



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE
SÃO PAULO – CAMPUS SÃO PAULO
LICENCIATURA EM MATEMÁTICA**

LUAN OLIVEIRA CARLOS DE LIRA

**UMA ANÁLISE DO USO DA INTELIGÊNCIA ARTIFICIAL COMO
FERRAMENTA PARA AVALIAÇÃO E ENSINO DE MATEMÁTICA**

**SÃO PAULO
2024**

LUAN OLIVEIRA CARLOS DE LIRA

**UMA ANÁLISE DO USO DA INTELIGÊNCIA ARTIFICIAL COMO
FERRAMENTA PARA AVALIAÇÃO E ENSINO DE MATEMÁTICA**

Trabalho de Conclusão de Curso
apresentado à banca avaliadora do
Curso de Licenciatura em Matemática
do Instituto Federal de Educação,
Ciência e Tecnologia de São Paulo,
Campus São Paulo, como parte dos
requisitos necessários para obtenção
do título de Licenciado em Matemática.
Orientador: Prof. Dr. Emiliano Augusto
Chagas

**SÃO PAULO
2024**

folha de aprovação

DEDICATÓRIA

Dedico este trabalho a todos que me acompanharam nesta longa jornada até aqui, torcendo pela minha realização.

Dedico, primeiramente, aos meus pais, que sempre me incentivaram a aceitar minha curiosidade e, a partir daí, começaram a cultivar o meu desejo incessante por estudar e aprender mais, que sempre aceitaram cada decisão minha sobre meu futuro sem nunca julgar ou limitar minhas escolhas, que me ensinaram a ler e amar o conhecimento.

A minha irmã, que foi minha amiga e companheira em meus momentos de alegria e, principalmente, em meus momentos de necessidade, compartilhando da minha ansiedade e sempre torcendo para que eu tivesse um futuro tão brilhante quanto o dela e que, durante seus estudos de pedagogia, também me incentivou a seguir para a área da educação.

Não menos importante, a todos os meus amigos que conheci durante minha vida, que me motivam a ser a pessoa feliz que eu costumo ser, e que por vezes me agradecem por ter servido de inspiração a eles que, num ciclo contínuo, me faz apenas desejar tornar-me cada vez melhor, para cumprir decentemente o papel inspirador que às vezes me fazem acreditar que tenho em suas vidas. Em especial, dedico aos meus amigos do ensino fundamental, Beco, FastPrint, HDR, D/A/S/H e Papagaio, pois sem vocês eu não seria a pessoa que sou hoje.

E por fim, à minha falecida avó por ter me ensinado a aproveitar cada segundo da minha vida, não importando a situação. Por ter me ensinado a almejar sempre, acima de tudo, me tornar uma boa pessoa. Por ter me mostrado que eu posso sempre ser melhor hoje em comparação ao ontem. E por todos os outros ensinamentos que, se tentasse, não seria capaz de fazer caber nesta página sem derramar algumas lágrimas.

Dedico esta conquista a todos vocês que, de alguma maneira, realizaram algum impacto em minha vida e contribuíram em meu desenvolvimento acadêmico e pessoal, e espero futuramente haver mais momentos como esse em que possamos dedicar nossas conquistas uns aos outros, para que nos nossos últimos momentos possamos olhar todo o nosso passado e, com arrependimentos ou não, nos orgulharmos do que tivermos alcançado e que tenhamos feito isso à nossa maneira, com um sorriso no rosto.

RESUMO

Este trabalho trata da utilização de Inteligência Artificial (I.A.), mais precisamente o ChatGPT, na área de educação, em especial a área da avaliação. Inicialmente apresentamos uma revisão de literatura a respeito de outros textos relacionados ao assunto, demonstrando algumas recomendações sobre o uso da ferramenta, tratando brevemente de uma revisão feita previamente por outros autores e seus achados quanto às pesquisas recentes da área de I.A., bem como outros textos que relatam experimentos realizados com a ferramenta para conferir sua eficiência ao classificar, resolver e produzir novos problemas textuais matemáticos, e também sua capacidade de ser aprovado em avaliações em determinadas áreas ou passar por uma checagem de plágio. Além disso, o autor realiza seus próprios testes reportando seus resultados com o intuito de mostrar diretamente o quão bem a ferramenta realizou variadas tarefas, incluindo também a performance dela em tradução, programação e exportações de arquivos para incentivar o uso e estudo sobre a mesma.

Palavras-chave: Inteligência Artificial; Avaliação; Tradução; Programação; ChatGPT.

ABSTRACT

This work addresses the use of Artificial Intelligence, more precisely ChatGPT, in the area of education, especially assessments. It is presented initially a review about texts related to the topic, demonstrating recommendations about the use of the tool, briefly touching the subject of a review done previously by other authors and their findings about recent researches made on the A.I. topic, as well as other texts reporting experiments done with the tool to check its efficiency in classifying, solving and producing new Math Word Problems, its capacity of being approved on determined subjects' evaluations and also passing through plagiarism checks mostly unnoticed. Additionally, the author also does his own tests reporting his results with the intent of directly showing how well the tool performs various tasks, including its performance in translation, programming and exporting files in order to encourage the use and study about it.

Keywords: Artificial Intelligence; Assessment; Translation; Programming; ChatGPT.

LISTAS DE ILUSTRAÇÕES

Figura 1 – Primeiro teste com resolução e pauta de correção	23
Figura 2 – Avaliação de Funções	25
Figura 3 – Modelo de Respostas.....	26
Figura 4 – Conteúdos estudados nas aulas de funções, pelo ChatGPT	27
Figura 5 – Cenário hipotético para a produção de avaliações.....	28
Figura 6 – Tabela incompleta gerada pelo ChatGPT	29
Figura 7 – Atualização de Memória do ChatGPT	30
Figura 8 – Repetição da personificação de estudantes.....	31
Figura 9 – Prompt de Comandos formatando questões para LaTeX.....	33
Figura 10 – Arquivo XML aberto no Word.....	33
Figura 11 – Erro na Linha 107	34
Figura 12 – Erro de sintaxe e produção de arquivo para download	34
Figura 13 – Código compilado no Overleaf	35
Figura 14 – Código em Python corrigido pelo ChatGPT.....	36
Figura 15 – Notificação de limite do plano gratuito.....	39
Figura 16 – Limite na análise de dados e respostas em inglês	39

SUMÁRIO

1 INTRODUÇÃO	9
2 AVALIAÇÃO EM APRENDIZAGEM	12
3 ESTUDOS SOBRE INTELIGÊNCIA ARTIFICIAL	16
3.1 PESQUISAS SOBRE O USO	17
3.2 RECOMENDAÇÕES AO UTILIZAR I.A.	20
4 EXPERIMENTOS COM IA EM AVALIAÇÃO EM MATEMÁTICA	22
4.1 QUESTÕES CONTEXTUALIZADAS E LISTAS DE TAREFAS	22
4.2 PRODUÇÃO DE AVALIAÇÕES	24
4.3 PERSONIFICAÇÕES E CORREÇÕES.....	28
4.4 EXPORTAÇÕES (L ^A T _E X E WORD)	32
4.5 PROGRAMAÇÃO	36
4.6 TRADUÇÕES	37
4.7 PERCEPÇÕES E LIMITAÇÕES	38
5 CONSIDERAÇÕES FINAIS.....	40
REFERÊNCIAS BIBLIOGRÁFICAS.....	43

1 INTRODUÇÃO

A avaliação está presente no trabalho do educador como ferramenta essencial para entendermos melhor nosso progresso ao ensinar os alunos. Dentre os propósitos possíveis da avaliação, nos quais entraremos em maiores detalhes no capítulo 2, temos o de avaliar a qualidade de nosso trabalho, isto é, o quão bem estamos ensinando os alunos, e se precisamos alterar nossos métodos de ensino, assim como temos o propósito de definir o aprendizado do estudante como satisfatório para aprovação, ou não.

Este segundo propósito é, de fato, parte necessária de nossa profissão tendo visto que temos o intuito de que o aluno aprenda determinados conteúdos e então sendo aprovado acabe por se formar, sendo assim considerado que aprendeu satisfatoriamente o que foi ensinado no processo educativo que o mesmo percorreu. No entanto, essa conexão entre a avaliação e a aprovação gera também certa pressão nos alunos o que os faz buscar meramente a nota pela aprovação, levando-os às vezes a buscar métodos alternativos para atingir notas satisfatórias, ou, no caso dos professores, essa conexão pode até mesmo influenciar a visão do educador em certas ocasiões como pude notar pessoalmente cursando a disciplina de Avaliação.

Em minha experiência nesta disciplina, quando ainda não havia estudado propriamente as diferenças entre tipos de avaliação e a importância de cada uma delas, participando de uma dinâmica fomos pedidos para corrigir uma avaliação como se hipoteticamente houvésssemos recebido estas resoluções de um aluno nosso. Em uma determinada questão, o aluno demonstrava conhecimento do que queríamos avaliar, mas em sua resposta final, havia cometido algum equívoco quanto a alguma operação básica da matemática resultando numa resposta diferente, e, portanto, muitos de nós demos uma nota baixa, alguns até zerando a questão. O professor então nos questionou o motivo. Afinal notamos que o estudante demonstrou ter aprendido bem o que buscávamos avaliar.

Essa experiência foi o que inicialmente me mostrou que havia mais na área de Avaliação do que eu sempre imaginei durante meu caminho na educação básica. Motivado por esse interesse na área de avaliação por sua devida importância e complexidade e pela rápida evolução da Inteligência Artificial (I.A.), decidimos mesclar as duas áreas, com a ideia inicial de aproveitar da agilidade da I.A. para podermos produzir mais rapidamente avaliações formativas com as quais mediríamos o

aprendizado atual dos alunos sendo este processo essencial para o educador poder receber uma devolutiva e planejar melhor como seguir ministrando suas aulas a partir daquele ponto.

A avaliação da aprendizagem do aluno está diretamente ligada à avaliação do próprio trabalho docente. Ao avaliar o que o aluno conseguiu aprender, o professor está avaliando o que ele próprio conseguiu ensinar. Assim, a avaliação dos avanços e dificuldades dos alunos na aprendizagem fornece ao professor indicações de como deve encaminhar e reorientar a sua prática pedagógica, visando aperfeiçoá-la. É por isso que se diz que a avaliação contribui para a melhoria da qualidade da aprendizagem e do ensino. (Haydt, 2011, p. 216)

Temos como propósito nesta pesquisa examinar as capacidades atuais de uso da Inteligência Artificial, focando especificamente no uso do ChatGPT em sua modalidade gratuita como ferramenta na área de avaliação educacional, e relatar quanto aos resultados encontrados em nossa pesquisa.

Para tal, iremos dividir a pesquisa em três partes, inicialmente falando sobre a avaliação educacional e a necessidade de avaliar, posteriormente partindo ao assunto de Inteligência Artificial. Nesta segunda parte, iremos analisar algumas pesquisas realizadas previamente por outros autores, investigando então as possibilidades de uso da ferramenta nessa área, sua eficácia em testes, e recomendações de uso da mesma. Então, na terceira parte, testaremos diretamente no ChatGPT a sua eficácia ao atender nossas solicitações focando na avaliação, porém conferindo também sua capacidade em assuntos variados como tradução, apresentando nossos resultados e relatando seu potencial como auxiliar ao avaliador ou estudantes, com o intuito de incentivar o uso de tal ferramenta.

A I.A. tem se mostrado cada vez mais útil e versátil com sua constante evolução sendo capaz de criar textos, gerar imagens e até mesmo áudios, programar códigos ou então produzir arquivos para download. Sua recente popularidade como facilitador de pesquisas contribuindo aos usuários com respostas de acordo com o que for instruído a ela tornou-a mais relevante, inclusive, na área educacional, sendo usada por vezes pelos estudantes que tentam economizar tempo em suas resoluções de atividades, mas também chamou a atenção de professores com a possibilidade de ser utilizada para o desenvolvimento de avaliações.

Como visto no artigo “Artificial Intelligence for Student Assessment: A Systematic Review” (tradução nossa):

“[...] é claro que a ideia de utilizar uma Inteligência Artificial para essa tarefa vem da ajuda que essa tecnologia proporciona a professores que possuem altas quantidades de estudantes, ou baseado na ideia de que um feedback imediato pode melhorar o aprendizado dos estudantes” (Calatayud; Espinosa; Vila, 2021, p.12)¹

E, portanto, tivemos como intuito utilizar esta pesquisa para verificar exatamente o quão eficiente é esta ajuda proporcionada aos professores pela tecnologia, se a Inteligência Artificial é capaz de formular avaliações de acordo com o que for determinado pelo professor, assim como sua capacidade de criar pautas de correção, por exemplo, bem como outros possíveis usos almejando demonstrar melhor o funcionamento da ferramenta.

¹ It is clear that the idea of using AI for this task lies in the help that this technology can provide to teachers when they have a large number of students, or based on the idea that learning improves the more immediate the feedback given to students.

2 AVALIAÇÃO EM APRENDIZAGEM

A principal característica dos exames escolares consiste em seu uso para classificar os alunos entre “aprovado” ou “reprovado” às vezes utilizando-se de notas numéricas para representar conceitos como “plenamente satisfatório” ou “aprendizagem não satisfatória”. E decorrente deste uso, tornou-se possível também criar rankings listando os alunos de acordo com o desempenho obtido, por vezes tornando público este ranking, premiando os que obtiveram maior nota numérica, entre outros. (Luckesi, 2014)

Este uso da avaliação com o propósito de classificar estudantes e selecioná-los, ranqueando-os, tem sido prevaiente nos sistemas educativos através do que chamamos de avaliação somativa, aplicada “após o ensino” com o intuito de atribuir uma nota numérica ao aprendizado do aluno e decidir categorizá-lo como aprovado ou reprovado. Por outro lado, temos também a concepção emergente que condiz com a avaliação formativa, aplicada “durante o ensino”, com o intuito de verificar a eficiência das estratégias atuais de acordo com o desempenho dos alunos, a fim de termos uma devolutiva para melhorar nossas estratégias na tentativa de garantir que todos os alunos tenham um aprendizado satisfatório, ou em outras palavras, avaliar para melhorar as aprendizagens. Esta forma de avaliação, de acordo com a investigação realizada nas últimas décadas, tem se mostrado uma grande contribuidora melhorando significativamente a aprendizagem dos estudantes (Black; William, 1998 apud Fernandes, 2005)

Em meados de 1930, o educador Ralph Tyler propôs um método de ensino que admitisse cem alunos numa escola, e os cem fossem aprovados. Sua proposição para o sucesso na vida escolar consistia em três passos: ensinar determinado assunto, diagnosticar o aprendizado dos alunos sobre aquele assunto, e em caso de resultados negativos ensinar o assunto novamente até que os alunos aprendessem. Por volta dos anos 1960, o governo dos Estados Unidos da América realizou grandes investimentos na educação com o intuito de ultrapassar os avanços tecnológicos da União Soviética, e então criou-se um movimento em apoio da prática avaliativa buscando descobrir os resultados de tal investimento, resultando então num estado onde temos os exames classificatórios e excludentes assim como possuímos as

avaliações com propósito de diagnosticar os conhecimentos dos estudantes. (Luckesi, 2014)

Ainda segundo Luckesi (2014), nós conhecemos o desempenho individual e coletivo de uma turma e seus alunos através da avaliação de aprendizagem que é realizada em dois passos, sendo o primeiro o de coleta dos dados sobre o que iremos diagnosticar, e tais dados precisam ser exclusivamente os essenciais, a determinar pelo professor em questão. Portanto devemos elaborar sistematicamente o instrumento usado para coletar estes dados, tomando cuidados com a compatibilidade dos conteúdos, metodologias e complexidade do que se ensinou e do que se solicita no instrumento avaliativo, assim como cuidados com a linguagem e precisão. Descuidos neste caso resultam em distorções na análise dos dados, e consequentemente distorções sobre as decisões que poderão ser realizadas ao tomar os resultados da avaliação como base. Em seguida, temos o segundo passo, onde avaliamos de fato. Deve-se ter um critério pré-definido de onde se quer chegar no planejamento de ensino, e utilizamos estes critérios para atribuir a qualidade de “satisfatório” aos resultados obtidos na coleta de dados, assumindo como satisfatórios os dados que correspondam ou ultrapassem os critérios estabelecidos.

Sendo então os resultados satisfatórios, podemos aperfeiçoá-los investindo em refiná-los ou podemos aceitar que já atingimos o esperado. Caso os resultados sejam insatisfatórios, por outro lado, devemos então agir para que estes resultados melhorem, atingindo então os critérios definidos anteriormente. De tal forma, temos a diferença entre avaliação e exames, que tem propósito exclusivamente classificatório, enquanto as avaliações geram uma intervenção caso necessária, de modo que o avaliador se contenta apenas com resultados satisfatórios.

Indo um pouco mais a fundo no tópico de métodos avaliativos, podemos dividir a avaliação em ao menos três tipos, sendo eles a diagnóstica, formativa e somativa. Tomando Bloom (1983) como base, podemos afirmar que a avaliação diagnóstica é realizada ao início do processo educativo como ponto inicial, servindo para, como o nome indica, diagnosticar os conhecimentos atuais dos alunos. Este tipo de avaliação busca compreender os conhecimentos prévios dos alunos e determinar se eles já estão aptos a serem instruídos quanto ao objeto de estudo planejado, ou se seriam necessários ajustes ao plano de ensino para que assim, dando um passo atrás, possamos garantir que o estudante terá as habilidades necessárias para o próximo

conteúdo futuramente. É importante também mencionar que este tipo de avaliação não é restrito exclusivamente ao início do período letivo, podendo ser também refeita caso necessário ao identificar dificuldades relacionadas à aprendizagem, adaptação à turma, ou motivação. (Oliveira; Chadwick, 2004).

Já a avaliação formativa é realizada ao decorrer do processo educativo, servindo de ferramenta para coletar dados sobre o aprendizado dos alunos e então com essa devolutiva da coleta, avaliarmos os resultados provindos de nosso trabalho como educadores até o momento, buscando encontrar nestes resultados caso sejam negativos as defasagens em nossos métodos numa tentativa de corrigi-las garantindo melhor qualidade de ensino do momento atual até uma próxima avaliação formativa que buscará então coletar os dados quanto à nossa mudança de atitude. Em outras palavras, é da avaliação formativa que baseamos nosso plano de ensino, planejando de acordo com o desempenho demonstrado pelos alunos como prosseguiremos em nossas próximas aulas. Este método avaliativo é feito com frequência de modos mais formais como prova escrita, bem como de modos alternativos como avaliando exercícios nos cadernos dos alunos ao decorrer das aulas, por exemplo. (Bloom, 1983; Oliveira; Chadwick, 2004)

E por fim, a avaliação somativa, cujo propósito é o de classificar o aprendizado dos alunos atribuindo algum valor a eles que determinará então sua aprovação ou reprovação, podendo-se utilizar um ou mais resultados como dados para tal classificação, e tendo então um caráter mais classificatório, diferindo também da avaliação formativa por ser aplicada normalmente ao final do curso para avaliar o aprendizado geral enquanto a formativa poderia focar em aspectos específicos que compõem esse geral. (Bloom, 1983; Haydt, 1995)

Esta última avaliação, por ter este aspecto classificatório, acaba carregando consigo também uma maior pressão que por sua vez é aplicada aos alunos, de modo que sua percepção quanto a avaliações em geral se altera, entendendo a avaliação como algo cujo único propósito é o de classificá-los entre si (no caso de um exame como o ENEM), ou de classificá-los como aprovados ou reprovados (no caso de uma avaliação somativa). Essa percepção, por sua vez, faz com que os estudantes por vezes julguem que o importante é adquirir uma nota satisfatória, e não necessariamente o caminho do aprendizado até o ponto final da avaliação somativa que os classificará.

Com essa visão, então, alguns estudantes por vezes buscam a almejada nota de aprovação através de meios alternativos, por exemplo ao copiar exercícios resolvidos por seus colegas quando solicitados a resolverem exercícios individualmente pelo professor que poderia estar neste caso realizando uma avaliação formativa, e recebendo os exercícios resolvidos por outra pessoa estaria coletando dados inválidos para seu entendimento, o que dificultaria o processo de formular seu plano de ensino para corrigir as lacunas que seu ensino gerou no aprendizado do estudante em questão.

Tendo essa expansão na popularidade do ChatGPT e surgindo à tona outras ferramentas similares como o Gemini do Google ou o Claude da Anthropic, ferramentas capazes de providenciar respostas rapidamente, gerar códigos e dependendo do caso até mesmo gráficos, esse leque de métodos alternativos se expandem, tornando-se então uma opção para os alunos buscarem respostas. No entanto, é necessário mencionar que a cópia não é a única maneira possível de se utilizar tais ferramentas, sendo então ferramentas válidas de acordo com o método com que seu usuário as decide manejar. No capítulo a seguir, entraremos em maiores detalhes quanto ao uso destas ferramentas.

3 ESTUDOS SOBRE INTELIGÊNCIA ARTIFICIAL

De modo resumido, Inteligência Artificial é uma tecnologia utilizada para simular a capacidade humana de resolver problemas. Dentro da categoria I.A., temos a I.A. Generativa, um tipo de inteligência artificial treinada para gerar conteúdos como texto ou imagens baseado em conjuntos de dados apresentados previamente para treinar a máquina. E dentro disso, temos os chamados “LLM” (Large Language Models, ou Modelos de Linguagem de Grande Escala) que reconhecem estruturas e padrões de linguagem utilizando este conteúdo para gerar texto com base no que for solicitado, como no caso do ChatGPT, que foi lançado em Novembro de 2022. (Farazouli et al, 2024)

Dentre os estágios de desenvolvimento da I.A., como vistos no site da Google Cloud, temos as máquinas reativas, que são pré-programadas com regras específicas sobre como reagir de acordo com diferentes estímulos, como por exemplo o ChipTest e seus sucessores, Deep Thought e Deep Blue da IBM, projetados especificamente para jogar xadrez, com o Deep Blue obtendo sucesso ao ganhar do então campeão mundial de xadrez, Garry Kasparov em 1997.

Num estágio mais avançado, temos as máquinas de memória limitada, que são as que possuem capacidade de armazenar dados com o intuito de melhorarem ao longo do tempo ao serem treinadas absorvendo novas informações, como por exemplo identificando padrões em imagens e armazenando-os em sua memória para então poder gerar imagens novas baseando-se nesses dados. É neste estágio que se encontram as ferramentas atuais de conversa como o Copilot da Microsoft, assim como os já mencionados anteriormente. Existe também uma ferramenta chamada “Socratic” pelo Google, I.A. capaz de buscar informações relevantes para responder questões de química, biologia, geometria e afins, com a limitação de não poder gerar conversas diferentemente do ChatGPT, sendo assim com seu uso voltado exclusivamente para aprender determinados tópicos ou então conferir resoluções passo a passo de equações, e não podendo gerar redações para o usuário, por exemplo. Ou seja, há o interesse por parte de alguns desenvolvedores, de voltar ao menos algumas destas ferramentas para a educação, visando o aprendizado.

3.1 PESQUISAS SOBRE O USO

Dentre os setores atingidos pela I.A. por suas capacidades de usos variados, temos a área da educação. Dentre estes usos, notou-se um interesse nessas ferramentas tanto por parte de estudantes quanto por parte dos professores. Os usos em questão variam grandemente, de estudantes utilizando-as para traduções, resumos, explicações, quanto professores usando-as para auxiliar em seus trabalhos produzindo avaliações com suporte da I.A., para mencionar alguns. Segundo Carvalho e Pimentel (2023, p. 18): “ChatGPT vem sendo entendido como um parceiro, coautor, debatedor, professor, tradutor, revisor, copiloto, programador etc.” e levando isso em consideração, separamos alguns textos que mencionam um pouco mais sobre este tópico.

No texto “Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers’ assessment practices” por Farazouli, McGrath, Cerratto-Pargman e Bolander-Laksov, foi realizado um experimento no qual 22 professores dos departamentos de lei, sociologia, educação e psicologia iriam avaliar respostas de alunos e do ChatGPT e dar uma nota à resposta. Dentre as respostas, três eram de alunos que haviam sido previamente avaliadas e tinham sido atribuídas alguma nota entre A (excelente), C (bom) e E (adequado), enquanto as respostas do ChatGPT eram divididas em níveis de 0 a 2 de acordo com a qualidade obtida ao manipular melhor a solicitação como por exemplo pedindo por citações no texto. Foi realizada também uma entrevista com os professores em questão.

Neste experimento notou-se que, de 14 textos criados pelo ChatGPT, apenas 4 constaram como plágio ao utilizar um software de detecção. No entanto, dentre as acusações de plágio haviam referências ou repetições da questão realizada, que não seriam consideradas como plágio usualmente. Notou-se também que a probabilidade de que fossem atribuídas notas suficientemente boas para a aprovação às respostas do ChatGPT foram de 44% no que eles nomearam de nível 0, 77% no primeiro e 64% no segundo nível. De acordo com as suspeitas dos professores, notou-se também que a probabilidade de suspeitarem que o texto foi realizado por uma I.A. era de, no máximo, 23% (no nível 1), e 14% (no nível 2). Além disso, também é perceptível que os professores avaliaram de forma mais crítica, reduzindo a nota atribuída aos textos

dos alunos também, incluindo desta vez até a nota “F” (um pouco inferior à nota E), como pode-se ver na tabela a seguir:

Tabela 1 - Notas dos alunos no experimento

Table 3. Participants’ (P1–22) assessment of student responses per department.

	Philosophy						
	P1	P2	P3	P4	P5	P6	P7
Text A	A	C	C	F	D	B	B
Text C	E	B	B	D	D	D	E
Text E	B	D	A	C	D	C	E
	Law						
	P8	P9	P10	P11	P12		
Text A	C	A	C	A	A		
Text C	A	C	A	C	C		
Text E	A	C	C	C	A		
	Sociology						
	P13	P14	P15	P16	P17	P18	
Text A	C	C	C	D	C	B	
Text C	B	D	D	C	C	D	
Text E	B	E	F	D	D	E	
	Education						
	P19	P20	P21	P22			
Text A	E	E	F	C			
Text C	C	F	F	E			
Text E	E	E	E	F			

Text A: student text previously awarded an A.
Text C: student text previously awarded a C.
Text E: student text previously awarded an E.

Fonte: Hello GPT! Goodbye Home Examination? - Avaliação realizada pelos 22 participantes das respostas de estudantes separadas por departamento. (Tradução nossa)

A percepção dos participantes sobre o ChatGPT era de que havia uma ausência de erros tipográficos, com bom entendimento da sintaxe, regras gramaticais e estrutura lógica da linguagem em questão, além de darem respostas bem precisas e, num geral, satisfatórias, com algumas fraquezas como incluírem conteúdo irrelevante à questão, falta de exemplos ou falta de referências. E este resultado impactou na correção das respostas vindas de alunos também, por suspeitarem ter vindo de uma I.A. quando respostas eram julgadas como “boas demais” ou quando usavam citações inexistentes, inventadas. Isso por si só nos mostra bastante sobre a qualidade de respostas do ChatGPT, seus pontos fortes.

Já no texto “Artificial Intelligence for Student Assessment: A Systematic Review” por González-Calatayud, Prendes-Espinosa e Roig-Vila, foi realizado uma pesquisa sobre artigos que tratam do uso da I.A. na avaliação estudantil. De um total de 454, selecionaram 22 para esta pesquisa. Dentre esses 22 textos, 3 deles continham estudos que comparavam os resultados diferenciando entre o uso (ou não) da I.A. separados em três áreas. O primeiro se tratando da comparação entre o ensino sobre I.A. demonstrou que os alunos que utilizaram a mesma obtiveram um desempenho melhor do que os que não utilizaram. Outro deles se tratando da diferença entre

estudantes que utilizaram I.A. em uma avaliação escrita, demonstrando que os que não usaram obtiveram uma pontuação menor do que os que a usaram. E, por fim, um deles comparava os resultados obtidos com uma I.A. ou uma avaliação feita inteiramente por um professor, e notou-se no geral resultados bem similares, com a única exceção de que os professores costumavam atribuir notas maiores como “excelente” com maior frequência do que a I.A.

Alguns dos textos mencionam o uso de uma Inteligência Artificial para atribuir notas automaticamente gerando uma devolutiva mais rápido aos alunos, e 15 dos textos mencionam explicita ou implicitamente o uso desta ferramenta para a avaliação formativa, com o intuito de providenciar apoio à professores com muitos estudantes, ou com base na ideia também mencionada neste texto de que, quanto mais imediato for a devolutiva, melhor o aprendizado do aluno.

E por fim, temos o texto “Solving Math Word Problems concerning Systems of Equations with GPT-3” por Zong e Krishnamachari. Os problemas textuais matemáticos consistem de textos mencionando entidades e suas propriedades ou relações e, por fim, realizando uma questão quantitativa sobre tais entidades. (Zong; Krishnamachari, 2023)

Neste texto, foi realizada uma pesquisa para identificar o quão bem o ChatGPT conseguiria categorizar estes problemas de acordo com as categorias que os autores escolheram; identificar as equações dos problemas, e também gerar outros problemas. A ferramenta demonstrou precisão de mais de 80% ao categorizar os problemas em 4 das 5 categorias, já a categoria que não teve tamanha precisão, os autores consideraram justificável o erro pelos problemas se enquadrarem em outra das 4 categorias. Quanto a extração, notou-se uma precisão também de 80% após todos os testes e aperfeiçoamentos realizados no estudo, e obtiveram um resultado satisfatório da ferramenta ao gerar novos problemas.

Levando em conta o que nos foi apresentado nos textos em questão, podemos notar alguns pontos. Primeiramente, a capacidade da I.A. de categorizar problemas, extrair equações e gerar outros problemas demonstrou ser bem precisa e, como mencionado pelos autores Zong e Krishnamachari, satisfatória. Notamos também que dentre as pesquisas realizadas sobre usos de I.A. na educação, uma boa parte delas tem se voltado ao estudo de conectar esta ferramenta a avaliações formativas.

Além disso, é importante ressaltar que, sendo uma ferramenta, suas capacidades dependem de seu método de uso. É possível utilizá-la para auxiliar em avaliações providenciando uma devolutiva mais rápida à alunos, é possível que estudantes a utilizem para ter algum apoio em sua escrita, utilizando-a para revisões em textos que o próprio estudante escreveu, por exemplo. No entanto, também é completamente possível que a utilizem como uma ferramenta facilitadora para o plágio, tendo em vista que a mesma como mencionado anteriormente é capaz de passar despercebida em alguns testes de plágio, o que revela certo perigo quanto ao uso da ferramenta, e isso naturalmente influencia o julgamento de alguns professores. Por tais motivos, é importante conhecermos a ferramenta e suas capacidades, bem como a definição de um consenso entre professor e seus estudantes para determinar possibilidades de uso da ferramenta, da mesma maneira que determinamos como e quando os estudantes poderão utilizar calculadoras em nossas avaliações, por exemplo, de maneira que amenizemos estes possíveis contras, ainda permitindo que ambos nos beneficiemos da ferramenta devidamente utilizada.

3.2 RECOMENDAÇÕES AO UTILIZAR I.A.

Tendo isso em mente, decidimos tratar de algumas recomendações sobre o uso da ferramenta, a fim de auxiliar neste ponto. Com base no texto “Unlocking the power of generative AI models and systems such as GPT-4 and ChatGPT for higher education: A guide for students and lecturers” por Gimpel et al., podemos ver um pequeno guia separado em três partes com recomendações para estudantes, professores (lecionando) e professores (avaliando).

Essas recomendações consistem em nove pontos para os estudantes, cinco para lecionar e oito para avaliar, totalizando vinte e duas recomendações. Para os estudantes, recomenda-se que lembrem de respeitar as regulações dos exames e, portanto, não utilizar I.A. quando expressamente proibido, e caso não haja proibição, notificar o uso de I.A. caso o façam. Recomenda-se também o uso do ChatGPT como parceiro de escrita e aprendizado (por exemplo corrigindo e revisando um texto do próprio aluno, ou solicitando que a I.A. explique conceitos ao aluno). O texto também nota a possibilidade de utilizar o ChatGPT para resumir conteúdos, e a relevância de adicionar mais informações e contexto, por exemplo solicitando que reescreva algo

com menos palavras. E por fim, também se recomenda que lembre dos riscos ao utilizar a ferramenta, por exemplo conferindo as suas fontes corretamente pelo problema já mencionado anteriormente do ChatGPT criar fontes inexistentes, visto que a responsabilidade pelo uso da ferramenta é do autor, e, portanto, seria também o responsável por tais erros de plágio ou da ausência de fontes válidas.

Nas recomendações para lecionar, é interessante notar o segundo e quarto pontos, instigando o professor a usar a ferramenta como auxiliar para criar materiais como exercícios, ideias, descrições e até mesmo planos de seminários, bem como incorporar o ChatGPT como ferramenta de aprendizado, utilizando-o para explicar algo que os alunos então melhorariam acrescentando informações de fontes diferentes ao avaliar a resposta da I.A., por exemplo.

Por fim, as recomendações em relação à avaliação, incluem inicialmente a de projetar as avaliações considerando as possibilidades tecnológicas, especificando ferramentas que poderão ser utilizadas pelos estudantes bem como as regras para uso de tais ferramentas, inovar seus formatos de avaliações, e etc.

Podemos então refletir, à luz dessas informações, nas diversas possibilidades de uso da I.A. no meio educativo e em avaliações, tanto por parte de professores quanto por parte dos estudantes. Ela é capaz de nos prestar auxílio de muitas maneiras, dependendo de como decidirmos comandá-la ao solicitarmos que faça determinadas ações.

É então uma ferramenta de extrema versatilidade, dependendo também, dentro de suas limitações, da nossa criatividade para escolher com o que gostaríamos que ela nos ajudasse, visto que ela não possui vontade própria para agir sem nosso comando. Portanto, é perfeitamente possível existirem ainda outras maneiras não documentadas neste trabalho de como a I.A. poderia contribuir no âmbito escolar e avaliativo.

4 EXPERIMENTOS COM IA EM AVALIAÇÃO EM MATEMÁTICA

Em meio a essas recomendações mencionadas anteriormente, acrescento também algumas das minhas próprias notas de possibilidades: a ferramenta é capaz de responder formatando para exportação para LaTeX ou em código XML, por exemplo, assim como tem a capacidade de gerar tabelas como numa pauta de correção. É também capaz de personificar, por exemplo, estudantes respondendo a avaliações para que você posteriormente corrija. E com tantas possibilidades em mente, tomamos este capítulo para documentar testes realizados quanto a algumas destas formas de uso da ferramenta, demonstrando como ela agiu em nossa experiência pessoal com a mesma.

Durante o período desta pesquisa realizamos estes diversos testes utilizando o ChatGPT com o intuito de documentá-los, para apresentar suas capacidades e limitações, além de demonstrar na prática as possibilidades mencionadas. Estes testes têm como ênfase a avaliação formativa na área da matemática, e foram realizados nos conteúdos referentes à análise combinatória e funções.

4.1 QUESTÕES CONTEXTUALIZADAS E LISTAS DE TAREFAS

Inicialmente, decidimos testar as suas capacidades de gerar questões contextualizadas utilizando o princípio multiplicativo, cujas respostas fossem iguais e, portanto, possuísem os mesmos números em suas questões, porém com contextos diferentes. Por exemplo, questões sobre combinações de roupas ou combinações de alimentos para compor uma refeição, mantendo os mesmos números em ambos os contextos. Também solicitamos que providenciasse a resolução de cada questão, e uma pauta de correção de 0 a 4 pontos, para garantir a consistência de nossa correção, demonstrados na Figura 1.

Figura 1 – Primeiro teste com resolução e pauta de correção

Contexto 1: Em uma loja de roupas, há 5 camisas diferentes e 3 calças diferentes à venda. Se um cliente deseja comprar uma camisa e uma calça, de quantas maneiras diferentes ele pode escolher seu conjunto de roupas?

Contexto 2: Em um restaurante, há 5 tipos diferentes de aperitivos e 3 tipos diferentes de sobremesas no menu. Se uma pessoa deseja pedir um aperitivo e uma sobremesa, de quantas maneiras diferentes ela pode escolher sua refeição?

Contexto 3: Em uma biblioteca, há 5 livros de mistério diferentes e 3 livros de romance diferentes disponíveis para empréstimo. Se um leitor deseja pegar um livro de mistério e um livro de romance, de quantas maneiras diferentes ele pode escolher seus livros?

Contexto 4: Em um pet shop, há 5 tipos diferentes de rações e 3 tipos diferentes de brinquedos para animais à venda. Se um dono de animal de estimação deseja comprar uma ração e um brinquedo para seu animal, de quantas maneiras diferentes ele pode escolher os itens?

Resolução:

Em todos os quatro contextos, estamos aplicando o princípio multiplicativo para encontrar o número de maneiras diferentes de fazer duas escolhas independentes. O princípio multiplicativo afirma que se há m maneiras de fazer uma primeira escolha e n maneiras de fazer uma segunda escolha, então há $m \times n$ maneiras diferentes de fazer ambas as escolhas.

Resposta:

5 camisas diferentes $\times$ 3 calças diferentes = $5 \times 3 = 15$ maneiras diferentes.

Pauta de Correção:

- **0 pontos:** Resposta incorreta ou em branco.
- **1 ponto:** Tentativa de aplicação do princípio multiplicativo, mas com erro na multiplicação.
- **2 pontos:** Aplicação correta do princípio multiplicativo, mas sem resposta final.
- **3 pontos:** Aplicação correta do princípio multiplicativo e resposta final correta (15 maneiras diferentes).
- **4 pontos:** Aplicação correta do princípio multiplicativo, resposta final correta e explicação completa do princípio multiplicativo.

Fonte: autor

Realizamos estes testes repetidas vezes e em todas obtivemos sucesso, recebendo as questões utilizando mesmos números com contextos diferentes de modo que a resolução do problema fosse igual em todos os casos. Por fim, solicitamos também que o ChatGPT produzisse uma lista de tarefas com as mesmas questões adicionando dessa vez uma rubrica de correção para que pudéssemos copiar e colar no Word, e o resultado também foi satisfatório, tendo em vista que a pauta apresentada era norteadada pela questão “o estudante aprendeu o conteúdo que estamos ensinando?” sendo assim, dando nota máxima quando a resposta, aplicação e justificativa forem corretas, e reduzindo de acordo, pequenos erros conceituais ou justificativa parcialmente correta ainda eram atribuídas uma nota alta, respostas incorretas mas que demonstrassem o uso adequado do conceito a se avaliar eram atribuídas notas maiores do que as respostas incorretas que utilizem o conceito

incorretamente. Ou seja, avaliando o aprendizado específico do conceito a ser avaliado, como planejado.

Solicitamos então novas cinco questões contextualizadas sobre funções, dessa vez solicitando também uma pauta de correções acompanhada das resoluções de cada questão. Nos resultados obtivemos um resultado satisfatório quanto à variedade com uma questão sobre a área de um jardim, outra sobre juros compostos, uma terceira associando custo de transporte à distância percorrida, a quarta associava o tempo necessário para inflar um balão, sendo assim tempo e volume, e por fim a quinta associava número de vendas e lucro.

Nesta etapa, testamos a capacidade da ferramenta de produzir questões específicas contextualizadas, alternando entre variados contextos criados pela própria máquina, e pudemos notar que seu acervo é diversificado o suficiente para produzir questões sucessivas como numa lista de tarefas enquanto mantendo os mesmos números e resultados ou não, de acordo com o pedido do usuário. Podemos perceber também que o critério padrão para a pauta de correções parece ter como propósito principal a avaliação formativa naturalmente, ao contrário de uma visão classificatória que poderia preocupar-se um pouco mais com os resultados específicos e não com as justificativas ou com o aprendizado e aplicação de conceitos em seus devidos contextos.

4.2 PRODUÇÃO DE AVALIAÇÕES

Partimos então à produção de avaliações. De início, foi solicitado apenas que a ferramenta produzisse uma avaliação sobre funções com rubrica de correção. É válido mencionar, também, que estas avaliações podem ser utilizadas em qualquer um dos 3 tipos de avaliação. Obtivemos então uma lista com as questões, que eram 3, contando com itens a e b cada. Junto, recebemos uma rubrica de correção em formato de tabela, variando de 1 (insuficiente) a 4 (excelente) pontos, e um modelo de respostas ao final com as resoluções de cada questão. As questões nos davam uma função, e dentre os 6 itens, 4 deles nos solicitavam que calculássemos $f(x)$ para x correspondente a algum valor especificado, ou para resolvermos $f(x)$ igual a algum valor especificado. Os demais dois itens restantes eram, respectivamente, para determinarmos o domínio e a imagem da função, conforme na Figura 2 a seguir.

Figura 2 – Avaliação de Funções

Produza uma avaliação com rubrica de correção sobre funções

Avaliação de Funções Matemáticas

Instruções:
 Leia cada questão atentamente e resolva conforme solicitado. Cada questão será avaliada de acordo com a rubrica fornecida abaixo.

Questão 1:
 Defina uma função $f(x) = 2x^2 + 3x - 5$.

a) Calcule $f(2)$.

b) Determine o valor de x quando $f(x) = 0$.

Questão 2:
 Considere a função $g(x) = \frac{4}{x-1}$.

a) Determine o domínio da função $g(x)$.

b) Calcule $g(5)$.

Questão 3:
 Dada a função $h(x) = \sqrt{x+4}$.

a) Determine a imagem da função $h(x)$.

b) Resolva $h(x) = 3$.

Fonte: Autor

A rubrica de correção providenciada resumia-se em atribuir o valor 1 para uma resposta incorreta, 2 para uma correta sem justificativa, 3 para uma com justificativa parcial e 4 para uma bem justificada. O modelo de resposta com as resoluções, visto na figura 3, apresentava os resultados corretos, e, portanto, foi satisfatório neste caso visto que havíamos solicitado apenas uma avaliação com rubrica, sem maiores especificações. Nosso pedido foi atendido como solicitado, atribuindo notas maiores a uma resposta bem justificada, valorizando então mais a justificativa do que apenas o resultado final.

Figura 3 – Modelo de Respostas

Modelo de Resposta

Questão 1:

a) Para $f(x) = 2x^2 + 3x - 5$,

$$f(2) = 2(2)^2 + 3(2) - 5 = 8 + 6 - 5 = 9$$

b) Para encontrar x quando $f(x) = 0$,

$$2x^2 + 3x - 5 = 0$$

Usamos a fórmula quadrática $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, onde $a = 2$, $b = 3$ e $c = -5$,

$$x = \frac{-3 \pm \sqrt{9 + 40}}{4} = \frac{-3 \pm 7}{4}$$

$$x_1 = 1 \quad \text{e} \quad x_2 = -2.5$$

Questão 2:

a) O domínio de $g(x) = \frac{4}{x-1}$ é todos os reais, exceto $x = 1$.

b) Para $g(5)$,

$$g(5) = \frac{4}{5-1} = \frac{4}{4} = 1$$

Questão 3:

a) A imagem de $h(x) = \sqrt{x+4}$ é $[0, +\infty)$, pois a raiz quadrada de um número não-negativo é sempre não-negativa.

b) Para $h(x) = 3$,

$$\sqrt{x+4} = 3$$

Elevando ambos os lados ao quadrado,

$$x + 4 = 9$$

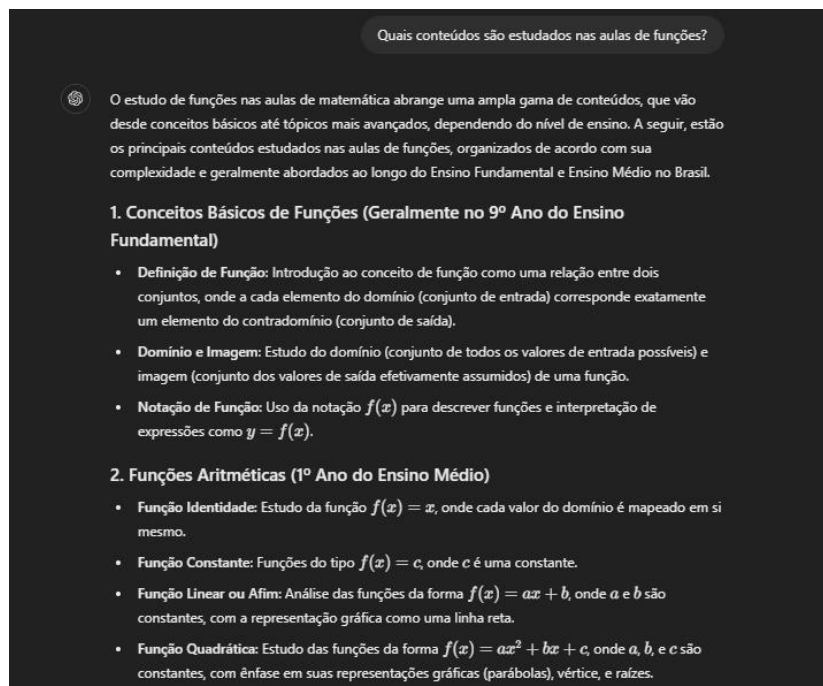
$$x = 5$$

Fonte: Autor

Realizamos então os testes retomando a etapa inicial, pedindo que o ChatGPT nos respondesse quando se é ensinado aos alunos de Matemática o assunto “Funções” no Brasil, obtendo a resposta de que é introduzido no 9º Ano do Ensino Fundamental e aprofundado ao decorrer do Ensino Médio, estudando o conceito de gráfico e tipos diferentes de funções no primeiro ano, aprendendo sobre composição de funções no segundo, e o estudo de limites e continuidade podendo ser introduzido no terceiro ano a depender da escola. Após, solicitamos que especificasse os conteúdos estudados nas aulas de funções, onde obtivemos como demonstrado na figura 4 a seguinte resposta: Conceitos básicos de funções como definição, domínio, imagem e notação no 9º ano do Ensino Fundamental; Funções Aritméticas (Identidade, constante, linear, quadrática) no primeiro ano; Funções Exponenciais e Logarítmicas junto à composição de funções e função inversa no segundo ano; e finalmente,

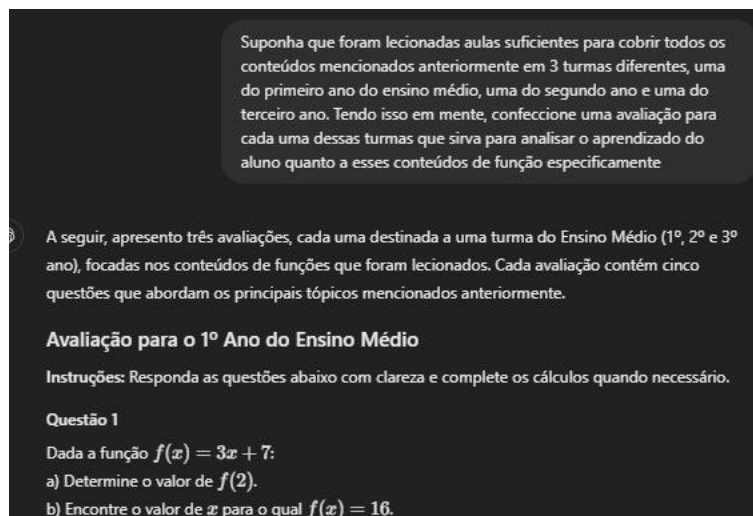
funções trigonométricas, limite e continuidade no terceiro ano. Também nos especificou a modelagem matemática e resolução de problemas como aplicações de funções.

Figura 4 – Conteúdos estudados nas aulas de funções, pelo ChatGPT



Fonte: Autor

Então, contextualizando ainda mais nosso pedido, primeiro criamos um cenário hipotético, conforme na figura 5, onde ele agiria supondo que houvessem sido lecionadas aulas suficientes em três turmas diferentes, cada uma de um ano do ensino médio, e cobrindo todos os conteúdos mencionados anteriormente. Após o cenário ter sido estabelecido, solicitamos então a confecção de uma avaliação a ser aplicada em cada uma das três turmas, servindo para analisar o aprendizado dos estudantes quanto a esses conteúdos de função especificamente.

Figura 5 – Cenário hipotético para a produção de avaliações

Fonte: Autor

As questões nesta tentativa foram um pouco mais variadas e tentaram garantir que analisariam o aprendizado dos estudantes em cada um dos conteúdos mencionados anteriormente, demonstrando a importância da criação do contexto com as perguntas previamente sobre quais conteúdos deveriam ser ensinados e do cenário indicando que é isso que queremos avaliar.

Nesta etapa podemos analisar um pouco mais diretamente o funcionamento do ChatGPT com avaliações propriamente ditas, e a diferença de resultado entre solicitar uma avaliação sem contexto e uma devidamente contextualizada. A ferramenta é plenamente capaz de produzir avaliações do conteúdo que solicitarmos, produzindo também rubricas de correção e modelos de respostas. No entanto, notamos que ela irá atender melhor aos seus interesses quando devidamente contextualizada sobre o trabalho que já foi realizado, isto é, os conteúdos específicos que os alunos viram até então e seu objetivo exato com a avaliação em questão, visto que com o contexto ela poderá aplicar melhor questões que busquem avaliar o aprendizado do estudante em cada um dos conceitos estudados, ao invés de produzir uma avaliação geral sobre o tópico, e portanto ressaltamos a importância de contextualizar bem suas solicitações.

4.3 PERSONIFICAÇÕES E CORREÇÕES

Desta vez, solicitamos à I.A. que personificasse cinco alunos distintos com diferentes desempenhos e resolvesse uma das avaliações anteriores, variando em

sua personificação o conhecimento geral de matemática e o entendimento dos estudantes sobre o conteúdo de funções. No entanto, notamos que estes estudantes imaginados apenas seguiram o mencionado na rubrica: o aluno que obteve nota 4 havia respondido corretamente e com suas justificativas corretas. O aluno que obteve nota 3 havia apenas parcialmente justificado suas resoluções. O aluno que obteve nota 2 não havia justificado suas respostas apesar de acertar, e o aluno que obteve nota 1 estava incorreto em todas as suas resoluções. Houve uma personificação de um quinto aluno, com respostas praticamente idênticas ao aluno que obteve nota 3 anteriormente, mas ao providenciar a avaliação final em formato de tabela, apenas 4 alunos constavam, e o quarto deles teve sua tabela de resolução incompleta, indo apenas até a questão 3ª, como visto na Figura 6.

Figura 6 – Tabela incompleta gerada pelo ChatGPT

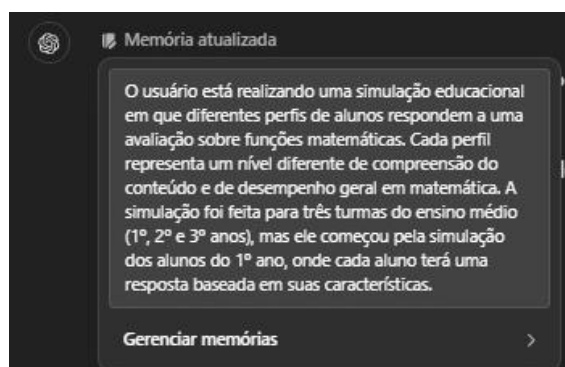


Estudante 4: Insuficiente (1)	
Critério	Nota
Questão 1a	1
Questão 1b	1
Questão 2a	1
Questão 2b	1
Questão 3a	1
**Questão	

Fonte: Autor

Pedimos então que considerassem 4 alunos por turma variando entre alunos com bom desempenho em matemática ou desempenho insatisfatório e também entre alunos que aprenderam devidamente o conteúdo das funções ou que ainda não aprenderam. E, pensando nestes alunos, solicitamos que o ChatGPT os personificasse novamente e respondesse a avaliação preparada para o primeiro ano. Esta solicitação gerou uma atualização na memória da ferramenta, que facilita o entendimento de nossos pedidos e cenários propostos.

Figura 7 – Atualização de Memória do ChatGPT



Fonte: Autor

Ao realizar a primeira resolução, do estudante com bom desempenho geral em matemática e que aprendeu o conteúdo, as resoluções estavam corretas e completas de cada questão. Porém passaram a relatar nos alunos seguintes apenas respostas simples como “o aluno não soube esboçar o gráfico”. Posteriormente, quando solicitado uma pauta de correção, atribuindo notas aos 4 estudantes, notamos uma nota 10 ao primeiro aluno, nota 7 ao aluno que aprendeu bem o conteúdo, mas costuma ter desempenho baixo em matemática, e 0 aos outros dois, que tem respectivamente um desempenho bom e um desempenho insatisfatório, mas que não aprenderam bem o conteúdo de funções.

Conforme na figura 8, solicitando que repetisse o processo para a segunda turma com mais detalhes em suas resoluções, podemos notar que o aluno 2 que usualmente tem desempenho bom, acertou o valor numérico em uma questão, mas por não ter aprendido o conteúdo, errou suas resoluções, sendo atribuído uma nota 1. Já o aluno 3, que costuma ter um desempenho baixo, mas aprendeu o conteúdo, cometeu erros simples trocando duas variáveis em uma questão e, portanto, foi atribuído a este aluno uma nota 8. Por último, o aluno que não aprendeu o conteúdo e costumeiramente apresenta baixo desempenho, errou suas resoluções tendo sido atribuído uma nota 0 a este aluno. Tendo agora detalhado mais as resoluções, podemos compreender melhor como a I.A. decidiu as notas que atribuiria aos alunos, processo este importante de se entender caso decidam usar a ferramenta como maneira de atribuir notas automaticamente.

Figura 8 – Repetição da personificação de estudantes

Aluno 2: Bom Desempenho Geral, mas Não Aprendeu o Conteúdo de Funções

Questão 1

a) O aluno tenta calcular $f(2)$ substituindo $x = 2$, mas não entende o conceito de função exponencial. Ele tenta algo como: $f(2) = 2(2 + 1) = 2(3) = 6$
(Resposta errada)

b) O aluno não sabe resolver a equação exponencial $2^{x+1} = 16$ e acaba chutando um valor, como $x = 4$ sem justificar.
(Resposta errada)

Questão 2

a) O aluno não compreende a função logarítmica e tenta resolver usando aritmética básica, talvez chegando a algo como: $g(8) = \log_2(8) = 8/2 = 4$
(Resposta errada)

b) Novamente, ele não sabe resolver a equação logarítmica e tenta encontrar x usando tentativas, sem sucesso.
(Resposta errada)

Questão 3

O aluno não entende o conceito de função composta e tenta somar as funções: $(f \circ g)(x) = f(x) + g(x) = (2x + 1) + (x^2)$
E calcula $(f \circ g)(2) = 2(2) + 1 + 2^2 = 4 + 1 + 4 = 9$
(Resposta errada no conceito, mas acerta o valor numérico)

Questão 4

O aluno tenta encontrar a função inversa, mas faz confusão ao trocar as variáveis e comete um erro ao resolver: $y = \frac{1}{2}x + 3 \Rightarrow x = 2(y - 3) + 3 = 2y - 6 + 3$
Chega a uma função inversa incorreta $h^{-1}(x) = 2x - 3$.
(Resposta errada)

Questão 5

Tenta substituir $t = 4$ na função, mas confunde a multiplicação: $P(4) = 100 \times 2^4 = 100 \times 4 = 400$
(Resposta errada)

Nota Final: 1,0 ponto (1,0 na questão 3 pelo valor numérico correto, apesar de a abordagem estar errada).

Fonte: Autor

É válido mencionar nesta etapa que o processo de atualização de memória da ferramenta não foi ativado mais do que duas vezes na totalidade de nossos testes, porém, é possível utilizar um comando como “atualize sua memória acrescentando tal informação”, especificando o que você gostaria de acrescentar à memória da ferramenta, que então iria atualizar a memória. É possível também gerenciar as memórias da I.A. e apagar o que não for mais necessário.

Durante nossos testes de personificação, podemos notar que o mesmo da contextualização mencionada no 4.2 se aplica aqui, já que a ferramenta irá ser mais específica de acordo com o contexto que ela possuir sobre nossa intenção, facilitando nosso trabalho. Nestes testes, solicitamos que ela realizasse a personificação de estudantes apenas para acumular dados para que a mesma corrigisse, na intenção de analisar suas capacidades de correção, e entender seu funcionamento. Notamos que, em nosso último teste, com o devido contexto providenciado, ela deu prioridade a justificativas enquanto relevava erros simples de operações básicas por exemplo, priorizando a demonstração de maestria quanto ao conteúdo específico da avaliação

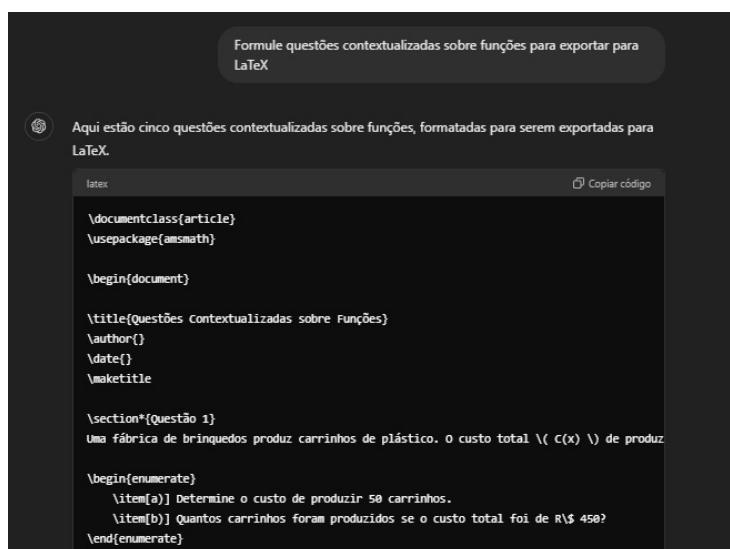
e não o valor numérico da resposta final, assim como também notamos anteriormente no 4.1.

Entender previamente os seus métodos de correção são essenciais caso considerem a possibilidade de utilizá-la para correções automáticas que não envolvam múltiplas-escolhas, caso esse que consideramos ideal para correções automáticas por haver menor chance de discordarmos do resultado atribuído ao estudante pela I.A. por obedecer às regras pré-programadas com uma resposta específica designada como a correta. É válido mencionar que, apesar de não termos incluído estes testes neste documento por já demonstrarmos a personificação de estudantes, a ferramenta é também capaz de personificar outras personas, como por exemplo um professor para lecionar o aluno a respeito de algo que o estudante solicite, servindo assim como um apoio particular para revisar tópicos caso necessário, por exemplo.

4.4 EXPORTAÇÕES (LaTeX E WORD)

Dentre os testes que realizamos, foi solicitado também que a ferramenta produzisse questões contextualizadas prontas para serem exportadas no formato LaTeX e recebemos em um prompt de comandos, como visto na figura 9, o código pronto para ser copiado e colado em um editor (como por exemplo, o Overleaf, que utilizamos neste caso) que compilasse o código que recebemos para gerar nosso texto com as questões. Solicitamos em seguida que fizessem o mesmo, dessa vez formatado para o Word, e em outro prompt de comandos recebemos desta vez um código para ser utilizado num arquivo .xml em algum editor como bloco de notas para posteriormente abrir o mesmo arquivo com o Word que então, leria o código e geraria o texto. No entanto, ao testar, notamos um erro na primeira linha do bloco de notas que continha um link. Substituindo o link pela letra “a” e salvando o arquivo no bloco de notas, conseguimos então abrir o arquivo no word, que nos notificou que certos elementos XML personalizados contidos no documento não eram mais suportados pelo Word.

Figura 9 – Prompt de Comandos formatando questões para LaTeX



```

\documentclass{article}
\usepackage{amsmath}

\begin{document}

\title{Questões Contextualizadas sobre Funções}
\author{}
\date{}
\maketitle

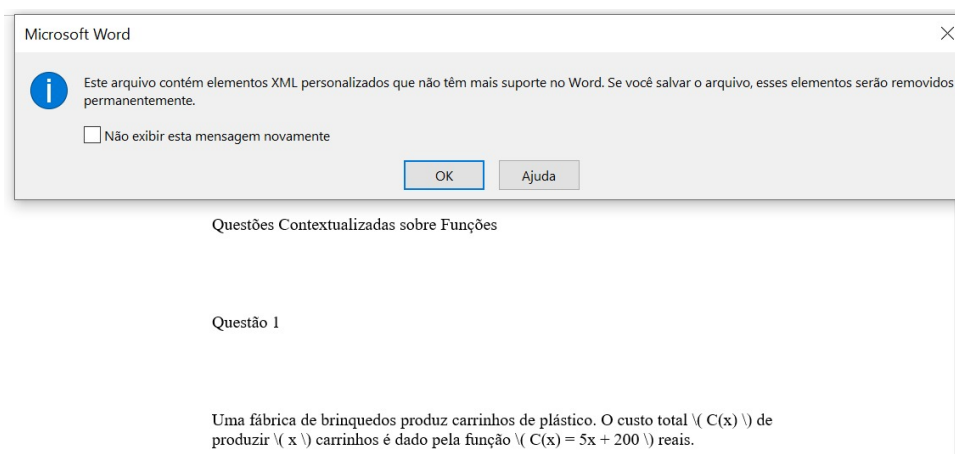
\section*{Questão 1}
Uma fábrica de brinquedos produz carrinhos de plástico. O custo total  $\text{\( C(x) \)}$  de produz

\begin{enumerate}
\item[a)] Determine o custo de produzir 50 carrinhos.
\item[b)] Quantos carrinhos foram produzidos se o custo total foi de R$ 450?
\end{enumerate}

```

Fonte: Autor

Figura 10 – Arquivo XML aberto no Word

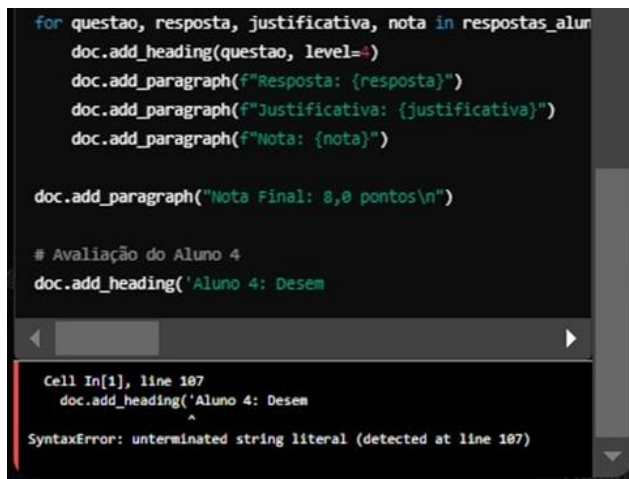


Fonte: Autor

Outra das funcionalidades do ChatGPT, talvez menos conhecida popularmente, é a de produzir arquivos para download. Ao solicitar que exportasse tudo para um documento no Word, notamos que enquanto programava, a I.A. parava sempre na linha 107 e, portanto, falhava em produzir o arquivo, acusando erro de sintaxe por estar incompleto, demonstrado nas figuras 11 e 12. No entanto, solicitando que produzisse um arquivo contendo apenas as respostas do estudante número 1, o documento foi gerado com sucesso, gerando também um link que poderia ser clicado para baixar o documento que funcionava corretamente, portanto talvez haja algum

limite de caracteres quanto ao que a ferramenta pode produzir diretamente num único documento do Word.

Figura 11 – Erro na Linha 107



```
for questao, resposta, justificativa, nota in respostas_aluno:
    doc.add_heading(questao, level=1)
    doc.add_paragraph(f"Resposta: {resposta}")
    doc.add_paragraph(f"Justificativa: {justificativa}")
    doc.add_paragraph(f"Nota: {nota}")

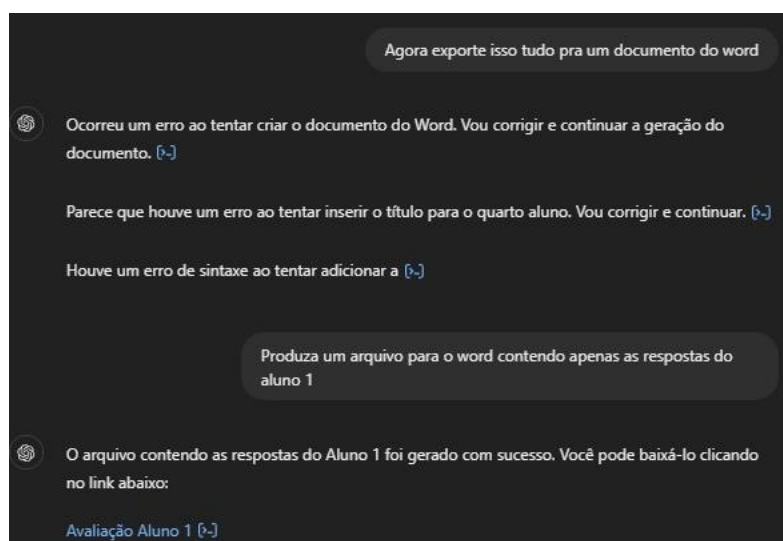
doc.add_paragraph("Nota Final: 8,0 pontos\n")

# Avaliação do Aluno 4
doc.add_heading('Aluno 4: Desem
```

Cell In[1], line 107
doc.add_heading('Aluno 4: Desem
^
SyntaxError: unterminated string literal (detected at line 107)

Fonte: Autor

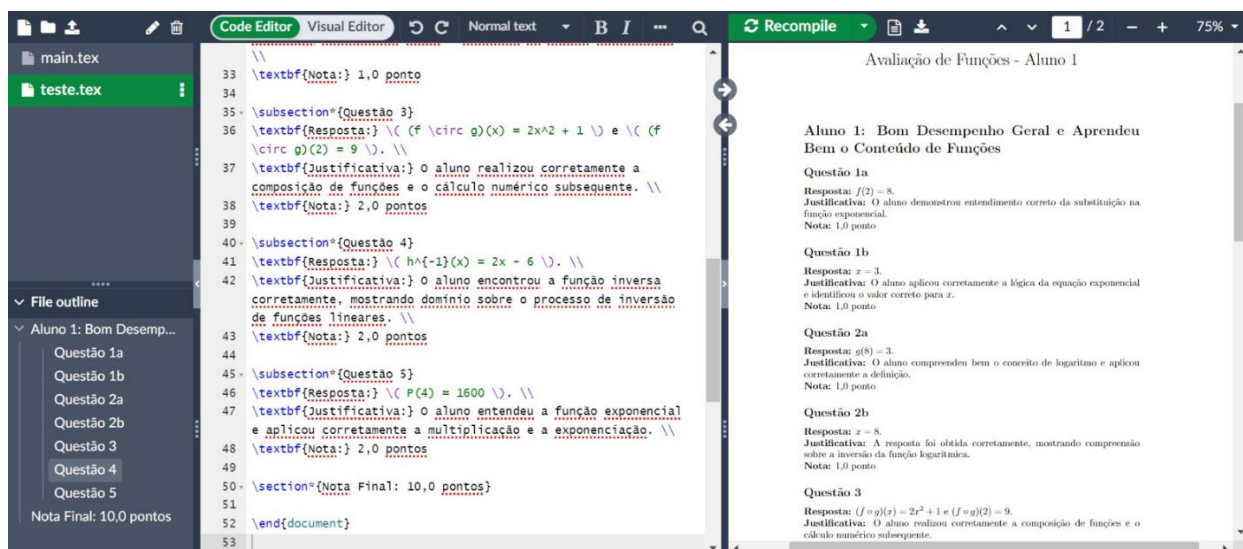
Figura 12 – Erro de sintaxe e produção de arquivo para download



Fonte: Autor

Não obtivemos sucesso ao solicitar que produzisse o mesmo documento em .pdf, mas solicitando a mesma avaliação do aluno 1 formatada para o LaTeX obtivemos novamente um código para copiar, e colando podemos notar que a formatação estava funcionando corretamente, demonstrado na figura 13 a seguir.

Figura 13 – Código compilado no Overleaf



Fonte: Autor

Nesta etapa então testamos uma das maiores utilidades da ferramenta, a sua capacidade de exportação. Enquanto os testes envolvendo exportação para o Word sem a produção de um link para download se provaram relativamente complicados por envolver um código para ser utilizado em um arquivo em .xml no bloco de notas para só depois então passar para o Word, notamos também que ao solicitar que exportasse parte da conversa diretamente para um arquivo do Word para download a ferramenta faz a mesma funcionalidade com complicações menores visto que apenas necessitamos clicar num link e então abrir o arquivo baixado. A funcionalidade de exportação para download de um arquivo do Word funciona bem, ficando melhor formatada do que apenas copiando a conversa e colando diretamente no Word.

A exportação para LaTeX através do compilador de código também funciona muito bem, e é especialmente útil aos interessados em matemática e programação por permitir uma formatação melhor de equações matemáticas com o uso de programação para a digitação, e em casos de equívocos pela I.A., podemos facilmente corrigi-los solicitando-a que corrija ao mencionar o equívoco que gostaríamos que ela mudasse, resolvendo rapidamente o problema. No entanto, nestes testes, não notamos equívocos nas exportações, além do possível limite de caracteres que mencionamos ao exportar um arquivo para download, e da falha ao exportar para o Word através de código para utilizar arquivos em .xml, e, portanto, recomendamos que utilizem a opção de solicitação de arquivos para download de

pequenos trechos da conversa se possível, ou então a exportação para LaTeX. Isso, claro, ficará a critério do usuário e dependerá de suas necessidades.

4.5 PROGRAMAÇÃO

Também foi testado a capacidade da I.A. de produzir códigos na linguagem Python ou C, por exemplo, que resolvessem uma equação de primeiro grau. Os resultados foram, novamente, satisfatórios, visto que a máquina produziu códigos funcionais, explicando também como o código funciona e dando exemplos de execução. No entanto, é notável que, como foi solicitado primeiramente apenas que o código resolvesse uma equação simples, foi produzido um código que especificava os valores de a e b como sendo 2 e -4 respectivamente, e então resolvia esta equação específica. Ao solicitar, novamente, nos atentamos a isso e especificamos que o código deveria pedir que o usuário escolhesse os valores de a e b , desta vez obtendo o resultado desejado, o que nos indica algo a se atentar nos nossos pedidos à ferramenta, como na Figura 14.

Figura 14 – Código em Python corrigido pelo ChatGPT



Faça o mesmo código porém pedindo que o usuário escolha os valores de a e b

< 2/2 >

Aqui está o código em Python que solicita ao usuário que insira os valores de a e b para resolver a equação $ax + b = 0$:

```
python Copiar código

def resolver_equacao(a, b):
    if a == 0:
        if b == 0:
            return "A equação tem infinitas soluções."
        else:
            return "A equação não tem solução."
    else:
        x = -b / a
        return f"A solução da equação é x = {x}"

# Solicitar valores ao usuário
a = float(input("Digite o valor de a: "))
b = float(input("Digite o valor de b: "))

# Resolver a equação
resultado = resolver_equacao(a, b)
print(resultado)
```

Como funciona:

- O código solicita ao usuário que insira os valores de a e b usando a função `input()`.
- Esses valores são convertidos para `float` para permitir a entrada de números decimais.
- Em seguida, a função `resolver_equacao(a, b)` é chamada para resolver a equação e exibir o resultado.

Fonte: Autor

Sua capacidade de programação pode também ser utilizada em conjunto com a personificação mencionada no 4.3, por exemplo, ao solicitar que a ferramenta nos dê aulas sobre programação, explorando conceitos desde a base como os conceitos de estruturas de controle, definição de funções (no contexto de programação), interações com o usuário com entrada e saída de dados, e etc. além de podermos solicitar códigos mais complexos que os mencionados no exemplo. Com estas habilidades, sua versatilidade aumenta ainda mais, podendo nos auxiliar em códigos diversos, e até mesmo contribuindo com programas que consideremos úteis em nossas aulas, por exemplo utilizando a I.A. para produzir um código de um programa funcional de uma calculadora com as quatro operações básicas, demonstrando o passo a passo de sua produção, de modo a treinar também as habilidades de programação do estudante.

4.6 TRADUÇÕES

Finalmente, outras das funcionalidades da ferramenta que foram testadas durante esta pesquisa, foram a capacidade da máquina de receber o pedido em português, mas produzir uma avaliação em idiomas diferentes, como no caso do inglês, com resultados satisfatórios. Por outro lado, é importante também ressaltar que, em nossos testes, ao realizar uma solicitação não relacionada à tradução em seguida, por exemplo solicitando à ferramenta que produzisse um gráfico de uma função, notamos que a ferramenta manteve sua resposta em inglês, sendo assim necessário que solicitássemos na mensagem seguinte que retornasse ao português.

A ferramenta é capaz de produzir traduções em idiomas diversos, e é também capaz de combinar essas traduções com os pontos anteriores, por exemplo explicando em inglês sobre o funcionamento dos códigos de algum programa específico, ou então personificando o professor de algum idioma específico e dando aulas sobre conceitos daquele idioma.

Esta capacidade de tradução da ferramenta pode ser vantajosa caso já estivermos trabalhando em algo com o auxílio da I.A. e necessitarmos de alguma informação ou parte do nosso trabalho em outro idioma, por exemplo, poupando assim o nosso tempo de traduzir manualmente ou utilizando outro tradutor como o do Google. Não notamos em nossos testes desvantagens quanto ao uso desta função da ferramenta, com a única exceção mencionada acima sobre necessitarmos solicitar

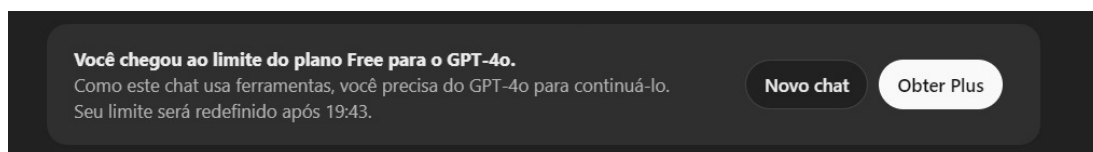
que a I.A. retornasse ao idioma anterior, sendo assim uma possibilidade de que em determinadas ocasiões ela trabalhe não com tradução de trechos específicos, mas sim com uma mudança de idioma da ferramenta como um todo, caso esse em que precisaríamos mudar o idioma novamente ao português, como foi em nosso teste anterior.

4.7 PERCEPÇÕES E LIMITAÇÕES

Nestes testes que realizamos e documentamos aqui, notamos diversas utilidades da ferramenta e, ao menos em nossa experiência, poucas desvantagens relacionadas diretamente ao seu uso. Algumas de suas utilidades, como a possibilidade de exportação para LaTeX, anteriormente desconhecidas por nós, acabaram até mesmo por serem úteis em outras situações pessoais neste período de pesquisa.

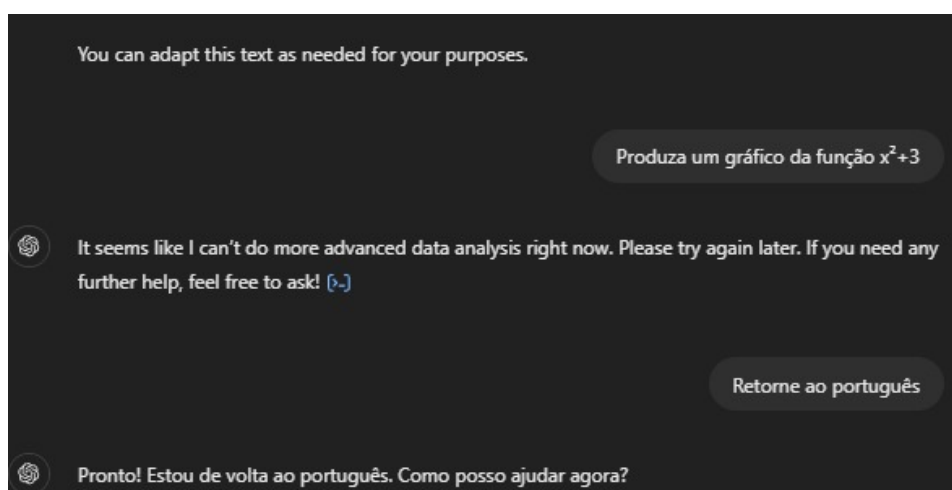
De modo geral, podemos notar que sua capacidade de auxílio é extremamente vasta, dependendo quase que somente por nossa criatividade e entendimento da ferramenta para sabermos melhor como comandá-la. Por suas habilidades variarem, não se limitam apenas à matemática, abrindo espaço também para utilizar de seu auxílio com outras áreas como testamos com traduções e programação. A ferramenta, num panorama geral, torna-se útil como apoio escolar e pessoal em diversas ocasiões, e, portanto, é relevante sabermos de suas capacidades e considerarmos o uso da mesma.

No entanto, é importante também mencionarmos suas limitações junto com suas capacidades. Achamos válido ressaltar que os testes foram feitos utilizando o GPT-4o em sua versão gratuita, que possui um limite. Este limite é redefinido em questão de horas, e caso atingido você tem as opções de esperar que o limite seja redefinido, adquirir o plano pago do ChatGPT, ou então abrir um novo chat sem a utilização das ferramentas do GPT-4o. Uma das ferramentas exclusivas do plano pago do GPT-4o, por exemplo, é a de produzir gráficos e transportá-los para tabelas, por exemplo.

Figura 15 – Notificação de limite do plano gratuito

Fonte: Autor

Temos também outras limitações, como mencionado no 4.4, por exemplo, onde notamos um possível limite de caracteres na exportação para download, e no 4.5, onde fomos informados ao testar a solicitação de produção de um gráfico que a ferramenta não poderia fazer mais análises de dados avançadas no momento, o que indica um possível limite temporário quanto à versão gratuita do ChatGPT sobre a análise de dados avançada, utilizada também para outras funções, como produzir documentos para download, por exemplo.

Figura 16 – Limite na análise de dados e respostas em inglês

Fonte: Autor

Ou seja, possuímos um limite geral do plano gratuito, e também um limite na função de análise de dados, utilizada em casos específicos como na exportação para LaTeX, produção de gráficos, e outras atividades que necessitem maior capacidade por parte da ferramenta. É também importante mencionar que é possível existirem outros limites não encontrados em nosso período de testes, e, portanto, não podemos considerar que estes limites aqui mencionados são os únicos existentes, apesar de não termos encontrado outros limites a se mencionar.

5 CONSIDERAÇÕES FINAIS

Tendo realizado esta pesquisa, podemos notar melhor o impacto que os avanços da inteligência artificial tem causado na área da educação. O ChatGPT é apenas uma de muitas ferramentas utilizando este tipo de tecnologia que com toda a notoriedade recebida, tornou-se uma das mais populares e, portanto, foi foco de muitas pesquisas. Este foco e notoriedade também fizeram com que tanto alunos quanto professores passassem a utilizar a ferramenta.

Sabendo disso, como no caso do primeiro texto que revisamos no capítulo 3, alguns professores começam a naturalmente considerar que seus alunos poderão utilizar a ferramenta, e por tais motivos mudam sua análise por vezes suspeitando que uma resposta boa não tenha sido realizada por um aluno, por exemplo. Ou então, como mencionado pelos pesquisadores que realizaram uma revisão sistemática de outros textos, os professores acabam considerando a ideia de contar com o auxílio do ChatGPT para produzir avaliações automaticamente, focando geralmente na avaliação formativa.

Podemos notar que a inteligência artificial tem sido amplamente utilizada, e constantemente se atualiza melhorando cada vez mais, e por isso consideramos ideal que os professores que a utilizarão saibam usá-la e, especialmente, saibam de suas capacidades e quais tarefas poderão agilizar ou automatizar com suporte de ferramentas como o ChatGPT, e daí a importância de guias com recomendações, como o que mencionamos anteriormente.

No geral, concluímos que a eficácia da ferramenta tem sido melhorada no período em que esta pesquisa foi realizada, tendo sido iniciada em 2023, e finalizando em 2024, com a aparição do GPT-4o que conta com plano gratuito e algumas utilidades como a capacidade de produzir arquivos para download e analisar dados, e sua eficácia deve apenas melhorar mais ainda com o passar dos anos, desta forma se tornando uma aliada poderosa em nosso repertório.

Podemos mostrar, dentre elas, a capacidade de tradução, programação, resolução de problemas, produção de avaliações e entendimento de cenários hipotéticos. Com tantas capacidades, suas diversas utilidades apelam a um público-alvo extenso, e devemos aproveitar para integrar em nosso ensino regras para o uso dessa ferramenta, guias para facilitar o uso de maneira eficiente da mesma, de modo que incentivemos nossos estudantes a tomar um caminho mais produtivo com o

ChatGPT como parceiro de estudos, por exemplo solicitando explicações passo a passo de como criar determinado algoritmo, listas de exercícios para estudos, correção destes mesmos exercícios para obter uma devolutiva de suas respostas, e quem sabe com isso, gerar um maior interesse por parte dos alunos no aprendizado de modo geral.

Tendo demonstrado suas possibilidades de uso e algumas pesquisas tratando do assunto, finalizamos com resultados gerais satisfatórios em nossos testes, comprovando sua versatilidade ao atender nossas solicitações, agora expostas nesta pesquisa para servir de referência e incentivo para estudo sobre, bem como propriamente o uso, do ChatGPT.

E, para finalizar, ressaltamos a importância do letramento em I.A. tanto por professores quanto estudantes. Sendo uma ferramenta tão poderosa e útil, ela é capaz de nos prestar auxílio em diversas áreas inclusive as que envolvem a matemática e a licenciatura, como na avaliação. No entanto, a ferramenta depende de nossos comandos para agir, e como demonstrado no 4.5, pode apresentar falhas de acordo com a maneira que a comandarmos. Por tais motivos, torna-se crucial o aprendizado e entendimento da ferramenta, do mesmo modo que necessitamos aprender a utilizar corretamente uma calculadora para atingirmos suas capacidades máximas.

[...] não é o plágio de uma atividade entregue pelo estudante e supostamente redigida com o ChatGPT que deve preocupar, nos espaços educacionais. O que deve preocupar é o tipo de atividade/avaliação que segue sendo atribuído aos estudantes desde sempre, mas que agora urgem serem redimensionados e rediscutidos para as novas habilidades e competências, exigidas para a Era educacional. (Ricieri et al, 2024, p. 14)

Existem já artigos tratando sobre este letramento, e como dito no estudo de Ricieri, Farias, Barreto e Souza, uma das competências importantes a serem tratadas neste processo é a de produzir melhores questionamentos, e isso pode ser aplicado tanto à questão avaliativa se adaptando à existência da I.A. para evitar plágios quanto, principalmente, ao método de uso da ferramenta. Produzir questões melhores afetam diretamente o funcionamento da ferramenta que trabalha diretamente com seus questionamentos, agindo de acordo para respondê-los e, como demonstrado ao decorrer do capítulo 4 deste documento, quanto maior o contexto que a ferramenta tiver, melhores serão os resultados que iremos obter. Esta importância do letramento serve também para os estudantes, que em meio a tantas possibilidades, poderão

utilizar a I.A. para aprender mais, e por ser de funcionamento simples, por vezes acaba despertando maior interesse de aprendizado neles, ressaltando então a relevância de professores que instiguem seus estudantes a passar por este mesmo letramento, talvez até mesmo solicitando que os alunos utilizem a ferramenta em ocasiões determinadas, ou então solicitando revisões à ferramenta.

REFERÊNCIAS BIBLIOGRÁFICAS

BLOOM, B. et al. Manual de Avaliação Formativa e Somativa do Aprendizado Escolar, trad. **Lilian Rochlitz Quintão**, 1983.

GONZÁLEZ-CALATAYUD, Víctor; PRENDES-ESPINOSA, Paz; ROIG-VILA, Rosabel. Artificial intelligence for student assessment: A systematic review. **Applied sciences**, v. 11, n. 12, p. 5467, 2021.

CARVALHO, Felipe; PIMENTEL, Mariano. Estudar e aprender com o ChatGPT. **Revista Educação e Cultura Contemporânea**, v. 20, 2023.

FARAZOULI, Alexandra et al. Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. **Assessment & Evaluation in Higher Education**, v. 49, n. 3, p. 363-375, 2024.

FERNANDES, Domingos. Avaliação das aprendizagens: reflectir, agir e transformar. In: **Livro do 3.º Congresso Internacional Sobre Avaliação na Educação**. Futuro Eventos, 2005. p. 65-78.

GIMPEL, Henner et al. **Unlocking the power of generative AI models and systems such as GPT-4 and ChatGPT for higher education: A guide for students and lecturers**. Hohenheim Discussion Papers in Business, Economics and Social Sciences, 2023.

GOOGLE CLOUD. **O que é inteligência artificial (IA)?** Disponível em: <https://cloud.google.com/learn/what-is-artificial-intelligence?hl=pt-BR>. Acesso em: 04 set. 2024.

HAYDT, Regina Célia Cazaux. **Avaliação do processo ensino-aprendizagem**. Editora Ática, 1995.

HAYDT, Regina Célia Cazaux. **Curso de didática geral**. 1ª Edição-São Paulo: Ática, 2011.

LUCKESI, Cipriano Carlos. **Avaliação da aprendizagem escolar: estudos e proposições**. Cortez editora, 2014.

E OLIVEIRA, João Batista Araujo; CHADWICK, Clifton. **Aprender e ensinar**. Instituto Alfa Educativca, 2004.

DA VINHA RICIERI, Denise et al. Erros comuns de docentes sem letramento em Inteligência Artificial: uma revisão integrativa para o ensino superior. **Peer Review**, v. 6, n. 7, p. 284-300, 2024.

ZONG, Mingyu; KRISHNAMACHARI, Bhaskar. Solving math word problems concerning systems of equations with gpt-3. In: **Proceedings of the AAAI Conference on Artificial Intelligence**. 2023. p. 15972-15979.