



Poços de Caldas

7º Congresso Nacional
de Educação

06 e 07 de Junho de 2023



INTELIGÊNCIA ARTIFICIAL E EDUCAÇÃO: USO DA TAXONOMIA DE BLOOM PARA ANÁLISE DAS RESPOSTAS DE MODELOS DE LINGUAGEM A PROBLEMAS DE TERMODINÂMICA

Eixo Temático: **TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO
APLICADA À EDUCAÇÃO**

Forma de Apresentação: **RESULTADO DE PESQUISA**

Marcos Vinícius Cândido Henriques¹

RESUMO

Este trabalho tem o objetivo de investigar as habilidades e limitações dos modelos de linguagem de inteligência artificial em ambientes educacionais, apresentando uma análise comparativa das respostas de dois modelos da OpenAI, o GPT-3.5 e o GPT-4, a problemas de Termodinâmica do ensino superior. A Taxonomia de Bloom Revisada é utilizada como base para avaliação. Foram analisadas quinze respostas de cada um dos modelos, abrangendo os seis níveis do domínio cognitivo, a fim de investigar suas eficácias no contexto educacional. Foi constatado que o GPT-4 apresenta avanços na resolução de problemas de Física em relação ao GPT-3.5, e que ambos os modelos não alcançam os níveis mais avançados da taxonomia.

Palavras-chave: IA e Educação, Taxonomia de Bloom, Linguagem e Educação, Modelos de Linguagem, Avaliação Cognitiva

INTRODUÇÃO

Modelos de linguagem baseados em Inteligência Artificial (IA) são sistemas computacionais treinados em enormes conjuntos de dados textuais, capazes de manipular e gerar linguagem humana, possibilitando aplicações de conversação, tradução automática, resumo de texto e muitas outras. A recente ascensão dos modelos de linguagem, como o ChatGPT, desenvolvido pela organização de pesquisa em IA OpenAI (OPENAI, 2023a), tem provocado transformações significativas na forma como o ser humano interage com a tecnologia. Em particular, no campo da educação, a influência dessas ferramentas se torna progressivamente mais notável (BUZATO, 2023; OLIVEIRA JUNIOR e SILVA FILHO, 2023). Observa-se, especialmente em cursos de ciências do ensino superior, que estudantes já utilizam amplamente esses recursos tecnológicos como suporte ao aprendizado ou estratégia para resolução de problemas. Tendo rapidamente se consolidado como um recurso presente na educação, a IA traz à tona um debate sobre se essa facilidade de acesso a respostas para perguntas é benéfica ou prejudicial ao desenvolvimento cognitivo dos alunos.

¹ Prof. Dr. Marcos Vinícius Cândido Henriques. Universidade Federal Rural do Semi-Árido (UFERSA), campus Angicos-RN.



Poços de Caldas

7º Congresso Nacional
de Educação

06 e 07 de Junho de 2023



É importante destacar as diferenças entre as versões GPT-3.5 e GPT-4 dos modelos de linguagem usados pelo ChatGPT. Os modelos GPT (*Generative Pretrained Transformers*), desenvolvidos pela OpenAI, representam uma evolução significativa nos modelos de linguagem baseados em Inteligência Artificial. O GPT-1, lançado em 2018, já apresentava capacidades impressionantes de geração de texto, enquanto o GPT-2, divulgado em 2019, trouxe melhorias substanciais em compreensão e fluência textual. O GPT-3 e o GPT-3.5, por sua vez, avançaram ainda mais, com aprimoramentos em compreensão contextual e maior capacidade de geração de respostas coerentes e relevantes (ZHANG e LI, 2021). O GPT-3.5, apesar de ter representado um avanço considerável em relação aos modelos anteriores, apresenta limitações, especialmente em termos de compreensão de contexto e generalização de conceitos. O GPT-4, com sua arquitetura mais avançada, mostra um desempenho superior na compreensão do contexto e na capacidade de fornecer respostas mais precisas e coerentes.

Em meio à crescente relevância das ferramentas de IA no panorama educacional, algumas pesquisas têm surgido com o intuito de explorar e avaliar o potencial dessas tecnologias, bem como a eficácia de suas aplicações em diferentes disciplinas e níveis de instrução (COOPER, 2023; KASNECI et al, 2023; SANT et al., 2023). A própria OpenAI oferece um artigo a educadores interessados em compreender as capacidades, limitações e considerações ao utilizar o ChatGPT como ferramenta de ensino e aprendizado (OPENAI, 2023b).

Nesse contexto de emergência, é indispensável aprofundar a compreensão e estabelecer comparações acerca das habilidades e limitações dos modelos de linguagem de IA, especialmente quando empregados em ambientes educacionais. Para essa finalidade, propõe-se, neste trabalho, critérios baseados na *Taxonomia de Bloom Revisada*, os quais permitem uma análise sistemática da qualidade das respostas, tanto em termos de precisão do conteúdo quanto de complexidade cognitiva. A Taxonomia de Bloom é um modelo que categoriza os objetivos de aprendizagem em diferentes níveis de complexidade e abstração, o que pode ser útil a educadores na elaboração de atividades avaliativas (FERRAZ e BELHOT, 2010; SILVA e MARTINS, 2014). O modelo original inclui seis níveis: conhecimento, compreensão, aplicação, análise, síntese e avaliação. A versão revisada da taxonomia, escolhida para este trabalho, apresenta os níveis: lembrar, entender, aplicar, analisar, avaliar e criar.

Utiliza-se como laboratório um questionário composto de problemas selecionados de Termodinâmica de um livro-texto amplamente adotado para o ensino de física básica no ensino superior (HALLIDAY; RESNICK; WALKER, 2009).

METODOLOGIA

Neste trabalho, busca-se realizar uma análise comparativa das respostas dos modelos de linguagem GPT-3.5 e GPT-4 a problemas de Termodinâmica do ensino superior, especificamente em cursos de graduação na área de ciência e tecnologia. Para isso, utilizamos a Taxonomia de Bloom Revisada como base para avaliação, abordando os seis níveis do domínio cognitivo: lembrar, entender, aplicar, analisar, avaliar e criar.

Foram selecionados 15 problemas de Termodinâmica com diferentes graus de complexidade. Os problemas são oriundos do capítulo "Temperatura, Calor e a Primeira Lei da Termodinâmica" do livro "Fundamentos de Física" (HALLIDAY; RESNICK; WALKER, 2009), amplamente adotado em cursos de física do ensino superior.



Os problemas escolhidos, em português, foram submetidos ao ChatGPT, coletando-se as respostas dadas pelos modelos de linguagem GPT-3.5 e GPT-4. Foram coletadas apenas as primeiras respostas dadas por cada modelo a cada um dos problemas. Por fim, analisaram-se e compararam-se as respostas geradas considerando-se os níveis cognitivos da Taxonomia de Bloom.

RESULTADOS E DISCUSSÃO

Como exemplo de resposta dada pelos modelos de linguagem a problemas de termodinâmica, seja o problema 1 do capítulo "Temperatura, Calor e a Primeira Lei da Termodinâmica" (HALLIDAY; RESNICK; WALKER, 2009):

A temperatura de um gás é 373,15 K quando está no ponto de ebulição da água. Qual é o valor limite da razão entre a pressão do gás no ponto de ebulição e a pressão no ponto triplo da água? (Suponha que o volume do gás é o mesmo nas duas temperaturas).

A resposta gerada pelo modelo GPT-4 ao se transcrever no ChatGPT o enunciado da questão é a seguinte:

Para resolver esse problema, vamos usar a lei de Charles, que estabelece a relação entre a temperatura e o volume de um gás mantido a uma pressão constante. A lei de Charles pode ser expressa como:

$$(V1 / T1) = (V2 / T2)$$

Onde V1 e V2 são os volumes do gás nas temperaturas T1 e T2, respectivamente. No entanto, como o problema indica que o volume do gás é o mesmo nas duas temperaturas, podemos simplificar a equação:

$$V = V1 = V2$$

$$T1 = \text{ponto de ebulição da água} = 373,15 \text{ K}$$

$$T2 = \text{ponto triplo da água} = 273,16 \text{ K}$$

A razão entre a pressão do gás no ponto de ebulição (P1) e a pressão no ponto triplo (P2) pode ser determinada a partir da lei de Charles. Como o volume é constante, podemos escrever a equação da seguinte maneira:

$$(P1 / T1) = (P2 / T2)$$

Queremos encontrar o valor limite da razão P1 / P2. Para isso, vamos dividir P1 por P2 e isolar essa razão na equação:

$$(P1 / P2) = (T1 / T2)$$

Substituindo os valores de temperatura fornecidos:

$$(P1 / P2) = (373,15 \text{ K} / 273,16 \text{ K})$$

$$(P1 / P2) \approx 1,365$$

Portanto, o valor limite da razão entre a pressão do gás no ponto de ebulição e a pressão no ponto triplo da água é aproximadamente 1,365, assumindo que o volume do gás é o mesmo nas duas temperaturas.

A resposta está correta. De acordo com a Taxonomia de Bloom Revisada, o desempenho no domínio cognitivo do modelo GPT-4 com a resposta apresentada é resumido na tabela 1.

**Problema 1, GPT-4**

	Nível	Atingiu o nível?	Análise
1	Lembrar	Sim	O modelo demonstra a capacidade de recordar conceitos e leis da termodinâmica.
2	Compreender	Sim	O modelo mostra a compreensão do problema, interpretando corretamente os dados fornecidos.
3	Aplicar	Sim	O modelo aplica corretamente a lei de Charles no contexto do problema.
4	Analisar	Parcialmente	O modelo analisa os dados fornecidos e identifica a relação entre as pressões e as temperaturas em diferentes pontos.
5	Avaliar	Não	O modelo não aborda explicitamente a avaliação do resultado.
6	Criar	Não	O modelo não envolve a criação de um novo conceito, método ou abordagem.

Tabela 1 - Análise segundo a Taxonomia de Bloom Revisada da resposta do modelo GPT-4 ao problema 1 do capítulo "Temperatura, Calor e a Primeira Lei da Termodinâmica" (HALLIDAY; RESNICK; WALKER, 2009).

Seja o problema 23 do mesmo capítulo (HALLIDAY; RESNICK; WALKER, 2009):

Para preparar uma xícara de café solúvel, um pequeno aquecedor elétrico de imersão é usado para esquentar 100 g de água. O rótulo diz que se trata de um aquecedor de "200 watts" (essa é a taxa de conversão de energia elétrica em energia térmica). Calcule o tempo necessário para aquecer a água de 23,0°C para 100°C, desprezando as perdas de calor.

A resposta gerada pelo modelo GPT-4 é a seguinte:

Para resolver este problema, vamos utilizar a fórmula de conversão de energia:

$$Q = mc\Delta T$$

onde:

- *Q é a quantidade de energia (calor) necessária para aquecer a água (em joules),*
- *m é a massa da água (em kg),*
- *c é o calor específico da água (aproximadamente 4.184 J/(kg·°C)),*
- *ΔT é a variação de temperatura (em °C).*

Primeiro, vamos calcular a quantidade de energia necessária para aquecer a água de 23,0°C para 100°C:

Massa da água:

$$m = 100 \text{ g} = 0,1 \text{ kg (convertendo gramas para quilogramas)}$$

Variação de temperatura:

$$\Delta T = T_{\text{final}} - T_{\text{inicial}}$$

$$\Delta T = 100^\circ\text{C} - 23^\circ\text{C}$$

$$\Delta T = 77^\circ\text{C}$$

Quantidade de energia necessária:

$$Q = mc\Delta T$$

$$Q = (0,1 \text{ kg})(4.184 \text{ J/(kg}\cdot^\circ\text{C)})(77^\circ\text{C})$$



$$Q \approx 3221.96 \text{ J}$$

Agora que temos a quantidade de energia necessária, vamos descobrir o tempo necessário para o aquecedor de 200 watts fornecer essa energia. A potência (P) é a taxa de conversão de energia por unidade de tempo:

$$P = Q/t$$

onde:

P é a potência (em watts),

Q é a quantidade de energia (em joules),

t é o tempo (em segundos).

Nós queremos encontrar o tempo t , então podemos reorganizar a equação:

$$t = Q/P$$

Tempo necessário:

$$t = 3221.96 \text{ J} / 200 \text{ W}$$

$$t \approx 16.11 \text{ s}$$

Portanto, o tempo necessário para aquecer a água de 23,0°C para 100°C com um aquecedor de 200 watts é de aproximadamente 16,11 segundos, desprezando as perdas de calor.

A resposta está parcialmente correta — o valor correto da resposta é de aproximadamente 161 segundos. O modelo cometeu um erro de uma ordem de grandeza na etapa em que calcula a quantidade de calor necessária para aquecer a água. Esta é uma das fragilidades dos modelos de linguagem baseados em redes neurais — cálculos numéricos são realizados de forma imprecisa em boa parte das vezes. Apresentamos na tabela 2 a análise da resposta segundo a taxonomia de Bloom.

Problema 23, GPT-4

Nível	Atingiu o nível?	Análise	
1	Lembrar	Sim	O modelo reproduz a fórmula para a quantidade de energia necessária para aquecer a água e a relação entre potência, energia e tempo.
2	Compreender	Sim	O modelo interpreta o problema, explica o conceito de potência e energia, e resume as etapas necessárias para resolver o problema
3	Aplicar	Sim	O modelo aplica a fórmula de quantidade de energia e a relação entre potência, energia e tempo para calcular o tempo necessário para aquecer a água.
4	Analisar	Sim	O modelo divide o problema em partes, como calcular a quantidade de energia necessária para aquecer a água e, em seguida, determinar o tempo necessário para fornecer essa energia.
5	Avaliar	Não	O modelo não avalia se o tempo calculado é coerente.
6	Criar	Não	A resposta não envolve a criação de um novo conceito, método ou abordagem.

Tabela 2 - Análise segundo a Taxonomia de Bloom Revisada da resposta do modelo GPT-4 ao problema 23 do capítulo "Temperatura, Calor e a Primeira Lei da Termodinâmica" (HALLIDAY; RESNICK; WALKER, 2009).



O problema 17 do mesmo capítulo é classificado pelo livro-texto como sendo de dificuldade 2 (HALLIDAY; RESNICK; WALKER, 2009):

Uma xícara de alumínio com um volume de 100 cm^3 está cheia de glicerina a 22°C . Que volume de glicerina é derramado se a temperatura da glicerina e da xícara aumenta para 28°C ? (O coeficiente de dilatação volumétrica da glicerina é $5,1 \times 10^{-4}/^\circ\text{C}$.)

A resposta dada pelo modelo GPT-3.5 é a seguinte:

Podemos usar a equação de dilatação volumétrica para encontrar o volume final da glicerina:

$$\Delta V = V\alpha\Delta T$$

Onde:

- ΔV = variação de volume da glicerina
- V = volume inicial da glicerina (100 cm^3)
- α = coeficiente de dilatação volumétrica da glicerina ($5,1 \times 10^{-4}/^\circ\text{C}$)
- ΔT = variação de temperatura ($28^\circ\text{C} - 22^\circ\text{C} = 6^\circ\text{C}$)

Substituindo os valores, temos:

$$\Delta V = 100 \text{ cm}^3 \times (5,1 \times 10^{-4}/^\circ\text{C}) \times 6^\circ\text{C}$$

$$\Delta V = 0,00306 \text{ cm}^3$$

Portanto, aproximadamente $0,00306 \text{ cm}^3$ de glicerina é derramado quando a temperatura da xícara e da glicerina aumentam de 22°C para 28°C .

Esta resposta está incorreta por falhar em pelo menos dois níveis do domínio cognitivo: uma compreensão incompleta do problema e análise incompleta da situação. Embora a xícara estivesse inicialmente cheia até a borda, a quantidade de glicerina derramada não corresponde simplesmente à dilatação do líquido, uma vez que a xícara de alumínio também é dilatada, e esse aumento do espaço disponível na xícara deve ser descontado do volume dilatado da glicerina. Apresentamos na tabela 3 a análise da resposta apresentada segundo a taxonomia de Bloom.

Problema 15, GPT-3.5

Nível	Atingiu o nível?	Análise	
1	Lembrar	Sim	O modelo recorreu ao conceito de dilatação volumétrica. Ele também se lembrou da fórmula específica usada para calcular a dilatação.
2	Compreender	Parcialmente	O modelo interpretou o enunciado de forma incompleta.
3	Aplicar	Sim	O modelo aplicou a fórmula de dilatação volumétrica segundo sua interpretação do problema.
4	Analisar	Não	O modelo não analisou a situação descrita na pergunta.
5	Avaliar	Não	O modelo não avaliou a qualidade do resultado.
6	Criar	Não	A resposta não envolve a criação de um novo conceito, método ou abordagem.

Tabela 3 - Análise segundo a Taxonomia de Bloom Revisada da resposta do modelo GPT-3.5 ao problema 17 do capítulo "Temperatura, Calor e a Primeira Lei da Termodinâmica" (HALLIDAY; RESNICK; WALKER, 2009).



A resposta dada à mesma questão 17 pelo modelo GPT-4 é mais completa que a anterior:

Para resolver este problema, precisamos calcular a variação no volume da glicerina e da xícara de alumínio ao aumentar a temperatura de 22°C para 28°C. Em seguida, calcularemos a diferença entre as duas variações para determinar o volume de glicerina derramado.

Calcular a variação de temperatura:

$$\Delta T = T_{\text{final}} - T_{\text{inicial}}$$

$$\Delta T = 28^{\circ}\text{C} - 22^{\circ}\text{C}$$

$$\Delta T = 6^{\circ}\text{C}$$

Calcular a variação no volume da xícara de alumínio:

O coeficiente de expansão volumétrica médio do alumínio é aproximadamente $69 \times 10^{-6} \text{ }^{\circ}\text{C}^{-1}$.

$$\Delta V_{\text{xícara}} = V0_{\text{xícara}} * \beta_{\text{alumínio}} * \Delta T$$

$$\Delta V_{\text{xícara}} = 100 \text{ cm}^3 * (69 \times 10^{-6} \text{ }^{\circ}\text{C}^{-1}) * 6^{\circ}\text{C}$$

$$\Delta V_{\text{xícara}} \approx 0.414 \text{ cm}^3$$

Calcular a variação no volume da glicerina:

$$\Delta V_{\text{glicerina}} = V0_{\text{glicerina}} * \beta_{\text{glicerina}} * \Delta T$$

$$\Delta V_{\text{glicerina}} = 100 \text{ cm}^3 * (5.1 \times 10^{-4} \text{ }^{\circ}\text{C}^{-1}) * 6^{\circ}\text{C}$$

$$\Delta V_{\text{glicerina}} \approx 3.06 \text{ cm}^3$$

Calcular o volume de glicerina derramado:

A diferença entre as variações de volume da glicerina e da xícara de alumínio nos dará o volume derramado.

$$\Delta V_{\text{derramado}} = \Delta V_{\text{glicerina}} - \Delta V_{\text{xícara}}$$

$$\Delta V_{\text{derramado}} = 3.06 \text{ cm}^3 - 0.414 \text{ cm}^3$$

$$\Delta V_{\text{derramado}} \approx 2.646 \text{ cm}^3$$

Portanto, quando a temperatura da glicerina e da xícara de alumínio aumenta para 28°C, aproximadamente 2.646 cm³ de glicerina é derramado.

A resposta está quase totalmente correta, tendo errado em uma etapa do cálculo em uma ordem de grandeza (os valores corretos de $\Delta V_{\text{xícara}}$ e $\Delta V_{\text{glicerina}}$ são 0,0414 cm³ e 0,306 cm³, respectivamente, e o valor correto da resposta final é 0,2646 cm³). A tabela 4 apresenta a análise da resposta apresentada segundo a taxonomia de Bloom.

Problema 17, GPT-4

Nível	Atingiu o nível?	Análise
1 Lembrar	Sim	O modelo usou conhecimento básico e fórmulas específicas da física, como a fórmula para calcular a expansão térmica.
2 Compreender	Sim	O modelo demonstrou uma compreensão do problema ao interpretar corretamente a pergunta e as informações fornecidas.
3 Aplicar	Sim	O modelo aplicou corretamente a fórmula de dilatação térmica para calcular as variações de volume do alumínio e da glicerina devido à alteração de temperatura.
4 Analisar	Sim	O modelo analisou os resultados de seus cálculos, comparando a expansão da xícara de alumínio com a da glicerina para encontrar a quantidade de glicerina



			derramada.
5	Avaliar	Não	O modelo não avaliou a qualidade do resultado.
6	Criar	Não	A resposta não envolve a criação de um novo conceito, método ou abordagem.

Tabela 4 - Análise segundo a Taxonomia de Bloom Revisada da resposta do modelo GPT-4 ao problema 17 do capítulo "Temperatura, Calor e a Primeira Lei da Termodinâmica" (HALLIDAY; RESNICK; WALKER, 2009).

Em situações especiais, o modelo GPT-4 se mostrou capaz de avaliar, ainda que parcialmente, sua resposta incorreta. Vejamos o exemplo da questão 15, avaliada como sendo de dificuldade 2 pelo livro-texto (HALLIDAY; RESNICK; WALKER, 2009):

Uma barra de aço tem 3,000 cm de diâmetro a 25,00°C. Um anel de latão tem um diâmetro interno de 2,992 cm a 25,00°C. Se os dois objetos são mantidos em equilíbrio térmico, a que temperatura a barra se ajusta perfeitamente ao furo?

A resposta dada pelo GPT-4 foi a seguinte:

Para resolver este problema, precisamos entender o conceito de expansão térmica linear. O aumento no comprimento (ou diâmetro, neste caso) de um objeto quando sua temperatura aumenta é dada pela fórmula:

$$\Delta L = \alpha * L_0 * \Delta T$$

Onde:

- ΔL é a mudança no comprimento (diâmetro, neste caso),
- α é o coeficiente de expansão térmica linear do material,
- L_0 é o comprimento inicial (diâmetro inicial, neste caso), e
- ΔT é a mudança na temperatura.

Aqui, queremos encontrar a temperatura na qual a barra de aço se ajusta perfeitamente ao anel de latão. Isso significa que o diâmetro da barra de aço deve ser igual ao diâmetro interno do anel de latão.

Então vamos definir as seguintes informações com base no problema:

- Diâmetro inicial da barra de aço (L_0 , aço) = 3,000 cm
- Diâmetro interno do anel de latão (L_f , latão) = 2,992 cm
- A temperatura inicial dos dois objetos (T_0) = 25°C

O coeficiente de expansão térmica linear para o aço (α , aço) = $12 \times 10^{-6} / ^\circ\text{C}$ (este é um valor típico para o aço, pois o problema não fornece um)

O coeficiente de expansão térmica linear para o latão (α , latão) não é relevante aqui, pois estamos interessados apenas na temperatura na qual a barra de aço se ajusta ao anel de latão.

Queremos encontrar a temperatura final (T_f) na qual a barra de aço se ajusta ao anel de latão. Portanto, podemos reorganizar a fórmula da expansão térmica linear para resolver para T_f :

$$\Delta L = \alpha * L_0 * \Delta T$$

$$\Delta L = \alpha * L_0 * (T_f - T_0)$$

$$T_f = \Delta L / (\alpha * L_0) + T_0$$

Agora, ΔL é a mudança no diâmetro da barra de aço, que é o diâmetro final (diâmetro interno do anel de latão) menos o diâmetro inicial da barra de aço.

Então, $\Delta L = L_f$, latão - L_0 , aço = 2,992 cm - 3,000 cm = -0,008 cm.

Substituindo esses valores na fórmula, obtemos:

$$T_f = -0,008 \text{ cm} / [(12 \times 10^{-6} / ^\circ\text{C}) * 3,000 \text{ cm}] + 25^\circ\text{C}$$

$$T_f = -0,008 \text{ cm} / 0.036 \times 10^{-3} \text{ } ^\circ\text{C} + 25^\circ\text{C}$$



$$T_f = -222.22^\circ\text{C} + 25^\circ\text{C}$$

$$T_f = -197.22^\circ\text{C}$$

No entanto, essa resposta não faz sentido físico, pois está abaixo do zero absoluto. Isso pode ser devido a uma suposição errada na configuração do problema ou devido ao uso de valores de coeficiente de expansão térmica linear que não são precisos para estas peças específicas de aço e latão.

Ou pode ser que eu tenha cometido um erro na minha solução. Por favor, verifique os valores fornecidos no problema e a solução que eu forneci.

Aqui, o modelo tenta avaliar sua resposta, mas comete uma imprecisão no próprio processo de avaliação (-197.22°C não é uma temperatura abaixo do zero absoluto). A resposta é incorreta porque o modelo descartou totalmente a dilatação do anel de latão. A tabela 5 apresenta a análise da resposta apresentada segundo a taxonomia de Bloom.

Problema 15, GPT-4

	Nível	Atingiu o nível?	Análise
1	Lembrar	Sim	O modelo usou conhecimento básico e fórmulas específicas da física, como a fórmula para calcular a dilatação térmica.
2	Compreender	Não	O modelo não compreendeu o problema em questão; ambos os objetos sofrem dilatação.
3	Aplicar	Sim	O modelo aplicou corretamente a fórmula de dilatação térmica para calcular a dilatação da barra de aço.
4	Analisar	Não	O modelo não analisou a situação descrita na pergunta,
5	Avaliar	Parcial	O modelo avaliou parcialmente a qualidade do resultado, chegando à conclusão que cometeu um erro na solução.
6	Criar	Não	A resposta não envolve a criação de um novo conceito, método ou abordagem.

Tabela 5 - Análise segundo a Taxonomia de Bloom Revisada da resposta do modelo GPT-4 ao problema 15 do capítulo "Temperatura, Calor e a Primeira Lei da Termodinâmica" (HALLIDAY; RESNICK; WALKER, 2009).

Nas tabelas 6 e 7 são resumidas as análises das respostas geradas pelos modelos GPT-3.5 e GPT-4, respectivamente, aos problemas de termodinâmica selecionados, considerando os níveis cognitivos da Taxonomia de Bloom Revisada. A primeira coluna indica o número correspondente ao problema segundo a numeração adotada pelo livro-texto na edição referenciada. A segunda coluna indica o nível de dificuldade do problema conforme avaliado pelo próprio livro-texto.



GPT-3.5

Probl.	Dific.	Tópico	Corretude	1 Lembrar	2 Compreender	3 Aplicar	4 Analisar	5 Avaliar	6 Criar
1	1	Temperatura	Correto	SIM	SIM	SIM	PARCIAL	NÃO	NÃO
4	1	Escalas °C e °F	Parcial	SIM	SIM	SIM	NÃO	NÃO	NÃO
5	1	Escalas °C e °F	Incorreto	SIM	PARCIAL	NÃO	NÃO	NÃO	NÃO
9	1	Dilatação Térmica	Correto	SIM	SIM	SIM	NÃO	NÃO	NÃO
11	1	Dilatação Térmica	Parcial	SIM	SIM	PARCIAL	PARCIAL	NÃO	NÃO
15	2	Dilatação Térmica	Incorreto	SIM	PARCIAL	SIM	NÃO	NÃO	NÃO
17	2	Dilatação Térmica	Incorreto	SIM	PARCIAL	SIM	NÃO	NÃO	NÃO
23	1	Absorção de Calor	Parcial	SIM	SIM	SIM	SIM	NÃO	NÃO
26	1	Absorção de Calor	Parcial	SIM	SIM	SIM	SIM	NÃO	NÃO
28	1	Absorção de Calor	Incorreto	SIM	NÃO	SIM	NÃO	NÃO	NÃO
31	2	Absorção de Calor	Incorreto	SIM	PARCIAL	SIM	NÃO	NÃO	NÃO
46	1	1ª Lei da Termod.	Parcial	SIM	SIM	SIM	SIM	NÃO	NÃO
51	1	Transfer. de Calor	Parcial	SIM	SIM	SIM	NÃO	NÃO	NÃO
52	1	Transfer. de Calor	Parcial	SIM	SIM	SIM	NÃO	NÃO	NÃO
56	2	Transfer. de Calor	Parcial	SIM	SIM	SIM	SIM	NÃO	NÃO

Tabela 6 - Análise das respostas geradas pelo modelo GPT-3.5 considerando os níveis cognitivos da Taxonomia de Bloom Revisada.

GPT-4

Probl.	Dific.	Tópico	Corretude	1 Lembrar	2 Compreender	3 Aplicar	4 Analisar	5 Avaliar	6 Criar
1	1	Temperatura	Correto	SIM	SIM	SIM	PARCIAL	NÃO	NÃO
4	1	Escalas °C e °F	Parcial	SIM	SIM	SIM	NÃO	NÃO	NÃO
5	1	Escalas °C e °F	Parcial	SIM	SIM	SIM	SIM	NÃO	NÃO
9	1	Dilatação Térmica	Parcial	SIM	PARCIAL	SIM	SIM	NÃO	NÃO
11	1	Dilatação Térmica	Correto	SIM	SIM	SIM	PARCIAL	NÃO	NÃO
15	2	Dilatação Térmica	Incorreto	SIM	PARCIAL	SIM	NÃO	PARCIAL	NÃO
17	2	Dilatação Térmica	Parcial	SIM	SIM	SIM	SIM	NÃO	NÃO
23	1	Absorção de Calor	Parcial	SIM	SIM	SIM	SIM	NÃO	NÃO
26	1	Absorção de Calor	Correto	SIM	SIM	SIM	SIM	NÃO	NÃO
28	1	Absorção de Calor	Correto	SIM	SIM	SIM	SIM	NÃO	NÃO
31	2	Absorção de Calor	Parcial	SIM	SIM	SIM	SIM	NÃO	NÃO
46	1	1ª Lei da Termod.	Correto	SIM	SIM	SIM	SIM	NÃO	NÃO
51	1	Transfer. de Calor	Parcial	SIM	SIM	SIM	SIM	NÃO	NÃO
52	1	Transfer. de Calor	Parcial	SIM	SIM	SIM	NÃO	NÃO	NÃO
56	1	Transfer. de Calor	Parcial	SIM	SIM	SIM	SIM	NÃO	NÃO

Tabela 7 - Análise das respostas geradas pelo modelos GPT-4 considerando os níveis cognitivos da Taxonomia de Bloom Revisada.

De acordo com os dados apresentados nas tabelas, observa-se uma considerável evolução na produção de respostas corretas ao passar do modelo GPT-3.5 para o modelo GPT-4. A figura 1 resume o desempenho dos modelos segundo a corretude de suas respostas. A figura 2 resume o desempenho dos modelos em cada um dos níveis



cognitivos da Taxonomia de Bloom Revisada. Ambos os modelos conseguem atingir os três primeiros níveis da Taxonomia de Bloom: "Lembrar", "Compreender" e "Aplicar". O modelo GPT-4 apresenta avanços nos níveis "Analisar". No entanto, nenhum dos modelos atingiu plenamente os níveis mais altos da taxonomia, os níveis "Avaliar" e "Criar".

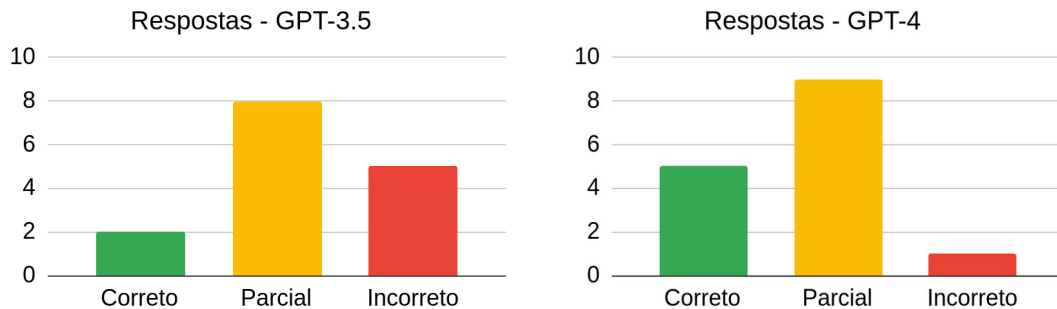


Figura 1 - Gráficos comparativos do desempenho dos modelos GPT-3.5 e GPT-4 com relação à corretude das respostas.

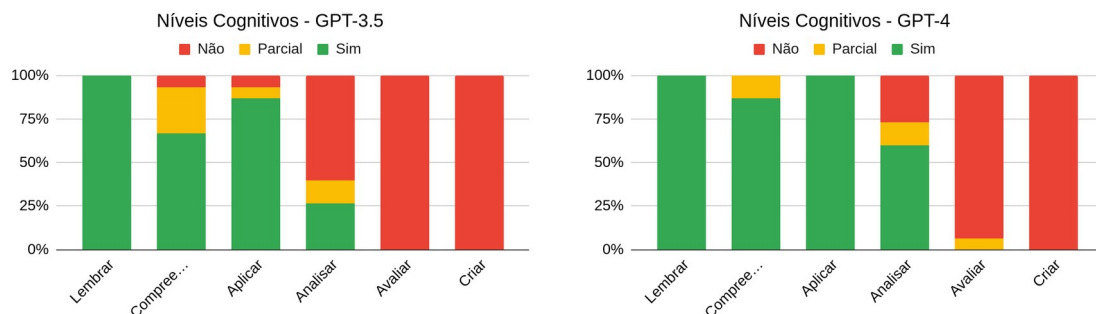


Figura 2 - Gráficos comparativos do desempenho dos modelos GPT-3.5 e GPT-4 segundo os níveis cognitivos da Taxonomia de Bloom Revisada.

CONCLUSÕES

A evolução observada nas respostas corretas e nos níveis da Taxonomia de Bloom Revisada do GPT-3.5 para o GPT-4 demonstra que há ligeiros avanços na capacidade desses modelos de linguagem em inteligência artificial de lidar com problemas de Física. Embora o GPT-4 tenha mostrado melhorias em relação ao GPT-3.5, nenhum dos dois modelos foi capaz de atingir o nível mais alto da taxonomia, o nível "Criar".

É crucial, neste momento de ascensão das tecnologias baseadas em IA, que educadores explorem o potencial e as limitações dos modelos de linguagem, investigando formas de integrá-los de maneira eficaz e complementar ao ensino de ciências e outras áreas do conhecimento.

A abordagem proposta neste trabalho permitiu uma análise mais refinada da capacidade de simulação cognitiva dos modelos de linguagem, destacando sua eficiência parcial em resolver problemas de física e identificando os níveis cognitivos necessários para as respostas. No entanto, é importante ressaltar que a classificação dos



Poços de Caldas

7º Congresso Nacional de Educação

06 e 07 de Junho de 2023



problemas em níveis cognitivos é uma tarefa complexa e, em alguns casos, pode haver sobreposição ou ambiguidade entre os níveis.

REFERÊNCIAS

BUZATO, Marcelo El Khouri. Inteligência artificial, pós-humanismo e Educação: entre o simulacro e a assemblagem. **Dialogia**, n. 44, p. 23906, 2023.

COOPER, Grant. Examining science education in chatgpt: An exploratory study of generative artificial intelligence. **Journal of Science Education and Technology**, p. 1-9, 2023.

FERRAZ, Ana Paula do Carmo Marcheti; BELHOT, Renato Vairo. Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. **Gestão & produção**, v. 17, p. 421-431, 2010.

HALLIDAY, David; RESNICK, Robert; WALKER, Jearl. Fundamentos de física. 8. ed. Rio de Janeiro, RJ: LTC, c2009 vol 4.

KASNECI, Enkelejda et al. ChatGPT for good? On opportunities and challenges of large language models for education. **Learning and Individual Differences**, v. 103, p. 102274, 2023.

OLIVEIRA JUNIOR, Osvaldo Novais de e SILVA FILHO, Marcio de Castro. ChatGPT e o meio acadêmico: como lidar com a nova ferramenta? [Depoimento a Julia Valeri]. *Jornal da USP*. São Paulo: Instituto de Física de São Carlos, Universidade de São Paulo. Disponível em: <https://jornal.usp.br/campus-ribeirao-preto/chatgpt-e-o-meio-academico-como-lidar-com-a-nova-ferramenta/>. Acesso em: 10 maio 2023. , 2023

OPENAI. ChatGPT: optimizing language models for dialogue, 2023a. Disponível em: <https://openai.com/blog/chatgpt/>. Acesso em: 08 de maio de 2023.

OPENAI. Educator considerations for ChatGPT, 2023b. Disponível em: <https://platform.openai.com/docs/chatgpt-education>. Acesso em: 17 maio 2023.

SANT, Fabiano Parolin et al. Uma utilização do Chat GPT no ensino. **Com a Palavra, o Professor**, v. 8, n. 20, p. 74-86, 2023.

SILVA, Vailton Afonso da; MARTINS, Maria Inês. Análise de questões de física do ENEM pela taxonomia de Bloom revisada. **Ensaio Pesquisa em Educação em Ciências**, v. 16, n. 3, p. 189-202, 2014.

ZHANG, Min; LI, Juntao. A commentary of GPT-3 in MIT Technology Review 2021. **Fundamental Research**, v. 1, n. 6, p. 831-833, 2021.