT que podem ser aproveitados positivamente pelos professores durante o preparo de suas aulas. Pensando nisso, o objetivo deste trabalho foi o de analisar a viabilidade do uso de IA generativa para a avaliação de Sequências Didáticas (SD) com enfoque no Ensino Médio, levantando algumas possibilidades e limitações nesse processo.

METODOLOGIA

O desenho metodológico proposto neste estudo consiste numa série de três etapas. Na *Etapa 1*, desenvolvemos uma SD intitulada “Sorria e diga pH”. Para isso, adotamos a seguinte estrutura: (1) Título e tema da SD; (2) Objetivos previstos em orientações curriculares oficiais; (3) Objetivos específicos; (4) Conteúdo programático; (5) Referencial teórico-metodológico; (6) Recursos necessários; (7) Metodologia; (8) Avaliação; e (9) Bibliografia do/a professor/a.

Em seguida, na *Etapa 2*, submetemos a SD à avaliação pelo ChatGPT (versão GPT-3.5, de acesso gratuito) (OpenAI, 2024). Muito embora a versão GPT-4, de licença paga, possua um rol de funcionalidades interessantes (OpenAI, 2023) e de necessária atenção no âmbito da pesquisa educacional, neste trabalho procuramos ater-nos à versão mais acessível aos professores e estudantes no contexto da Educação Básica brasileira. Para maior organização, a *Etapa 2* foi dividida em duas subetapas. Na *Subetapa 2.1*, elaboramos o *prompt* inicial e solicitamos ao ChatGPT que fizesse uma primeira avaliação da SD na sua íntegra considerando cinco aspectos do planejamento didático-pedagógico da SD, a saber: (1) contexto; (2) objetivos de aprendizagem; (3) adequação teórico-metodológica; (4) estratégias didáticas; e (5) estratégias avaliativas. A resposta gerada nessa interação com o ChatGPT foi denominada “Resposta A”.

Na *Subetapa 2.2*, por sua vez, treinamos o ChatGPT para a identificação e avaliação desses mesmos aspectos, fornecendo definições para eles. Ao final do treinamento, solicitamos uma reavaliação, de modo que a nova resposta gerada nesta segunda interação com o ChatGPT foi identificada como “Resposta B”. Neste momento, utilizamos três diretrizes apontadas por Ekin (2023) como boas práticas na engenharia de *prompts*: (a) teste e refinamento iterativos; (b) equilíbrio entre a intenção do usuário e a criatividade do modelo; e (c) garantia de uso ético, evitando-se vieses.

Por fim, na *Etapa 3*, submetemos as Respostas A e B à Análise Textual Discursiva (ATD) (Moraes; Galiazzi, 2016), a qual realizamos com o auxílio do *software* MAXQDA v. 24.2.0. As Unidades de Significado (US) obtidas foram agrupadas em seis categorias *a posteriori*. O **Quadro 1** mostra os *prompts* das *Subetapas 2.1* e *2.2*.

Quadro 1: Apresentação dos prompts utilizados e respostas geradas pelo GPT-3.5 nas Subetapas 2.1 e 2.2

Subetapa	Prompt	ID da resposta
2.1	A seguir, forneço uma Sequência Didática (SD) para você. Gostaria que você avaliasse minha SD detalhadamente quanto aos seguintes aspectos: (1) contexto; (2) objetivos de aprendizagem; (3) adequação teórico-metodológica; (4) estratégias didáticas; e (5) estratégias avaliativas. Justifique cada apontamento seu sobre a SD. [inserção da SD aqui]	A
2.2	Agora, apresento as definições de cada aspecto que mencionei em nosso diálogo anterior: (1) Contexto: O contexto abrange os níveis local, regional e nacional. Leva em consideração o ano escolar da turma, seu perfil socioeconômico, a localização da escola (estado e cidade, e também se fica em zona urbana, rural, periférica, central etc.) e especificidades diversas, como a infraestrutura escolar e a presença ou não de alunos de inclusão. Também envolve a caracterização de aspectos como os conhecimentos prévios dos estudantes. (2) Objetivos de aprendizagem: São os objetivos que o professor busca atingir com aquela aula ou conjunto de aulas. Estão frequentemente associados à aprendizagem de competências e habilidades específicas a depender das orientações curriculares consideradas (municipais, estaduais e/ou nacionais). (3) Adequação teórico-metodológica: Trata-se da coerência global da SD em termos dos referenciais teóricos e/ou metodológicos selecionados. É importante verificar se há explicitação desses referenciais na SD, e se as estratégias didáticas e avaliativas estão de acordo com as propostas desses referenciais. (4) Estratégias didáticas: São o conjunto de ações didáticas intencionadas e planejadas do professor para a consecução dos objetivos de aprendizagem propostos, ou seja, trata-se do elemento do planejamento responsável pela consecução dos objetivos. (5) Estratégias avaliativas: São o conjunto de ações avaliativas intencionadas e planejadas do professor para a consecução dos objetivos de aprendizagem propostos, ou seja, trata-se do elemento do planejamento responsável pela consecução dos objetivos. As estratégias didáticas e avaliativas não são mutuamente excludentes, ou seja, uma estratégia pode ser didática e avaliativa ao mesmo tempo. De posse dessas definições, proponha uma reavaliação da mesma SD, considerando estes cinco aspectos que mencionei.	B

Fonte: elaboração própria.

Especificamente para a Subetapa 2.2, tomamos por base Alves e Bego (2020) para a fundamentação dos conceitos apresentados à máquina. Por fim, o **Quadro 2** traz uma síntese de todas as etapas propostas.

Quadro 2: Esquema das etapas do desenho metodológico

Etapas / subetapas		Descrição
1		Elaboração da SD “Sorria e diga pH”.
2		Engenharia de <i>prompts</i> para a avaliação da SD.
	2.1	Solicitação de avaliação da SD apresentada na sua íntegra.
	2.2	Treinamento e solicitação de reavaliação pelo ChatGPT considerando os seguintes aspectos: (1) contexto; (2) objetivos de aprendizagem; (3) adequação teórico-metodológica; (4) estratégias didáticas; e (5) estratégias avaliativas.
3		Análise Textual Discursiva (ATD) (Moraes; Galiuzzi, 2016) das respostas geradas pela IA generativa ChatGPT (GPT-3.5).

Fonte: elaboração própria.

Destacamos nosso reconhecimento de que o ChatGPT, assim como outras IA generativas, funciona com base no “diálogo” com a máquina, requerendo a solicitação constante de reformulação, e isso não se dá por meio de um único *prompt*. Neste estudo, porém, optamos por testar a IA com *prompts* iniciais a fim de avaliarmos a perspectiva de um usuário estreante no uso desse tipo de ferramenta.

RESULTADOS E DISCUSSÃO

Inicialmente, apresentamos, por meio do **Quadro 3**, a SD “Sorria e diga pH”.

Quadro 3: Apresentação da Sequência Didática (SD) “Sorria e diga pH”

1. Título e tema da SD: “Sorria e diga pH”: uma aula sobre teoria ácido-base para o Ensino Médio
2. Objetivos previstos em orientações curriculares oficiais Nesta SD, propomos que o trabalho seja alinhado com o modelo da Base Nacional Comum Curricular (BNCC) (Brasil, 2018), voltando-se para a Competência Específica 3 de Ciências da Natureza e suas Tecnologias. As duas habilidades a serem desenvolvidas são: (EM13CNT302) Comunicar, para públicos variados, em diversos contextos, resultados de análises, pesquisas e/ou experimentos, elaborando e/ou interpretando textos, gráficos, tabelas, símbolos, códigos, sistemas de classificação e equações, por meio de diferentes linguagens, mídias, Tecnologias Digitais de Informação e Comunicação (TDIC), de modo a participar e/ou promover debates em torno de temas científicos e/ou tecnológicos de relevância sociocultural e ambiental (Brasil, 2018, p. 559). (EM13CNT306) Avaliar os riscos envolvidos em atividades cotidianas, aplicando conhecimentos das Ciências da Natureza, para justificar o uso de equipamentos e recursos, bem como comportamentos de segurança, visando à integridade física, individual e coletiva, e socioambiental, podendo fazer uso de dispositivos e aplicativos digitais que viabilizem a estruturação de simulações de tais riscos (Brasil, 2018, p. 559).

Quadro 3 (continuação): Apresentação da Sequência Didática (SD) “Sorria e diga pH”

3. Objetivos específicos

A partir dos aspectos científicos dos saberes, como a teoria de Arrhenius e a definição de pH, pretende-se que o estudante interprete a linguagem química, sendo capaz de explicar propriedades químicas através das fórmulas químicas apresentadas. Dessa forma, objetiva-se discutir aspectos ambientais e sociais de problemáticas cotidianas que envolvam esse conteúdo. Para tal, sugerimos o uso de Peruzzo e Canto (2006) como fonte de preparo de aula por parte do professor.

4. Conteúdo programático

Ácidos e bases de Arrhenius e suas propriedades químicas, nomenclatura dos principais compostos, definição de pH e sua presença no cotidiano.

5. Referencial teórico-metodológico

Como referencial teórico-metodológico, serão utilizados os Três Momentos Pedagógicos, que se baseiam na possibilidade de estabelecer uma maior dinâmica de atuação entre o docente e os estudantes (Delizoicov; Angotti; Pernambuco, 2018). Estes são caracterizados por:

1) Problematização inicial: consiste em apresentar situações reais dentro do tema que será trabalhado, conectando o conhecimento científico à sua realidade, de forma a despertar sua curiosidade e motivação para a aprendizagem. Inicialmente, estes problemas são discutidos em grupos pequenos e, posteriormente, são explorados por toda a sala de aula.

2) Organização do conhecimento: o conteúdo científico selecionado para a compreensão do primeiro momento é estudado, com intervenção docente. É nesta etapa que a função formativa do livro didático e exercícios é desempenhada.

3) Aplicação do conhecimento: destina-se a abordar sistematicamente o conhecimento incorporado pelo aluno, para analisar e interpretar tanto as situações iniciais que determinaram seu estudo quanto outras que, embora não estejam diretamente ligadas ao momento inicial, possam ser compreendidas pelo mesmo conhecimento.

6. Recursos necessários

Lousa, giz e/ou pincéis, projetor, *notebook*, cabo HDMI, material em *slides* do professor, acesso à *internet*, celulares dos estudantes. Para o experimento investigativo, serão utilizadas estantes com tubos de ensaio, copos descartáveis, água de torneira, pistilo ou pilão pequeno, e feijão preto cozido.

7. Metodologia**1) Problematização Inicial:**

Organização de grupos de 3 alunos e leitura da notícia “Erosão ácida: uma ameaça à saúde bucal” (Bianco, 2018), formulando sugestões para responder a questão geradora “Como entendemos o conceito de ácidos e bases cotidianamente?”. Neste momento, é importante que o professor se antecipe quanto às principais respostas possivelmente sugeridas (1 aula presencial de 45 minutos).

2) Organização do Conhecimento:

A partir de uma aula expositiva dialogada (2 aulas presenciais de 45 minutos), propõe-se a conceituação do conteúdo de ácidos e bases de Arrhenius, destacando as suas diferenças, propriedades químicas, nomenclatura e definição de pH.

Quadro 3 (continuação): Apresentação da Sequência Didática (SD) “Sorria e diga pH”

3) Aplicação do Conhecimento: - Realização de um experimento investigativo (Bruck; Bretz; Towns, 2008) buscando-se analisar a mudança de cor de uma solução de extrato de feijão preto cozido macerado com a adição de produtos ácidos ou básicos do cotidiano (Pereira; Viturino; Assis, 2017). Segundo os autores, ele apresenta coloração vermelha clara ($0 \leq \text{pH} \leq 3$), passando pela incolor ($4 \leq \text{pH} \leq 7$), verde clara ($8 \leq \text{pH} \leq 9$), amarela ($10 \leq \text{pH} \leq 12$), e amarela intensa ($13 \leq \text{pH} \leq 14$) (<u>1 aula presencial de 45 minutos</u>). - Discussão das sugestões da questão geradora e esquematização das diferenças entre a percepção cotidiana e o conteúdo científico (<u>1 aula presencial de 45 minutos</u>).
8. Avaliação A. Proposta somativa. Confecção de uma tabela contendo alguma solução utilizada no cotidiano, o pH da substância utilizada, a coloração obtida e a classificação de sua acidez. B. Proposta somativa / formativa. Produção de um texto no formato de <i>post</i> de rede social contendo orientações em relação ao consumo de alimentos ácidos a fim de se evitar o desgaste dos dentes.
9. Bibliografia do/a professor/a BIANCO Oral Care. Erosão ácida: uma ameaça à saúde bucal, G1, Triângulo Mineiro-MG, 30 jul. 2018. Disponível em: https://g1.globo.com/mg/triangulo-mineiro/especial-publicitario/bianco-oral-care/noticia/2018/07/30/erosao-acida-uma-ameaca-a-saude-bucal.shtml. Acesso em: 7 mar. 2024. BRASIL. Ministério da Educação. Base Nacional Comum Curricular: Educação é a Base. Brasília, DF: MEC/CONSED/UNDIME, 2018. Disponível em: http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518-versaofinal_site.pdf. Acesso em: 6 mar. 2024. BUCK, L. B.; BRETZ, S. L.; TOWNS, M. H. Characterizing the level of inquiry in the undergraduate laboratory. Journal of College Science Teaching, v. 38, n. 1, p. 52–58, 2008. Disponível em: https://www.chem.purdue.edu/towns/Towns%20Publications/Bruck%20Bretz%20Towns%202008.pdf. Acesso em: 7 mar. 2024 DELIZOICOV, D.; ANGOTTI, J. A.; PERNAMBUCO, M. M. Ensino de Ciências: fundamentos e métodos. 5. ed. Cortez: São Paulo, 2018. PEREIRA, A. de S.; VITURINO, J. P.; ASSIS, A. O uso de indicadores naturais para abordar a experimentação investigativa problematizadora em aulas de Química. Educação Química em Ponto de Vista, v. 1, n. 2, 2017. Disponível em: https://revistas.unila.edu.br/eqpv/article/view/891. Acesso em: 6 mar. 2024. PERUZZO. F. M.; CANTO. E. L. Química na abordagem do cotidiano. v. 1, 4. ed. São Paulo: Editora Moderna, 2006.

Fonte: elaboração própria.

A partir da SD apresentada e utilizando-nos do primeiro *prompt* fornecido no **Quadro 1**, obtivemos a Resposta A. De posse dela, inserimos o segundo *prompt* do quadro e obtivemos a Resposta B. Ambas foram utilizadas como *corpus* de análise da ATD, passando pelas etapas de unitarização, categorização e comunicação (Moraes; Galiazzi, 2016). As etapas de unitarização e categorização geraram um total de 60 unidades, das quais 13 foram Unidades Não Categorizadas (NC). Como mencionamos previamente, as outras 47 US foram agrupadas em seis categorias *a posteriori*. As unidades consideradas NC foram os próprios *prompts* inseridos e os subtítulos de cada

seção das respostas geradas. O **Quadro 4** mostra os números de US obtidas na análise juntamente da distribuição de percentuais relativos para cada categoria, desconsiderando-se as unidades NC no cálculo percentual.

Quadro 4: Categorias, números (N) e percentuais relativos (P) de Unidades de Significado (US) obtidas a partir da Análise Textual Discursiva (ATD) das Respostas A e B

Categorias	Resposta A		Resposta B	
	Nº US (N)	P (%)	Nº US (N)	P (%)
Afirmações vagas (C1)	11	42,3	3	14,3
Críticas pertinentes (C2)	0	0	12	57,1
Críticas não pertinentes (C3)	0	0	4	19,0
Descrição acrítica (C4)	12	46,2	0	0
Linguagem taxativa (C5)	3	11,5	1	4,8
Persistência na resposta após iteração (C6)	0	0	1	4,8
Não categorizado (NC)	6	–	7	–
Total (excluindo-se as unidades de NC)	26	100	21	100

Fonte: elaboração própria.

Com o auxílio do *software* MAXQDA v. 24.2.0, obtivemos os retratos de documento das Respostas A e B, contendo 30 x 30 círculos cada. Um retrato de documento é uma forma de visualização das unidades atribuídas na codificação. Nele, é possível verificar a distribuição das US ao longo dos documentos sob análise, de modo que cada cor representa uma categoria, e cada círculo ou sequência de círculos de uma determinada cor representam uma US em tamanho proporcional àquele do trecho codificado. Isso posto, a **Figura 1** mostra os retratos de cada uma das respostas.

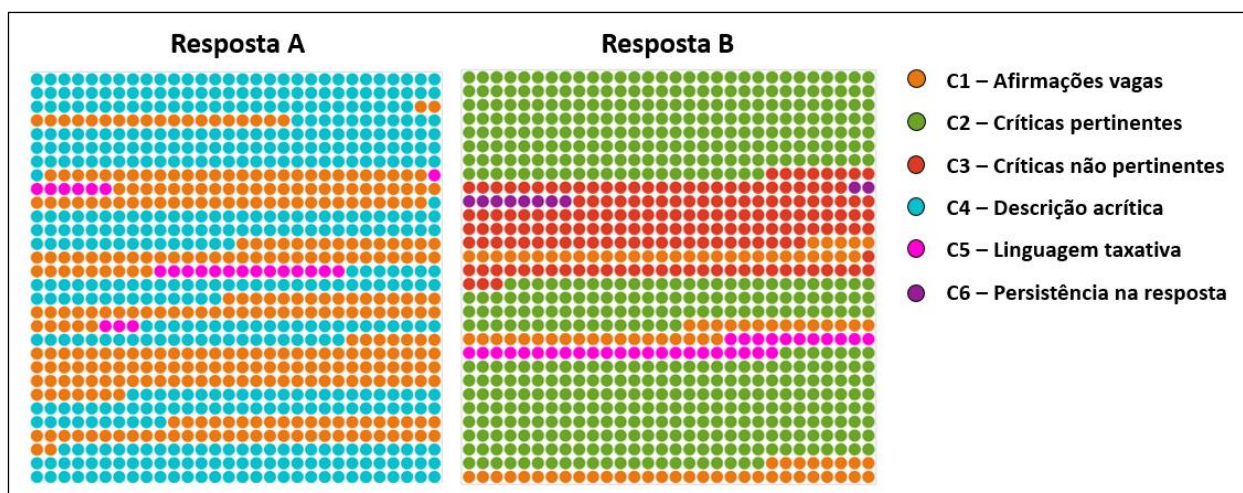


Figura 1: Retratos de documento, contendo 30 x 30 círculos, das Respostas A e B

Conforme apontamos na seção de metodologia, elegemos cinco aspectos do planejamento didático-pedagógico que consideramos relevantes para a boa qualidade de uma SD: (1) contexto; (2) objetivos de aprendizagem; (3) adequação teórico-

metodológica; (4) estratégias didáticas; e (5) estratégias avaliativas. Ao submetermos a SD à avaliação do GPT-3.5, notamos que 46,2% das respostas do sistema se enquadraram na categoria “Descrição acrítica” (C4). Essa categoria incluiu respostas que parafrasearam informações da SD sem oferecer uma avaliação crítica. Na Resposta A, por exemplo, a US em que se lê “Os objetivos de aprendizagem estão alinhados com as orientações curriculares oficiais, particularmente as competências da BNCC” limita-se a reiterar o conteúdo já presente na SD. Seria mais adequado se o GPT-3.5 incluísse informações adicionais sobre a BNCC, uma vez que a base de dados fornecida para seu treinamento data de abril de 2023 (OpenAI, 2024). Um fator que muito possivelmente obstaculiza o acesso do GPT-3.5 à BNCC em sua íntegra é o fato de que esta se encontra em um documento PDF disponível para *download* no *website* do Ministério da Educação (MEC) (Brasil, 2018). Como muito possivelmente a plataforma do GPT-3.5 não tem acesso a esses arquivos, ela recorre à reprodução de trechos da BNCC que foram disponibilizados para seu treinamento prévio.

“Afirmações vagas” (C1) e termos que podem ser associados a uma “Linguagem taxativa” (C5) foram outros pontos que nos chamaram a atenção no processo analítico da ATD. O texto gerado na Resposta A tende ao uso de palavras que não dão margem para a relativização ou subjetividades que são tão presentes na dinâmica escolar. Isso ocorre, por exemplo, ao afirmar que determinada sequência de objetivos promove a aprendizagem, como na US em que se lê “Portanto, os objetivos propostos são adequados e fornecem uma direção clara para o ensino e aprendizagem”. Levando-se em consideração o contexto educacional e as teorias de aprendizagem contemporâneas comumente aceitas pela comunidade acadêmica, afirmar categoricamente que os objetivos promoverão a aprendizagem é, no mínimo, equivocado. A aprendizagem é algo que ocorre na estrutura cognitiva do estudante, e o desejável é que o professor fomente e faça a mediação desse processo (Moreira, 2006), o que de forma alguma implica garantia de aprendizagem.

Em uma análise mais abrangente, a primeira tentativa do GPT-3.5 em elicitar uma crítica detalhada da SD não atingiu um alto nível de profundidade. Isso se deu conforme esperado, pois a literatura tem apresentado diversos exemplos de equívocos cometidos pela plataforma (Baidoo-Anu; Owusu Ansah, 2023; Emenike; Emenike, 2023; Leite, 2023). A resposta fornecida consistiu predominantemente em comentários que não ultrapassaram a esfera descritiva, limitando-se a reiterar o conteúdo já existente na SD sem introduzir uma avaliação crítica substancial. Uma crítica efetiva deveria envolver a avaliação da relevância dos objetivos de aprendizagem, da pertinência das estratégias didáticas e avaliativas em relação ao público-alvo, ou mesmo da verificação se houve a consideração do contexto.

Com relação à Resposta B, vale ressaltar que a utilização da IA favorece um processo iterativo com o sistema, isto é, um processo que se constitui em repetições sucessivas de diálogos com a máquina (neste caso, interações via *chat*). Trata-se de um método particularmente desejável, uma vez que a máquina aprende com as orientações do usuário, o que permite, em última instância, treinar cada vez mais o algoritmo para reconhecer aspectos relevantes solicitados nos *prompts* inseridos anteriormente. Para favorecer um direcionamento sistemático de uma análise mais crítica por parte da IA, especificamos elementos relevantes do planejamento didático-pedagógico que requeriam uma avaliação mais aprofundada, detalhando cada um dos tópicos inicialmente apresentados no primeiro *prompt*. A partir deles (contexto, objetivos de aprendizagem, adequação teórico-metodológica, estratégias didáticas e estratégias avaliativas), cada comentário fornecido sobre a SD deveria ser justificado, de forma a

alinhar a crítica da IA com a intencionalidade pedagógica e necessidades específicas do professor usuário da plataforma.

Analisando-se o retrato do documento da Resposta B (**Figura 1**) e considerando-se o acesso a ele por meio da ATD, houve um aumento expressivo na qualidade geral das críticas produzidas pelo GPT-3.5, o que consideramos reflexo direto da engenharia do segundo *prompt*, em que acrescentamos informações instrucionais importantes para a máquina. Neste caso, o GPT-3.5 passou a reconhecer o que cada tópico significava para o usuário, aumentando a probabilidade de respostas mais complexas e pertinentes, e também menos descritivas.

No que se refere às categorias contempladas pela Resposta B, obtivemos 57,1% das respostas do GPT-3.5 incluídas na categoria “Críticas pertinentes” (C2), destacada em verde na **Figura 1**. A título de exemplo, a primeira série de críticas aborda a ausência de contextualização na sequência, como aponta a US em que se lê “A SD não detalha aspectos relevantes como o perfil socioeconômico dos estudantes, a infraestrutura disponível na escola, ou a inclusão de alunos com necessidades especiais”. Este apontamento é de considerável importância, visto que, no caso de um estudante em formação no campo da licenciatura, tais elementos contextuais são cruciais e poderiam modificar significativamente as escolhas pedagógicas adotadas na SD, influenciando a abordagem metodológica escolhida e as respectivas estratégias e materiais didáticos utilizados (Bego; Ferrarini; Morales, 2021).

Em se tratando do aspecto de adequação teórico-metodológica, a resposta do GPT-3.5 criticou a falta de detalhamento presente na sequência, como aponta a seguinte US pertencente à categoria “Críticas não pertinentes” (C3): “Embora a SD mencione os Três Momentos Pedagógicos como referencial teórico-metodológico, faltam explicações mais detalhadas sobre como esses momentos serão implementados na prática”. Não consideramos tal crítica pertinente porque este detalhamento foi fornecido no primeiro *prompt* em que a SD fora apresentada. Quanto às estratégias didáticas, o GPT-3.5 apresentou uma crítica válida na US “No entanto, seria útil fornecer mais detalhes sobre como cada atividade será conduzida, incluindo a distribuição de tempo, o papel do professor e dos alunos, e os recursos necessários”, incluída na categoria “Críticas pertinentes” (C2). Esta observação decorre da falta de contextualização que mencionamos previamente.

Notamos também que as expressões de natureza taxativa — incluídas na categoria “Linguagem taxativa” (C5) — as quais não permitem flexibilidade interpretativa, permaneceram após a iteração, embora em menor número. Houve uma queda de 3 US na Resposta A para 1 US na Resposta B, isto é, de 11,5% para 4,1% das US totais em cada caso. Por exemplo, no segmento que se refere às estratégias didáticas, temos a seguinte US: “Isso garantiria uma implementação mais eficaz das estratégias e uma melhor compreensão por parte dos envolvidos”. A utilização sem referenciais comparativos de termos absolutos, taxativos ou que atribuem valoração, como é o caso das expressões “mais eficaz” e “uma melhor compreensão”, pode ser relativamente problemático em termos de escrita acadêmica. E, novamente, alegar que uma estratégia didática específica irá garantir um entendimento mais aprofundado por parte dos estudantes é inadequado, pois tal garantia não pode ser assumida aprioristicamente.

Cabe ressaltarmos uma grande potencialidade formativa das ferramentas de IA. No caso de um licenciando em Química (e, por que não, em outras disciplinas escolares), argumentamos que, quando os *prompts* são construídos por meio de um detalhamento segundo o conhecimento do usuário e fazendo considerações relevantes para a máquina, este se torna, sobretudo, um momento de oportunidade excepcional para sua

formação inicial. Isso ocorre muito em conta da necessidade de escolhas conscientes por parte do licenciando a respeito do que o *chatbot* deve ou não considerar para a elaboração das próximas interações dialógicas. Esse tipo de conhecimento docente é altamente complexo, pois envolve a mobilização diferencial de componentes de conhecimentos diversos relacionados ao conteúdo, às estratégias didáticas e avaliativas, ao currículo, ao contexto etc., integrando-se, ainda, aos conhecimentos tecnológicos dos licenciandos (Mishra; Warr; Islam, 2023). Para que essa oportunidade seja mais bem aproveitada, é muito importante que os professores formadores das licenciaturas façam a mediação adequada com os licenciandos, intervindo conforme necessário e fazendo provocações para com o conteúdo do *feedback* obtido, seja este mais recheado de equívocos ou não.

Nessa esteira, a formação do formador é um tema de destaque, pois envolve uma formação de professores específica para as TDIC. Segundo Duque *et al.* (2023), há uma série de barreiras que podem dificultar essa formação. De pronto, os autores apontam a resistência à mudança por parte dos educadores, que estão familiarizados com métodos mais tradicionais, por exemplo. Além disso, os autores complementam que a falta de recursos adequados e a não equidade de acesso às tecnologias digitais também constituem problemas graves na realidade educacional brasileira. Para que possamos debater e, eventualmente, superar algumas dessas questões, ações em larga escala são basilares. É importante que a legislação brasileira leve em consideração a ruptura de paradigmas que as IA vêm proporcionando, incluindo-as no debate político-educacional. Somente através disso será possível a promoção consistente de ações em caráter formativo e continuado para os professores em atuação, por exemplo. Em outra instância, é também necessário que as Instituições de Ensino Superior (IES) revejam seus currículos das licenciaturas a fim de considerar uma maior inclusão ou redistribuição de cargas horárias em relação às TDIC, acrescentando o tópico de IA no campo educacional e trazendo discussões pertinentes quanto ao seu funcionamento e à ética em seu uso.

Para finalizar a seção, voltamos-nos ao objetivo de nosso trabalho e elencamos, por meio do **Quadro 5**, algumas possibilidades e limitações que pudemos notar para o uso de IA generativa na obtenção de *feedback* para Sequências Didáticas (SD).

Quadro 5: Possibilidades e limitações no uso de Inteligência Artificial (AI) generativa para a avaliação de Sequências Didáticas (SD)

Possibilidades	Limitações
Produzem <i>feedbacks</i> relativamente coerentes, desde que treinadas com <i>prompts</i> adequados	Requerem competências digitais dos professores para a obtenção de resultados assertivos
Permitem que professores de Química e/ou Ciências problematizem aspectos do planejamento, tanto a partir de críticas feitas de forma acertada pela IA como a partir daquelas equivocadas	Há risco de obtenção de <i>feedbacks</i> altamente generalistas, no caso de <i>prompts</i> mal formulados ou que não passaram por adequada engenharia de <i>prompts</i>
Quando adequadamente mediado, o uso de IA no <i>feedback</i> de SD pode se constituir em excelentes oportunidades de formação inicial e continuada	Quando o <i>feedback</i> por meio da IA é utilizado acriticamente, os professores podem considerar que determinadas SD não são passíveis de aprimoramento quando o são (e vice-versa)

Fonte: elaboração própria.

Ainda que nosso enfoque tenha sido o Ensino Médio e a temática ácido-base, acreditamos que os itens listados no **Quadro 5** possam ser extrapolados para sequências de diferentes níveis escolares nas disciplinas de Química e/ou Ciências, aspecto que acreditamos ser importante para investigações futuras.

CONSIDERAÇÕES FINAIS

Em suma, consideramos a interação com a IA através do *chatbot* GPT-3.5 uma forma interessante e relativamente inédita para possíveis aprimoramentos na análise crítica dos elementos didático-pedagógicos presentes nas SD, desde que esta não se restrinja unicamente à máquina. A melhoria expressiva na qualidade das críticas oferecidas pelo *chatbot*, impulsionada pela clareza e detalhamento dos *prompts* fornecidos pelo usuário, sublinha a importância de uma comunicação efetiva e bem-informada para o refinamento da ferramenta. Entretanto, a presença de afirmações taxativas em respostas posteriores à nossa interação também evidencia a necessidade de constante vigilância quanto à qualidade do *feedback* gerado pela IA, procurando sempre alinhá-la às expectativas acadêmicas e à realidade educacional por meio de novas e sucessivas interações dialógicas.

O uso das ferramentas de IA em outros tópicos da Química para além da temática ácido-base também é desejável, pois isso nos permitirá o acesso a cada vez mais dados empíricos sobre o tema, estabelecendo as bases fundamentais para que desenvolvamos em nossos estudantes boas práticas escolares e acadêmicas a respeito do uso desse tipo de plataforma. Ademais, também é de grande valia que novos estudos sejam realizados utilizando-se a Língua Portuguesa, uma vez que a maior parte das pesquisas a respeito do ChatGPT no Ensino de Química se encontra em inglês.

Em tempo, ressaltamos que nosso estudo não advoga pelo uso indiscriminado, impensado ou com perspectiva milagrosa das IA no campo educacional, e sim por um uso crítico-reflexivo, cuidadoso e com viés formativo. Acreditamos que os equívocos cometidos pelas plataformas são oportunidades de debate frutífero tanto para os estudantes em formação inicial como para professores com ampla experiência de sala de aula. Para nós, enquanto professores, este desenvolvimento profissional é de grande importância, pois em um cenário futuro de idealidade e com acesso pleno às TDIC, é muito provável que recebamos turmas de estudantes que também se utilizarão das ferramentas de IA cotidianamente.

Por fim, discutir o papel das IA e seus impactos na educação permite-nos refletir aberta e criticamente sobre aspectos de relevância ímpar para o desenvolvimento de habilidades como a ética e a autocriticidade em nossos estudantes, tão necessárias aos cidadãos do século XXI.

REFERÊNCIAS

ALVES, Milena; BEGO, Amadeu Moura. A celeuma em torno da temática do planejamento didático-pedagógico: definição e caracterização de seus elementos constituintes, **Revista Brasileira de Pesquisa em Educação em Ciências**, v. 20, p. 71–96, 2020. Disponível em: <https://doi.org/10.28976/1984-2686rbpec2020u7196>. Acesso em: 29 mar. 2024.

BAÏDOO-ANU, David; OWUSU ANSAH, Leticia. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. **Journal of AI**, v. 7, n. 1, p. 52–62, 2023. Disponível em: <https://doi.org/10.61969/jai.1337500>. Acesso em: 7 mar. 2024.

BEGO, Amadeu Moura; FERRARINI, Francisco Otávio Cintra; MORALLES, Vagner Antonio. Ressignificação dos estágios curriculares supervisionados por meio da implementação de Unidades Didáticas Multiestratégicas. **Educação Química em Ponto de Vista**, v. 5, n. 1, p. 5–28, 2021. Disponível em: <https://doi.org/10.30705/eqpv.v5i1.2530>. Acesso em: 29 mar. 2024.

BRASIL. Ministério da Educação. **Base Nacional Comum Curricular**: Educação é a Base. Brasília, DF: MEC/CONSED/UNDIME, 2018. Disponível em: http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518-versaofinal_site.pdf. Acesso em: 6 mar. 2024.

DUQUE, Rita de Cássia Soares; MONTEIRO, Rhadson Rezende; DE OLIVEIRA FILHO, Fernando Luiz Cas; LOUREIRO, Valéria Jane Siqueira; DO NASCIMENTO, Isidro José Bezerra Maciel Fortaleza; PLACIDO, Reginaldo Leandro; DA SILVA, Cícero José; DA SILVA, Joel Manga; SARAIVA, Maria do Socorro Gomes; SILVA, Aline Maria de Brito. Formação de professores para o uso de tecnologia: a inteligência artificial (IA) e os novos desafios da educação. **Caderno Pedagógico**, v. 20, n. 2, p. 838–852, 2023. Disponível em: <https://ojs.studiespublicacoes.com.br/ojs/index.php/cadped/article/view/1607>. Acesso em: 7 mar. 2024.

EKIN, Sabit. Prompt engineering for ChatGPT: a quick guide to techniques, tips, and best practices. **Authorea Preprints**. 2023. Disponível em: <https://www.techrxiv.org/doi/pdf/10.36227/techrxiv.22683919.v2>. Acesso em: 7 mar. 2024.

EMENIKE, Mary E.; EMENIKE, Bright U. **Journal of Chemical Education**, v. 100, n. 4, p. 1413–1418, 2023. Disponível em: <https://pubs.acs.org/doi/10.1021/acs.jchemed.3c00063>. Acesso em: 7 mar. 2024.

LEITE, Bruno Silva. Inteligência Artificial e Ensino de Química: uma análise propedêutica do ChatGPT na definição de conceitos químicos. **Química Nova**, v. 46, n. 9, p. 915–923, 2023. Disponível em: <http://dx.doi.org/10.21577/0100-4042.20230059>. Acesso em: 7 mar. 2024.

MCCARTHY, John; MINSKY, Marvin L.; ROCHESTER, Nathaniel ; SHANNON, Claude E. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. **AI Magazine**, v. 27, n. 4, p. 12, 2006. Disponível em: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1904>. Acesso em: 7 mar. 2024.

MISHRA, Punya; WARR, Melissa; ISLAM, Rezwana. TPACK in the Age of ChatGPT and Generative AI. **Journal of Digital Learning in Teacher Education**, v. 39, n. 4, p. 235-251, 2023. Disponível em: <https://doi.org/10.1080/21532974.2023.2247480>. Acesso em: 29 mar. 2024.

MORAES, Roque; GALIAZZI, Maria do Carmo. **Análise Textual Discursiva**. 3. ed. rev. e aum. Ijuí: Editora Unijuí, 2016. 264 p.

MOREIRA, Marco Antônio. **Aprendizagem Significativa: da visão clássica à visão crítica**. Conferência de encerramento do V Encontro Internacional sobre Aprendizagem Significativa, Madrid, Espanha, setembro de 2006 e do I Encuentro Nacional sobre Enseñanza de la Matemática, Tandil, Argentina, abril de 2007. Disponível em: www.if.ufrgs.br/~moreira/visaoclasicavisaocritica.pdf. Acesso em: 29 mar. 2024.

OKONKWO, Chinedu Wilfred; ADE-IBIJOLA, Abejide. Chatbots applications in Education: a systematic review. **Computers & Education: Artificial Intelligence**, v. 2, p. 1–10, 2021. Disponível em: <https://doi.org/10.1016/j.caeai.2021.100033>. Acesso em: 7 mar. 2024.

OPENAI. **GPT-4 Technical Report**. 2023. Disponível em: <https://arxiv.org/abs/2303.08774>. Acesso em: 7 mar. 2024.

OPENAI. **ChatGPT**. 2024. Disponível em: <https://openai.com/chatgpt>. Acesso em: 7 mar. 2024.