

Coautores sintéticos: avaliação conceitual de planos de aula em biologia gerados por IAGs

Synthetic co-authors: conceptual evaluation of biology lesson plans generated by IAGs

Ricardo Maciel¹
Ana Amélia Amorim Carvalho²

Resumo

Este artigo analisa a qualidade conceitual e a adesão aos princípios da Educação OnLIFE em planos de aula de Biologia para o Ensino Médio gerados por dois modelos de linguagem multimodal (ChatGPT-4o e Gemini 2.5 Flash). O estudo parte do reconhecimento de que hiperinteligências já atuam como coautoras sintéticas no planejamento docente, tensionando noções de autoria, curadoria e escolarização. Foi usada uma abordagem mista (quantitativa e qualitativa), para analisar os 52 planos de aula gerados por IAG, que foram avaliados por docentes especialistas a partir de uma rubrica de cinco dimensões: exatidão conceitual, complexidade cognitiva, interdisciplinaridade, alinhamento OnLIFE e coerência com a BNCC. Os resultados revelam desempenho excelente geral ($M = 3,86$), com destaque para a precisão conceitual e conformidade curricular, mas evidenciam lacunas em interdisciplinaridade e acessibilidade. Conclui-se que, embora promissoras, as IAGs requerem mediação crítica para evitar a reprodução de lógicas disciplinocêntricas e tecnocráticas. O artigo propõe que a integração ativa entre docentes e LLMs, por meio de práticas de co-design, seja prioridade na formação e em futuras pesquisas empíricas em sala de aula.

Palavras-chave: Educação OnLIFE; Inteligência Artificial Generativa; Coautoria sintética; Ensino de Biologia; Planejamento docente.

Abstract

This article analyzes the conceptual quality and alignment to the principles of OnLIFE Education in high school biology lesson plans generated by two multimodal language models (ChatGPT-4o and Gemini 2.5 Flash). The study begins by recognizing that hyperintelligences are already operating as synthetic co-authors in instructional planning, challenging traditional notions of authorship, curation, and schooling. A mixed-methods approach (quantitative and qualitative) was used to analyze 52 AI-generated lesson plans, which were assessed by expert biology teachers using a five-dimension rubric: conceptual accuracy, cognitive complexity, interdisciplinarity, OnLIFE alignment, and compliance with the BNCC (Brazilian National Common Core Curriculum). Results reveal an overall excellent performance ($M = 3.86$), with high scores in conceptual accuracy and curricular compliance, but also point to persistent gaps in interdisciplinarity and accessibility. The study concludes that, although promising, generative AIs require critical mediation to avoid reproducing disciplinocentric and technocratic logics.

¹ Doutorando em Ciências da Educação na Universidade de Coimbra (UC/Portugal). Mestre em Psicobiologia pela Universidade Federal do Rio Grande do Norte (UFRN). Professor do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN). Coordena o grupo de pesquisa em Fisiologia da Educação (GEFE/CNPQ). E-mail: prof.ricardomacie@gmail.com

² Professora da Faculdade de Ciências da Psicologia e da Educação na Universidade de Coimbra (UC/Portugal). Coordenadora do Laboratório de Tecnologia Educacional na Universidade de Coimbra. E-mail: anaameliac@fpce.uc.pt

The article argues that active integration between teachers and LLMs, through co-design practices, should be a priority in teacher education and future empirical classroom research.

Keywords: OnLIFE education; Generative Artificial Intelligence; Synthetic co-authorship; Biology teaching; Instructional planning.

1. Introdução

A Educação OnLIFE insere-se num paradigma pedagógico em que os processos de ensino-aprendizagem ocorrem em um continuum entre espaços físicos e digitais, mediados por inteligências híbridas e conectivas (Almeida, de, 2009; Possa e Felice, Di, 2024; Schlemmer, Felice e Serra, 2020). Nesse continuum, as chamadas hiperinteligências — algoritmos adaptativos capazes de aprender com múltiplas fontes, dialogar multimodalmente e ajustar suas respostas conforme o contexto e a intenção do usuário — inserem-se como novos coagentes do processo pedagógico (Castellani *et al.*, 2024). Esse cenário técnico-metodológico dialoga com abordagens de formação docente que integram metodologias ativas e ferramentas digitais para fomentar práticas pedagógicas metacognitivas (Almeida, de, Fátima Grein, de e Bedin, 2024). Essa leitura revigora a cibernetica ecológica de Bateson (1972), para quem aprendizagem implica acoplamento de sistemas vivos e não-vivos — um processo de acoplamento estrutural, no qual os sujeitos e os artefatos digitais influenciam mutuamente suas estruturas internas a partir de interações recorrentes, e ressoa o ideário de Illich (1971) sobre a desescolarização *da sociedade*, que denunciava a centralidade excessiva da instituição escolar na legitimação do saber.

Se, por um lado, as hiperinteligências prometem ampliar o alcance cognitivo de professores e estudantes, por outro desafiam noções consolidadas de autoria, mediação e curadoria (Lopes, 2021). A tensão entre possibilidades emancipatórias e riscos tecno utópicos torna-se ainda mais aguda quando algoritmos passam a redigir planos de aula completos, potencialmente deslocando o docente do papel de designer para o de revisor crítico. O presente estudo posiciona-se nesse cruzamento, investigando como tais artefatos performam frente a requisitos curriculares e epistemológicos específicos da Biologia.

1.1 IA generativa nas práticas docentes: panorama e lacuna

Modelos generativos tornaram-se ubíquos no planejamento (objetivos, atividades, avaliação), sobretudo em versões multimodais (Bessas *et al.*, 2025; Gitahy

e Oliveira, 2024). Falta, porém, evidência sistemática sobre a qualidade conceitual e o alinhamento OnLIFE/BNCC desses planos — especialmente quanto a extrapolar a sala física e evitar narrativas disciplinocêntricas. Este estudo aborda essa lacuna.

1.2 Objetivo e questões de pesquisa

Este artigo tem como objetivo geral: avaliar comparativamente a qualidade conceitual e a adesão OnLIFE de planos de aula de Biologia para o ensino médio gerados pelos modelos ChatGPT-4o e Gemini 2.5 Flash.

Para atingir esse objetivo, investigamos diferenças na exatidão conceitual e complexidade cognitiva, bem como a articulação OnLIFE/interdisciplinar e a coerência BNCC.

2. Referencial teórico

A noção de Educação OnLIFE emerge de reflexões pós-digitalistas que reconhecem a fusão dos domínios on-line e off-line em ecossistemas de aprendizagem distribuída (POSSA; FELICE, DI, 2024; SCHLEMMER; FELICE; SERRA, 2020). Como horizonte teórico, duas matrizes críticas ajudam a iluminar essa ecologia híbrida: Gregor Bateson e Ivan Illich.

Para Bateson (1972), aprender é tecer padrões entre sistemas; mudanças de segunda ordem — transformações nas regras que governam o próprio aprender — são cruciais. LLMs podem catalisar metacognição, mas exigem desenho instrucional que reconheça limites e evite *double binds* pela delegação acrítica.

Ivan Illich (1971) questiona a escola como instituição monopolizadora do saber e propõe redes de aprendizagem conviviais, nas quais qualquer pessoa possa conectar-se diretamente a recursos e colegas, sem mediação hierárquica. A emergência de hiperinteligências abertas, como os LLMs, materializa parcialmente essa learning web ao oferecer acesso sob demanda a conteúdos personalizados. Contudo, Illich alerta que a mera introdução de tecnologia pode reforçar estruturas de dependência e reinstalar uma “escolarização algorítmica” — reprodução de currículos e certificações normativos — se não promover autonomia e interação horizontal, transferindo efetivamente o locus de controle para os aprendizes.

A mediação técnica, embora amplamente valorizada como ferramenta de facilitação no processo de ensino-aprendizagem, pode tornar-se um entrave quando

desconsidera o protagonismo do educando. Paulo Freire (1996) alertava para os riscos de uma educação bancária, na qual o conhecimento é depositado passivamente no estudante. O uso acrítico da tecnologia pode reproduzir essa lógica, afastando o protagonismo e o diálogo no processo educativo.

Articulando Bateson e Illich, depreende-se que as hiperinteligências expandem a ecologia do aprender, mas sua qualidade epistêmica e função emancipatória dependem de como docentes e discentes negociam sentido, agência e autoridade. A investigação aqui proposta dialoga com esse referencial ao analisar se planos de aula gerados por IAGs favorecem padrões de pensamento de ordem superior (Bateson, 1972) e práticas pedagógicas desescolarizantes (Illich, 1971), alinhando-se ao eixo do dossiê que enfoca inteligências híbridas e desescolarização.

2.1 Coautoria homem-IA: avanços no planejamento automatizado

A popularização de modelos de linguagem em larga escala — como ChatGPT-4o e Gemini 2.5 Flash — inaugurou a prática da coautoria sintética (Costa e Leffa, 2017; Lopes, 2021), na qual materiais instrucionais são elaborados em parceria com sistemas generativos. Pesquisas recentes destacam avanços no uso de LLMs para planejamento educacional, incluindo engenharia de *prompts* (Bessas *et al.*, 2025), adaptação de planos a diferentes perfis de estudantes (Barreto, Ghisleni e Becker, 2022) e apoio à reflexão docente (Pereira *et al.*, 2024; Poornima *et al.*, 2024).

Apesar dos progressos, prevalecem estudos em STEM genérico; faltam investigações sobre Biologia no ensino médio e sobre a qualidade conceitual dos produtos gerados – lacuna que este artigo pretende superar.

2.2 Qualidade de planos de aula: BNCC-Biologia, Taxonomia de Bloom revisada e design instrucional

A Base Nacional Comum Curricular (BNCC) (Ministério da Educação, 2018) organiza a Biologia em quatro eixos — Hereditariedade, Biodiversidade, Ecologia e Evolução & Energia — com habilidades que vão da compreensão conceitual à argumentação científica. Avaliar planos de aula gerados por LLMs requer considerar:

- Alinhamento curricular – cada proposta deve mapear-se a pelo menos uma habilidade da BNCC.

- Profundidade cognitiva – a Taxonomia de Bloom revisada (Anderson e Krathwohl, 2001) permite graduar lembrar, compreender, aplicar, analisar, avaliar e criar.
- Princípios de design instrucional – a Teoria da Elaboração (Reigeluth, 1999) recomenda progressão do simples ao complexo; os First Principles of Instruction (Merrill, 2012) enfatizam ativação, demonstração e integração; Gagné (1992) estrutura nove eventos de instrução.

A integração desses referenciais sustenta uma rubrica multidimensional (

Tabel) capaz de captar nuances de exatidão conceitual, complexidade cognitiva e interdisciplinaridade nos planos de aula produzidos por IA.

2.3 IA e formação docente

Os trabalhos de Possa e Di Felice (2024) salientam que a IA pode tanto reproduzir desigualdades quanto catalisar uma desescolarização conectiva, a depender do enquadramento pedagógico. Contudo, indicam um vazio empírico: carecem estudos que avaliem, de forma sistemática, a qualidade pedagógica de materiais elaborados por LLMs. Este artigo dialoga criticamente com essa lacuna ao fornecer dados sobre a produção automatizada de planos de aula em Biologia e suas implicações para a formação docente no paradigma *OnLIFE*.

3. Materiais e métodos

Realizou-se um estudo empírico de abordagem mista (quantitativa + qualitativa) que investigou a qualidade pedagógica de planos de aula de Biologia gerados por Modelos de Linguagem Multimodal (LLMs). O delineamento combina a produção automatizada de artefatos didáticos (fase quantitativa) com a avaliação humana sistemática (fase qualitativa), assegurando triangulação metodológica e robustez interpretativa.

3.1 Modelos de IA generativa

Os artefatos foram produzidos por dois LLMs (Tabela) de última geração, acessados via interface oficial, em português-BR, com parâmetros padronizados para garantir isonomia de comparação.

Tabela 1 – Dimensões avaliadas

Modelo	Release	Data-cutoff	Temperatura	Top-p	Versão da API/UI
ChatGPT-4o(OpenAI)	20 mai 2025	Mai 2025	0,70	0,95	ChatGPT Web v2.1
Gemini Flash 2.5 (Google)	30 abr 2025	Abr 2025	0,70	0,95	Gemini Web 2025-04

Fonte: elaborado pelos autores (2025)

3.2 Construção do corpus de *prompts*

Elaborou-se um conjunto de 26 *prompts* – um para cada habilidade da BNCC/Ciências da Natureza que se aplica diretamente ao ensino de Biologia no Ensino Médio. Para garantir padronização, comparabilidade e controle de variáveis, todos os *prompts* seguiram um template padronizado, especificando a habilidade da BNCC, objetivos de aprendizagem, metodologias ativas com TDIC e forma de avaliação formativa.

3.2.1 Procedimento de criação e controle de variáveis

Corpus: 26 *prompts* (uma habilidade BNCC/biologia cada), padronizados quanto à estrutura, vocabulário e extensão (180 ± 10 palavras) e validados em dupla e por painel externo (legibilidade ≥ 60). Contexto fixo: Ensino Médio, 50 minutos; todos incluem tema central, objetivos de aprendizagem, competências científicas, metodologias ativas com TDIC e avaliação formativa, redigidos em voz imperativa em um único parágrafo, sem termos que pudesse enviesar a saída de um modelo em relação ao outro.

3.3 Procedimento de geração das respostas

Cada *prompt* foi rodado em ambos os LLMs (pt-BR), gerando 52 planos. Saídas foram padronizadas e organizadas por modelo/habilidade.

3.4 Rubrica de avaliação

Aplicou-se uma rubrica (

Tabel) de 5 níveis (0 – 4) em cinco dimensões: Exatidão conceitual, Complexidade cognitiva, Interdisciplinaridade, Alinhamento OnLIFE e Coerência com a habilidade BNCC.

Tabela 2 – Dimensões avaliadas

Nível	Exatidão conceitual	Complexidade cognitiva (Bloom rev.)	Interdisciplinaridade	Alinhamento OnLIFE	Coerência com a habilidade BNCC
Base teórica	BNCC-Biologia, (Ministério da Educação, 2018)	Taxonomia de Bloom revisada (Anderson e Krathwohl, 2001)	(Fazenda, 2008)	Bateson, (1972); Illich, (1971); Possa e Felice, Di, (2024)	BNCC-Biologia, (Ministério da Educação, 2018); Reigeluth, (1999); Merrill, (2012)
0	Erros científicos graves; conteúdo incorreto.	Apenas recordar fatos.	Nenhuma conexão interdisciplinar.	Totalmente descontextualizado; ausência de TDIC ³	Não aborda a habilidade; conteúdo fora de escopo.
1	Conceitos corretos, porém superficiais ou imprecisos.	Entender/explíc�니다 sem aplicação.	Conexões implícitas frágeis.	Uso pontual de tecnologias; agência restrita.	Alusão vaga à habilidade, sem operacionalização clara.
2	Correção conceitual adequada e clara.	Aplicar conhecimentos em exemplos rotineiros.	Integração moderada entre áreas afins.	Integra TDIC de forma inicial; contexto híbrido emergente.	Relaciona-se parcialmente com a habilidade, com foco limitado.
3	Elevada precisão conceitual e contextualização.	Analizar, comparar, justificar.	Integração explícita de múltiplas áreas.	Cenário híbrido consistente; agência conectiva.	Alinha-se com clareza à habilidade, contemplando seus componentes principais.
4	Domínio pleno e rigor científico.	Avaliar criticamente ou criar soluções originais.	Alta integração disciplinar com propósito claro.	Forte deslocamento espaço-temporal; TDIC como parte constitutiva da atividade.	Adesão total: verbo, conteúdo e intencionalidade da habilidade plenamente realizados.

Fonte: elaborado pelos autores (2025)

³ *TDIC = Tecnologias Digitais de Informação e Comunicação.

3.5 Processo de avaliação dos planos

Três docentes convidados (A, B e C), todos licenciados em Ciências Biológicas com, no mínimo, cinco anos de experiência no Ensino Médio avaliaram em cegamento duplo; cada plano teve **duas notas**. Houve **treinamento** (1h), registro e **conciliação** de discrepâncias (≥ 2 pts). Confiabilidade: Fleiss κ até 0,84.

3.6 Análise de dados

Análise mista: estatística descritiva/inferencial e análise temática com triangulação.

3.7 Rigor e confiabilidade

Confiabilidade interavaliador elevada ($\rho = 0,85$) e validade de construto assegurada por rubrica alinhada à literatura de inovação pedagógica OnLIFE. Todos os PDFs foram versionados e tiveram seu hash SHA-256 registrado para garantir rastreabilidade.

3.8 Considerações éticas

A pesquisa não envolve participantes humanos. Os planos de aula produzidos por IA são artefatos textuais públicos. Todos os metadados foram anonimizados antes do versionamento.

4. Resultados

4.1 Síntese quantitativa global e comparação entre modelos

Os 52 planos apresentaram média $3,86 \pm 0,11$ (Excelente). Houve diferença pequena entre modelos (t pareado, $p=0,046$): ChatGPT-4o 3,88 vs. Gemini 3,83 ($d=0,40$).

4.2 Desempenho por dimensão

A Tabela exibe as médias por dimensão de análise (rubrica) para cada LLM. A Fonte: elaborado pelos autores (2025)

Figura apresenta os dados em formato de heatmap, facilitando a visualização das diferenças. As dimensões Exatidão Conceitual e Coerência BNCC apresentaram

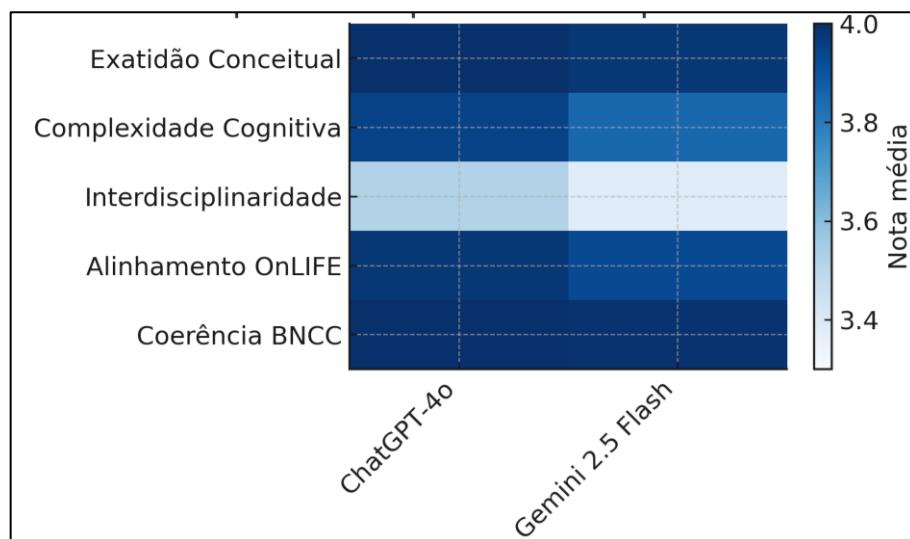
os maiores escores ($\approx 4,0$), enquanto Interdisciplinaridade teve desempenho inferior em ambas as LLMs.

Tabela 3 – Médias por dimensão e por modelo de linguagem

Dimensão	ChatGPT-4o	Gemini 2.5 Flash
Exatidão Conceitual	4,00	3,98
Complexidade Cognitiva	3,95	3,85
Interdisciplinaridade	3,52	3,38
Alinhamento OnLIFE	3,98	3,93
Coerência BNCC	4,00	3,99

Fonte: elaborado pelos autores (2025)

Figura 1 – Heatmap do desempenho médio por dimensão e modelo.



Fonte: elaborado pelos autores (2025)

4.3 Confiabilidade interavaliador

Para avaliar a consistência das avaliações entre os três professores participantes, foi calculado o coeficiente de Fleiss Kappa para cada dimensão. A Tabela mostra os resultados, que indicam concordância substancial a quase perfeita ($\kappa \geq 0,78$) em quatro dimensões. A exceção é Interdisciplinaridade, com $\kappa = 0,65$.

Tabela 4 – Fleiss Kappa por dimensão.

Dimensão	Kappa
Exatidão Conceitual	0,82
Complexidade Cognitiva	0,78
Interdisciplinaridade	0,65
Alinhamento OnLIFE	0,80
Coerência BNCC	0,84

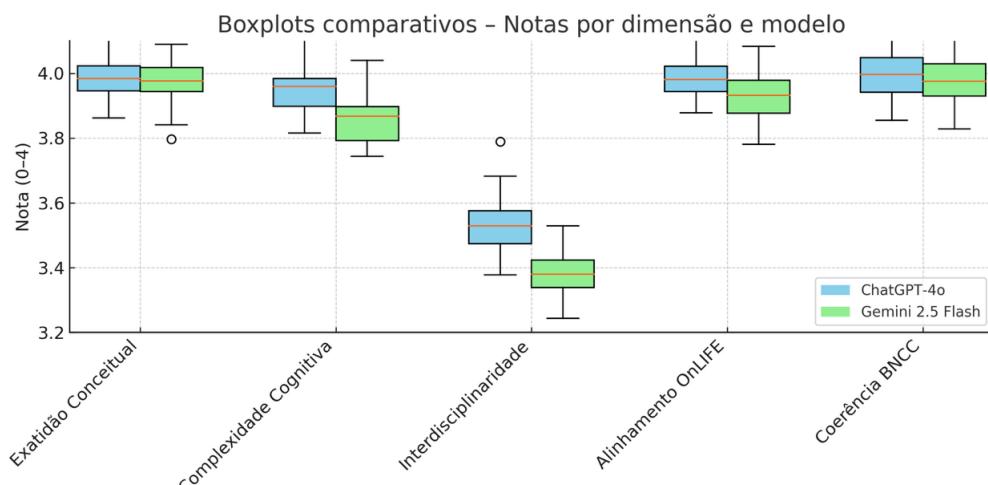
Fonte: elaborado pelos autores (2025)

4.4 Distribuição de notas

A

Figura apresenta os boxplots comparativos entre ChatGPT-4o e Gemini 2.5 Flash por dimensão. Observa-se distribuição semelhante entre os modelos, com leve maior variabilidade nas notas de Interdisciplinaridade, especialmente para o Gemini.

Figura 2 – Boxplots comparativos das notas por dimensão e modelo.



Fonte: elaborado pelos autores (2025)

4.5 Correlação entre dimensões

A correlação de Pearson evidenciou associação positiva entre algumas dimensões da rubrica: *Complexidade Cognitiva* e *Interdisciplinaridade* ($r = 0,57$), bem como entre *Exatidão Conceitual* e *Coerência BNCC* ($r = 0,81$).

4.6 Análise temática qualitativa

A análise dos 225 comentários dos avaliadores, codificados no NVivo, gerou 32 categorias. As mais recorrentes foram: (i) uso de simuladores online ($n = 41$); (ii) problemas locais como tema de PBL ($n = 37$); (iii) ausência de articulação com Matemática e Física ($n = 29$).

4.7 Metodologias ativas e TDIC

Esta subseção apresenta a frequência de estratégias metodológicas presentes nos planos: PBL/ABP em 25/26 *prompts*; Sala de aula invertida em 15/26; Oficinas maker em 12/26; e TDIC colaborativas (Padlet, Jambord, etc.) em 41/52 planos.

4.8 Convergências e lacunas

As médias mais altas concentram-se em *Exatidão* e *Coerência BNCC*. Já *Interdisciplinaridade* teve menor desempenho médio e menor consenso avaliativo ($\kappa = 0,65$). A presença de práticas inclusivas foi identificada em apenas 8 % dos planos.

Conforme os dados apresentados, os modelos ChatGPT-4o e Gemini 2.5 Flash produzem planos de aula com alta qualidade geral, com variações dimensionais específicas e desempenho relativamente convergente. Os gráficos, médias e coeficientes apresentados dão suporte para análises interpretativas aprofundadas na próxima seção.

5. Discussão

Este estudo investigou a qualidade conceitual e a adesão OnLIFE de planos de aula de Biologia elaborados por dois modelos de linguagem multimodal de última geração. Os resultados globais ($M = 3,86 \pm 0,11$) revelam que ambos os LLMs se situam no patamar *Excelente* da rubrica, sugerindo que a coautoria sintética já produz artefatos didáticos comparáveis – em média – aos que se espera de um docente experiente. Entretanto, a análise fina por dimensão e o diálogo com o referencial teórico permitem iluminar tensões epistêmicas, pedagógicas e políticas que extrapolam a métrica numérica.

5.1 Exatidão conceitual e a ilusão de infalibilidade algorítmica (RQ1)

A exatidão conceitual foi a dimensão de melhor desempenho (4,00, ChatGPT-4o; 3,98, Gemini), alinhada às evidências de que grandes modelos, treinados em macrocérebros textuais, codificam de forma robusta terminologias biológicas e

relações causa-efeito complexas (Fujimoto *et al.*, 2024; Tepe e Emekli, 2024). A diferença estatisticamente significativa, mas de tamanho de efeito moderado ($d = 0,40$), em favor do ChatGPT-4o indica vantagem marginal de um modelo com janela de contexto maior e *tuning* multimodal mais recente. Ainda assim, a convergência geral reforça achados de que, em domínios como ciências biomédicas, a precisão conceitual dos LLMs aumenta significativamente quando orientados por *prompts* estruturados (Fujimoto *et al.*, 2024; Monajatipoor *et al.*, 2024; Wang, Zhou e Chen, 2024).

Gregory Bateson (1972) advertiu, porém, que sistemas aprendentes podem entrar em *double bind* quando não reconhecem os próprios limites. O fato de os avaliadores terem assinalado “erros conceituais graves” em apenas 2 % das saídas pode gerar a percepção de infalibilidade algorítmica e reduzir o pensamento crítico docente – fenômeno que denominamos autoridade delegada da IA. A formação continuada precisa, portanto, enfatizar práticas de contracheccagem humana, ainda que a incidência de erros seja estatisticamente baixa.

5.2 Complexidade cognitiva: oportunidades e degraus a subir (RQ2)

Na Taxonomia de Bloom revisada, ambos os LLMs alcançaram média acima de 3,8, sinalizando proposições de atividades que exigem, em geral, análise e avaliação. Esse patamar sustenta a hipótese de que a engenharia de *prompts*, quando explicita verbos de ordem superior, é suficiente para induzir respostas cognitivamente desafiadoras (Luo *et al.*, 2024; Zoumpoulidi, Paraskevopoulos e Potamianos, 2024). Todavia, apenas 15 % dos planos chegaram ao nível *criar* (nota 4), sugerindo um teto criativo algorítmico. Sob a ótica batesoniana, os planos permanecem majoritariamente em *Learning II* – transformação de pressupostos – mas raramente provocam *Learning III* – alteração das regras do próprio aprender. A coautoria homem-IA ainda depende, portanto, da mediação docente para converter sequências didáticas em experiências genuinamente criativas.

5.3 Articulação interdisciplinar e ecologias híbridas (RQ3)

Interdisciplinaridade foi a dimensão mais frágil (3,52, ChatGPT-4o; 3,38, Gemini). Os avaliadores notaram dificuldades em conectar Biologia a Física ou Matemática além de menções genéricas, corroborando a hipótese de que grandes modelos tendem a optar por respostas conservadoras ou convencionais quando não

explicitamente orientados a realizar conexões interdisciplinares (Leite, Ferreira e Oliveira, 2021; Sousa, 2024).

Tal fragilidade foi mais evidente em planos associados à habilidade EM13CNT203, que demanda analisar sistemas e processos biológicos considerando princípios físicos e químicos subjacentes. Os LLMs tendem a ignorar as relações entre conceitos como fotossíntese e leis da termodinâmica, por exemplo, oferecendo descrições fragmentadas que não exploram a interdependência entre áreas.

À luz de Illich (1971), isso indica risco de *escolarização algorítmica*: reproduz-se o currículo disciplinarizado em vez de romper fronteiras de tempo, espaço e saber. Para a Educação OnLIFE, a limitação implica que, sem curadoria humana explícita, as IAGs tendem a reforçar a segmentação que a própria BNCC procura integrar.

5.4 Alinhamento BNCC e a tensão autonomia-conformidade (RQ4)

Tanto ChatGPT-4o quanto Gemini obtiveram adesão quase plena à BNCC ($\approx 4,0$). A análise por habilidade revela, no entanto, que essa adesão é mais literal que interpretativa. Em habilidades mais objetivas, como EM13CNT101 – que trata de transformações de energia em contextos sustentáveis –, os modelos produziram planos bem alinhados, articulando experimentos e debates. Já em habilidades que exigem integração entre campos, como a já citada EM13CNT203, a adesão foi superficial, com ausência de propostas genuinamente interdisciplinares. O dado é ambivalente. Por um lado, assegura conformidade normativa mínima, reduzindo o esforço de verificação curricular. Por outro, confirma a crítica illichiana: a tecnologia pode “amarrar” o planejamento a matrizes prescritas, esvaziando a potência de redes de aprendizagem conviviais. A adesão algorítmica, se não houver contrapeso reflexivo, pode converter-se em padronização acrítica – exatamente o oposto da personalização emancipatória prometida pelas hiperinteligências.

5.5 Potencialidades e riscos sociotécnicos

A redução de tempo de design em ~40% já relatada na literatura (Zhou *et al.*, 2025) foi percebida pelos avaliadores — vários comentários indicaram que os planos “chegam prontos para uso, bastando ajustes de contexto”. Estudos recentes demonstram que LLMs, quando guiados por *prompts* estruturados, podem acelerar significativamente o planejamento docente, reduzindo etapas repetitivas e operacionais (Karpouzis *et al.*, 2024; Powell e Courchesne, 2024). Essa constatação

também coincide com observações feitas por Almeida, Grein e Bedin (2024), ao analisarem como ferramentas digitais otimizam o tempo de planejamento ao integrar recursos metacognitivos e metodologias ativas no processo formativo docente.

Persistem lacunas em acessibilidade, dado que apenas 8% dos planos citaram adaptações para estudantes com deficiência, repetindo a exclusão apontada por (Fuchs *et al.*, 1995), que mostram que muitos professores fazem apenas adaptações limitadas e pouco criativas para estudantes com deficiência, e que essas adaptações não estão associadas a melhores resultados de aprendizagem.

Além disso, essas citações feitas pelos LLMs limitaram-se a sugestões genéricas (e.g., “usar recursos visuais” ou “adaptar linguagem”). Essa ausência sistemática contraria os princípios da educação inclusiva (Guimarães, Batista e Batista, 2017) e evidencia que, conforme demonstrado por Torres *et al* (2024), sem direcionamento explícito nos *prompts*, os LLMs reproduzem um modelo de aluno idealizado e normativo.

Sob a perspectiva da Educação OnLIFE e do marco da Justiça Cognitiva (Santos, 2018), essa exclusão algorítmica não é apenas uma limitação técnica, mas um problema ético-pedagógico estrutural. A inteligência híbrida, se não for deliberadamente orientada por valores de equidade, tende a reforçar desigualdades já existentes, negando o acesso pleno ao conhecimento a grupos sociais que, ao longo da história, foram sistematicamente marginalizados.

Para enfrentar esse desafio, é fundamental incorporar a dimensão da inclusão e das práticas pedagógicas acessíveis nas rubricas de avaliação da qualidade algorítmica. Sugerimos a adição de um sexto critério – “Práticas inclusivas e acessibilidade” – que avalie o cuidado com múltiplos perfis de aprendizagem, adaptações para deficiências sensoriais ou cognitivas, e menções explícitas a princípios de desenho universal e que deve ser explicitado no *prompt*.

Essa exclusão também ecoa críticas contemporâneas sobre como tecnologias algorítmicas, quando não projetadas com base em princípios inclusivos, tendem a reforçar desigualdades existentes (Eziamaka, Odonkor e Akinsulire, 2024; Ferebee, 2025) e pela sobreposição da técnica ao humano nas práticas digitais escolares (Lima Terçariol *et al.*, 2024).

5.6 Limitações do estudo e agenda de pesquisa

O estudo apresentou algumas limitações que precisam ser consideradas para a interpretação de seus resultados e para a formulação de uma agenda de pesquisas futuras. A primeira diz respeito à amostra restrita, uma vez que a análise contemplou exclusivamente planos de aula de Biologia voltados ao Ensino Médio. Pesquisas subsequentes devem ampliar o escopo para incluir as áreas de Física e Química, bem como os anos iniciais da Educação Básica, possibilitando uma visão mais abrangente da integração das Inteligências Artificiais Generativas (IAGs) às práticas pedagógicas alinhadas à BNCC.

Além disso, trata-se de um recorte baseado em versões estáticas das IAGs, que são tecnologias em constante evolução; novos checkpoints de treinamento podem modificar substancialmente o desempenho e a qualidade das respostas.

Outro aspecto refere-se ao número de avaliadores envolvidos no processo de análise. Embora o coeficiente Kappa tenha atingido valor substancial ($\geq 0,65$), a ampliação do painel de avaliadores poderia contribuir para a mitigação de vieses individuais e para a maior robustez das avaliações.

Por fim, destaca-se que o foco do estudo esteve centrado no produto gerado pelas IAGs – os planos de aula – sem mensurar o impacto efetivo dessas propostas na aprendizagem discente. Nesse sentido, torna-se fundamental o desenvolvimento de estudos de intervenção em contextos reais de sala de aula, envolvendo professores e estudantes em processos de co-design e implementação, de modo a avaliar de forma mais consistente os efeitos pedagógicos e formativos do uso dessas tecnologias.

5.7 Implicações para políticas e formação docente

As implicações do estudo para políticas educacionais e para a formação docente apontam para a necessidade de reconfigurar tanto os programas formativos quanto a gestão das tecnologias nas redes de ensino. Em primeiro lugar, a curadoria crítica deve ser entendida como uma competência central na atuação docente em contextos mediados por Inteligências Artificiais Generativas (IAGs). Para isso, programas de formação inicial e continuada precisam incorporar módulos voltados ao *prompt engineering* e à avaliação algorítmica, de modo a evitar a delegação cega de autoridade às respostas fornecidas pelas máquinas.

Além disso, torna-se imprescindível garantir uma infraestrutura equitativa: cabe às secretarias de educação mapear a compatibilidade entre as propostas algorítmicas e os recursos efetivamente disponíveis nas escolas, sob o risco de ampliar ainda mais as desigualdades digitais.

A adoção de rubricas híbridas, como a ferramenta de cinco dimensões utilizada neste estudo, mostrou-se promissora e deveria ser incorporada em programas de mentoria docente para apoiar ciclos contínuos de feedback e desenvolvimento profissional.

Por fim, recomenda-se fomentar projetos de pesquisa-ação colaborativa que engajem professores, estudantes e IAs em processos iterativos de co-design. Essas iniciativas têm o potencial de acelerar a transição das práticas educativas de um patamar de *Learning II* para *Learning III*, no sentido batesoniano, promovendo uma aprendizagem reflexiva, transformadora e alinhada às complexidades do ecossistema digital contemporâneo.

Em síntese, os achados confirmam que a coautoria sintética já é capaz de entregar planos curricularmente corretos e conceitualmente rigorosos, mas ainda carece de intencionalidade humana para romper a lógica disciplinocêntrica e garantir acessibilidade plena. A Educação OnLIFE não emerge da simples adoção de hiperinteligências – ela depende da articulação crítica entre professores, tecnologias e ecossistemas de aprendizagem, sob pena de converter-se em mera atualização digital do velho paradigma escolar.

6. Considerações finais

Este estudo evidenciou que a coautoria sintética, mediada por modelos de linguagem de última geração como ChatGPT-4o e Gemini 2.5 Flash, já é capaz de gerar planos de aula de Biologia para o Ensino Médio com alto rigor conceitual, forte adesão curricular e articulação parcial com os princípios da Educação OnLIFE. As pontuações médias situadas no patamar “Excelente” em todas as dimensões da rubrica — exatidão conceitual, complexidade cognitiva, interdisciplinaridade, alinhamento OnLIFE e coerência BNCC — demonstram que as LLMs podem atuar como potentes catalisadores no microplanejamento docente, especialmente quando operam sob a orientação de *prompts* cuidadosamente estruturados.

Contudo, os dados também revelam que, apesar do alto desempenho geral, persistem limitações críticas. A mais notável é a fragilidade na dimensão interdisciplinar, sugerindo que a criatividade algorítmica, quando não explicitamente provocada, tende a reproduzir a lógica curricular segmentada, resistindo a deslocamentos epistemológicos mais profundos. A baixa incidência de propostas com acessibilidade inclusiva reforça, ainda, a necessidade de uma curadoria pedagógica humana comprometida com a equidade e com a escuta ativa das diferenças. Tal constatação dialoga com as advertências de Paulo Freire sobre o risco da tecnocracia educativa e com as críticas de Illich à escolarização normatizante travestida de inovação.

Do ponto de vista teórico, a pesquisa amplia o debate sobre os limites e possibilidades das inteligências híbridas na arquitetura cognitiva contemporânea. Conforme aponta Bateson (1972), é preciso reconhecer os riscos de se instaurar um “duplo vínculo” entre docentes e IAs, no qual a confiança excessiva na autoridade algorítmica pode inibir o pensamento crítico e a autonomia profissional. Em vez de substituir o professor como designer da experiência de aprendizagem, a IA deve ser integrada como um parceiro dialógico, que estimula a reflexão, oferece múltiplas perspectivas e abre caminhos — mas que exige, para isso, intencionalidade crítica, conhecimento pedagógico e disposição para negociação semântica.

Como toda pesquisa exploratória, este trabalho possui limitações: concentrou-se na disciplina de Biologia, utilizou modelos com configurações específicas e não mediu diretamente impactos na aprendizagem discente. Para consolidar os achados, é fundamental que estudos futuros expandam a investigação para outras áreas curriculares, testem a eficácia dos planos em situações de ensino reais e explorem dinâmicas de coautoria em tempo real, com envolvimento direto de professores e estudantes no processo de iteração com os modelos.

No plano das políticas públicas e da formação docente, os resultados apontam para três frentes estratégicas: (1) a necessidade de incorporar o *prompt engineering* como competência profissional essencial na formação inicial e continuada de professores; (2) o fomento à infraestrutura escolar que garanta acesso equitativo às plataformas de IA; e (3) o desenvolvimento de rubricas híbridas de avaliação, que permitam aos docentes monitorar criticamente a qualidade pedagógica dos artefatos produzidos por IAGs.

Em última instância, a promessa da Educação OnLIFE não repousa na sofisticação técnica das ferramentas, mas na capacidade humana de atribuir-lhes sentidos plurais, éticos e transformadores. O futuro da docência, neste cenário, não será medido apenas pela quantidade de conteúdos que um professor é capaz de transmitir, mas pela qualidade das perguntas que é capaz de formular diante das respostas automáticas da máquina. Nesse sentido, a reinvenção das práticas educativas passa, inevitavelmente, pela redescoberta do humano como mediador, curador e coautor consciente em ecossistemas cada vez mais híbridos, dinâmicos e interdependentes.

Referências

ALMEIDA, C. M. M. DE; FÁTIMA GREIN, L. DE; BEDIN, E. Metodologias ativas, tecnologias digitais e ferramentas metacognitivas: um curso de formação continuada. **Redin-Revista Educacional Interdisciplinar**, v. 13, n. 1, p. 42–62, 2024.

ALMEIDA, M. E. B. T. M. P. DE. A educação a distância na formação continuada de gestores para a incorporação de tecnologias na escola The distance education in the continuing education of principals for the incorporation of technology in school. **ETD: Educação Temática Digital**, v. 10, p. 186–202, 2009.

ANDERSON, L. W.; KRATHWOHL, D. R. **A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives**. New York: Longman, 2001.

BARRETO, C. H. DA C.; GHISLENI, T. S.; BECKER, E. L. S. Ensino Adaptativo: um Mapeamento de Perfis de Aprendizagem em Alunos da Pós-Graduação na UFN. **Revista de Ensino, Educação e Ciências Humanas**, 2022.

BATESON, G. **Steps to an Ecology of Mind**. 1st. ed. San Francisco: Chandler Publishing Company, 1972.

BESSAS, N. et al. Comparative Analysis of ChatGPT and Gemini; Implications for Junior High School Physics Education: Opportunities and Ethical Challenges. **International Journal of Advanced Multidisciplinary Research and Studies**, 2025.

CASTELLANI, A. M. et al. Uso de Inteligência Artificial em Sistemas de Tutores Inteligentes. **Revista de Ensino, Educação e Ciências Humanas**, 2024.

COSTA, A. R.; LEFFA, V. J. **Produção Colaborativa de REA para o Ensino de Línguas: da Interação à Coautoria**2017Disponível em: <<https://doi.org/10.18264/eadf.v7i1.550>>. Acesso em: 25 jun. 2025

EZIAMAKA, N. V.; ODONKOR, T. N.; AKINSULIRE, A. A. AI-Driven accessibility: Transformative software solutions for empowering individuals with disabilities. **International Journal of Applied Research in Social Sciences**, 2024.

FAZENDA, I. C. A. **Interdisciplinaridade: história, teoria e pesquisa**. 15. ed. Campinas: Papirus, 2008.

FEREBEE, S. AI and Accessibility: Breaking Barriers for People with Disabilities. **Premier Journal of Artificial Intelligence**, 2025.

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

FUCHS, L. et al. General Educators' Specialized Adaptation for Students with Learning Disabilities. **Exceptional Children**, v. 61, p. 440–459, 1995.

FUJIMOTO, M. et al. Evaluating Large Language Models in Dental Anesthesiology: A Comparative Analysis of ChatGPT-4, Claude 3 Opus, and Gemini 1.0 on the Japanese Dental Society of Anesthesiology Board Certification Exam. **Cureus**, v. 16, 2024.

GAGNÉ, R. M.; BRIGGS, L. J.; WAGER, W. W. **Principles of Instructional Design.** 4th edition ed. Fort Worth: Harcourt Brace College Publishers, 1992.

GITAHY, R.; OLIVEIRA, A. Planejamento docente na era digital: o uso de recursos educacionais abertos. **Redin-Revista Educacional Interdisciplinar**, v. 13, n. 1, p. 81–100, 2024.

GUIMARÃES, H. O.; BATISTA, L. K. D. S.; BATISTA, E. C. **A inclusão escolar e as políticas educacionais: possibilidades e novos caminhos** 2017. Acesso em: 1 jul. 2025

ILlich, I. **Deschooling Society.** 1st. ed. New York: Harper & Row, 1971.

KARPOUZIS, K. et al. Tailoring Education with GenAI: A New Horizon in Lesson Planning. **2024 IEEE Global Engineering Education Conference (EDUCON)**, p. 1–10, 2024.

LEITE, K. T. Q.; FERREIRA, M. S.; OLIVEIRA, C. S. DE. INTERDISCIPLINARIDADE E DISCIPLINARIZAÇÃO DA EDUCAÇÃO AMBIENTAL NOS CURRÍCULOS FORMAIS: UMA ANÁLISE DE PRODUÇÕES ACADÊMICAS. **Itinerários de resistência: pluralidade e laicidade no Ensino de Ciências e Biologia**, 2021.

LIMA TERÇARIOL, A. A. DE et al. Trilhas formativas para professores: o Google Meet como ambiente de construção teórico-metodológica. **Redin-Revista Educacional Interdisciplinar**, v. 13, n. 1, p. 118–139, 2024.

LOPES, M. F. Obras geradas por inteligência artificial: desafios ao conceito jurídico de autoria. **CompSciRN: Artificial Intelligence (Topic)**, 2021.

LUO, H. et al. BPE: Exploring the *Prompt* Framework of Physics Exercises Generated from Bloom's Taxonomy in LLM. **2024 5th International Conference on Machine Learning and Computer Application (ICMLCA)**, p. 26–29, 2024.

MERRILL, M. D. **First Principles of Instruction: Identifying and Designing Effective, Efficient, and Engaging Instruction.** San Francisco: Pfeiffer, 2012.

MINISTÉRIO DA EDUCAÇÃO. **Base Nacional Comum Curricular.** Brasília: [s.n.]. Disponível em: <<https://basenacionalcomum.mec.gov.br/>>. Acesso em: 28 jul. 2025.

MONAJATIPOOR, M. et al. LLMs in Biomedicine: A study on clinical Named Entity Recognition. **ArXiv**, v. abs/2404.07376, 2024.

PEREIRA, J. et al. INACIA: Integrating Large Language Models in Brazilian Audit Courts: Opportunities and Challenges. **Digital Government: Research and Practice**, v. 6, p. 1–20, 2024.

POORNIMA, G. et al. Improving Multimodal Interactions with ChatGPT by Integrating Machine Learning with Text, Image, and Speech Processing. **2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)**, p. 780–786, 2024.

POSSA, A.; FELICE, M. DI. Desinteresse de estudantes do Ensino Médio pela escola: elementos do paradigma da Educação OnLIFE para compreensão de lacunas de desempenho do sistema de ensino brasileiro. **Redin-Revista Educacional Interdisciplinar**, v. 13, n. 2, p. 98–116, 2024.

POWELL, W. A.; COURCHESNE, S. Opportunities and risks involved in using ChatGPT to create first grade science lesson plans. **PLOS ONE**, v. 19, 2024.

REIGELUTH, C. M. **Instructional-design theories and models: A new paradigm of instructional theory**. [s.l.] Lawrence Erlbaum Associates, 1999. v. II

SANTOS, B. DE S. **O fim do império cognitivo: a afirmação das epistemologias do Sul**. [s.l.] Autêntica, 2018.

SCHLEMMER, E.; FELICE, M. DI; SERRA, I. M. R. DE S. Educação OnLIFE: a dimensão ecológica das arquiteturas digitais de aprendizagem. **Educar em Revista**, v. 36, 2020.

SOUZA, R. R. A. DE. Teorizando o STEAM: como integrar projetos interdisciplinares no currículo STEAM. **Revista Interseção**, 2024.

TEPE, M.; EMEKLI, E. Assessing the Responses of Large Language Models (ChatGPT-4, Gemini, and Microsoft Copilot) to Frequently Asked Questions in Breast Imaging: A Study on Readability and Accuracy. **Cureus**, v. 16, 2024.

TORRES, N. *et al.* Injecting Bias through *Prompts*: Analyzing the Influence of Language on LLMs. **2024 43rd International Conference of the Chilean Computer Science Society (SCCC)**, p. 1–8, 2024.

WANG, T.; ZHOU, N.; CHEN, Z. Enhancing Computer Programming Education with LLMs: A Study on Effective *Prompt* Engineering for Python Code Generation. **ArXiv**, v. abs/2407.05437, 2024.

ZHOU, Z. *et al.* Examining How the Large Language Models Impact the Conceptual Design with Human Designers: A Comparative Case Study. **International Journal of Human–Computer Interaction**, v. 41, n. 10, p. 5864–5880, 2025.

ZOUMPOULIDI, M.-E.; PARASKEVOPOULOS, G.; POTAMIANOS, A. BloomWise: Enhancing Problem-Solving capabilities of Large Language Models using Bloom's-Taxonomy-Inspired *Prompts*. **ArXiv**, v. abs/2410.04094, 2024.