

Homem ou Máquina? Um estudo exploratório do desempenho do Chat GPT 3.5 no Exame de Suficiência do CFC

JOSÉ CARLOS RAMOS DE OLIVEIRA JUNIOR
FACULDADE FIPECAFI (FIPECAFI)

AHMED SAMEER EL KHATIB
CENTRO UNIVERSITÁRIO ÁLVARES PENTEADO (FECAP)

Homem ou Máquina? Um estudo exploratório do desempenho do *Chat GPT* 3.5 no Exame de Suficiência do CFC

1 Introdução

Com um cenário de evolução constante à frente, a IA é uma disciplina da ciência da computação que tem como objetivo a reprodução das reações da mente humana, incluindo habilidades como reconhecimento de fala, pensamento lógico e aprendizado (Madhavi & Kumar, 2021).

Em 1950, no seminal “*Computing Machinery and Intelligence*”, Alan Turing propôs um teste que parte do pressuposto de que se uma pessoa, ao interagir remotamente com um computador, não conseguir dizer se é um computador ou outra pessoa, o sistema é descrito como inteligente (George & S. Gillis, 2023). Alçado em 1955 por John McCarthy, Nathaniel Rochester, Claude Shannon e Marvin Minsky, o conceito de *Artificial Intelligence*, em português, Inteligência Artificial (IA), foi descrito em um documento que propôs a criação de uma máquina com capacidade cognitiva similar à humana (Garattoni, 2023).

A IA generativa é um campo de inteligência artificial que se concentra na geração de informações novas e originais por aprendizado de máquina em enormes bancos de dados. Existem vários modelos de linguagem e de IA disponíveis na internet. Porém o ChatGPT (sigla para “*Generative Pre-Trained Transformer Chatbot*”) ganhou notoriedade por ter alcançado mais de um milhão de assinantes na primeira semana após sua estreia no mundo e tem aumentado sua popularidade com grande entusiasmo e expectativa (Baidoo-Anu & Ansah, 2023; Garattoni, 2023).

O ChatGPT pode ajudar na análise de dados, examinando rapidamente grandes quantidades de dados financeiros e gerando *insights* significativos para ajudar os profissionais da contabilidade e usuários das demonstrações financeiras na tomada decisões, economizando um tempo valioso e aumentando a produtividade (Alshurafat, 2023; Guo, 2019).

A avaliação mais abrangente da prontidão de capacidade para trabalhadores do conhecimento profissional contábil no Brasil é o Exame de suficiência do Conselho Federal e Contabilidade (CFC), que é constituído de prova objetiva para Bacharéis em Ciências Contábeis. Segundo a Resolução CFC nº 1.486/15, o exame de suficiência é uma prova com o objetivo de comprovar a obtenção de conhecimentos médios obtido no curso de Ciências Contábeis e por meio dele, é possível que o bacharel e estudantes do último ano letivo do curso de Ciências Contábeis obtenha o registro na categoria Contador. Desta forma, o exame torna-se um requisito necessário e obrigatório para a obtenção do registro profissional em um Conselho Regional de Contabilidade (CRC) (Conselho Federal de Contabilidade, 2015).

Cabe ressaltar que, para poder atuar como responsável técnico na auditoria de entidades com atividades reguladas, o auditor deve prestar exame e estar habilitado no Cadastro Nacional de Auditores Independentes (CNAI) de Qualificação Técnica Geral, que confere ao contador o reconhecimento de capacitação geral para atuação em atividades de Auditoria Independente e qualificação técnica específica para a sua área de atuação (Conselho Federal de Contabilidade, 2020).

De acordo com Aydın e Karaarslan (2023), alguns autores limitaram suas análises do ChatGPT a apenas perguntas e respostas, enquanto outros consideraram os possíveis efeitos e fizeram avaliações mais abrangentes. As áreas de estudo dessas análises incluem educação geral, escrita acadêmica, escrita em geral, saúde, direito, economia e finanças.

Ante ao exposto, este trabalho pretende responder à questão de pesquisa: Qual o nível de precisão do *chatbot* para responder a questões dos exames de suficiência e de qualificação técnica do CFC?

Para atingir o objetivo geral, têm-se os seguintes objetivos específicos: i) abordar brevemente a definição das pesquisas recentes que tratam da utilização da tecnologia de IA; ii) apresentar pesquisas recente encontradas sobre utilização do ChatGPT e seu modelo de PLN; iii) demonstrar o desempenho do ChatGPT 3.5 ao responder as questões do exame de suficiência e de qualificação técnica do CFC; iv) demonstrar o desempenho do ChatGPT na produção das redações para o exame de qualificação técnica do CFC v) comparar o desempenho entre as versões 3 e 3.5 do *Chat GPT* no exame de suficiência; e vi) explorar as limitações e os principais desafios na utilização e no desenvolvimento deste tipo de tecnologia para o profissional de contabilidade.

Isto posto, esta pesquisa tem o objetivo de explorar o grau de assertividade do *chatbot* no 1º Exame de Suficiência de 2022 e no 1º Exame de Qualificação Técnica para registro no Cadastro Nacional de Auditores Independentes (CNAI) de 2023. O trabalho foi dividido, além desta introdução, em mais quatro seções principais, a saber: Referencial Teórico, Metodologia, Análise dos Resultados e Considerações Finais.

2 Referencial teórico

2.1 Chat GPT

A IA é um ramo abrangente da ciência da computação preocupado com a construção de máquinas inteligentes capazes de racionalizar, tomar ações que tenham as melhores chances de atingir um objetivo específico de realizar tarefas que normalmente requerem inteligência humana e pode ser descrita como uma simulação da inteligência humana em máquinas que são programadas para pensar e agir como humanos. Podendo aprender automaticamente e melhorar seu desempenho em uma tarefa específica (Frankenfield & Scott, 2023; Schroer, 2023).

Como esses poderosos sistemas de IA são construídos e usados será muito importante para o futuro do nosso mundo e de nossas próprias vidas. Todas as tecnologias têm consequências positivas e negativas, mas com a IA, o alcance dessas consequências é extraordinariamente grande: a tecnologia tem imenso potencial para o bem. Ainda assim, vem com grandes desvantagens e altos riscos (Giattino et al., 2022).

Os sistemas de IA podem executar tarefas comuns associadas às funções cognitivas humanas, como interpretar a fala, jogar e identificar padrões processando grandes quantidades de dados para modelar em seus processos de tomada de decisão e possui diversas aplicações em vários campos, como: saúde, finanças, transporte e entretenimento (Copeland, 2023; Schroer, 2023).

O *ChatGPT* foi treinado com um grande conjunto de dados abertos disponíveis e que permite criar uma linguagem coerente, contextualmente apropriada, com uma habilidade verbal impressionante e muitas vezes indistinguível de coisas escritas por humanos. (Alshurafat, 2023; Aydin & Karaarslan, 2023; Lopes, 2022). Após diversos treinamentos, o ChatGPT foi refinado e ajustado para melhorar a qualidade de suas respostas e minimizar erros. Desde então, o modelo tem sido usado em uma variedade de aplicativos, incluindo assistentes virtuais, *chatbots* e outras interfaces de conversação, para fornecer respostas precisas e naturais a perguntas e outras entradas de texto. Segundo o site da OpenAI, modelo interage de forma conversacional com uma interface de bate papo o que permite ao *Chatbot*

responder a perguntas, admitir seus erros, contestar premissas incorretas e rejeitar solicitações inadequadas. O ChatGPT é um modelo de linguagem de última geração que utiliza técnicas de aprendizado profundo para gerar texto humano (Hutanu, 2023). É um produto da OpenAI, uma organização de pesquisa dedicada ao avanço da inteligência artificial e ao desenvolvimento de tecnologias de ponta que beneficiam a sociedade. O núcleo da arquitetura do ChatGPT é um *Transformer*, uma arquitetura de rede neural que permite ao modelo analisar sequências de dados, como, por exemplo, entradas de texto. O *Transformer* foi introduzido em 2017 e, desde então, revolucionou o campo do processamento de linguagem natural (PLN) (Heaven, 2023).

Uma das principais vantagens do ChatGPT é sua capacidade de ser ajustado para uma ampla gama de tarefas relacionadas ao idioma. Foi demonstrado que modelos de linguagem bem ajustados podem ser aprendizes contínuos, entregando uma indicação das capacidades futuras de inovações como o ChatGPT (Scialom et al., 2022). Ao treinar o modelo em uma tarefa específica com dados adicionais, como classificação de texto ou tradução automática, o modelo pode se adaptar a novos domínios e ter um bom desempenho em várias tarefas de PLN.

O processo de pré-treinamento envolve prever a próxima palavra em uma frase, dadas as palavras anteriores, permitindo que o modelo desenvolva uma compreensão profunda da estrutura e do significado da linguagem, que é então usada para gerar um texto coerente e contextualmente relevante (Pavlik, 2023).

Com o desenvolvimento de modelos de linguagem mais avançados, como o ChatGPT, espera-se que o futuro da PLN traga maneiras mais inovadoras e eficientes de se comunicar e interagir com as máquinas. Até o momento, a OpenAI lançou várias versões do ChatGPT, como GPT-2, GPT-3 e GPT-4 (a partir de 14 de março de 2023), que diferem em tamanho, número de parâmetros e número de idiomas incluídos no seus dados pré-treinamento. Por exemplo, o GPT-3, que conquistou o público, foi treinado em um conjunto diversificado de textos em 95 idiomas e com 175 bilhões de parâmetros (Farseev, 2023).

2.2 Estudos sobre aplicações do ChatGPT

De acordo com o relatório *Artificial Intelligence Index Report 2023* publicado pela universidade de Stanford, de 2010 até 2021, o número de publicações sobre AI aumentou de 200 mil para 500 mil. Sendo que em 2021, 60% de todos os documentos de IA publicados foram em artigos de periódicos, 17% em artigos de conferências e 13% em envios de repositórios (Maslej et al., 2023).

A literatura recente contém diversos estudos acerca do ChatGPT, e o número dessas publicações cresceu, especialmente no final de 2022 e início de 2023, quando o modelo passou a ser utilizado gratuitamente por uma ampla base de usuários. (Aydin & Karaarslan, 2023) afirmam que o ChatGPT tem a capacidade de gerar textos em qualquer área de conhecimento e assunto, a partir de um *prompt* - termo utilizado em informática que se refere a um sinal que indica ao usuário que o sistema está pronto para receber uma entrada - fornecido pelo usuário. Sua estrutura modular possibilita a execução de diversas tarefas, como responder perguntas, traduzir idiomas, modelar linguagem, dentre outras. Além disso, o ChatGPT pode ser utilizado no desenvolvimento de assistentes virtuais e *chatbots*, bem como na geração de diálogos semelhantes aos humanos.

O ChatGPT foi desenvolvido com uma técnica chamada *Reinforcement Learning Human Feedback* (Aprendizagem por Reforço com Feedback Humano, em tradução livre) para treinar o modelo de linguagem. Todavia, como afirma o site da

OpenAI, “o ChatGPT às vezes escreve respostas que parecem plausíveis, mas incorretas ou sem sentido”. Vários exemplos mostram erros gritantes que ele pode cometer, inclusive referenciando um estudo científico que não existe (Thorp, 2023).

Bommarito et al. (2023), avaliaram a capacidade do modelo Text-DaVinci-003 da OpenAI e versões anteriores do GPT para avaliação dos exames aplicados pelo AICPA. Os resultados obtidos pelos autores da pesquisa demonstraram um desempenho inferior em raciocínio numérico em *prompts* não treinados, mas se aproximou do desempenho humano nos níveis de habilidade “lembrar e compreender” e “aplicação”. Segundo os autores, as últimas gerações do GPT-3 apresentaram uma melhora significativa, passando de 30% no Text-DaVinci-001 para 57% no Text-DaVinci-003.

Na seção de múltipla escolha do *Bar Exam* (exame de qualificação profissional aplicado nos Estados Unidos e em alguns outros países para avaliar a qualificação de um indivíduo para exercer a advocacia), Bommarito e Katz (2022) testaram o modelo GPT-3.5 da OpenAI e concluíram que, apesar de pessoas com 7 anos de formados no ensino superior e treinamento específico responderem apenas 68% das perguntas corretamente, GPT-3.5 obteve uma taxa de precisão de 50,3% e que a tendência atual de melhoria do modelo GPT prevê que um modelo de linguagem possa passar no *Bar Exam* em breve.

Na área das finanças, Dowling e Lucey (2023) concluem em seu estudo que o ChatGPT pode contribuir significativamente para o desenvolvimento da pesquisa financeira. Os pesquisadores afirmaram que o *chatbot* tem desempenho satisfatório em termos de geração de ideias e identificação de dados, mas não alcança a mesma eficiência na síntese literária e na construção de estruturas de testes. Zaremba e Demir (2023) afirmam que o ChatGPT tem potencial para melhorar as aplicações financeiras, mas também apresenta preocupações éticas e regulatórias.

Com o objetivo de investigar o aumento do uso da IA na governança corporativa, categorizando seus níveis de autonomia e explorando os papéis potenciais da IA em vários tipos de organizações, Mertens (2023) afirma em seu artigo que as aplicações conhecidas de inteligência artificial, como o ChatGPT, são apenas alguns exemplos de como a IA pode ser usada nos negócios. O autor analisa as possíveis aplicações da IA no mundo corporativo, especificamente na governança corporativa e afirma que a IA pode melhorar a tomada de decisões e desafiar o pensamento de grupo, mas os quadros legais das empresas ainda consideram a tomada de decisões exclusivamente humana, o que causa incertezas sobre a adoção da IA no ambiente corporativo.

Alshater (2023) investigou a aplicação da inteligência artificial, mais especificamente o processamento de linguagem natural, na melhoria do desempenho acadêmico, usando a economia e as finanças como exemplos e concluiu que ela tem o potencial de melhorar a pesquisa em economia e finanças. No entanto, ele também observou limitações, como a dependência da qualidade e variedade dos dados, a falta de experiência, a compreensão limitada do contexto e considerações éticas, que devem ser levadas em consideração ao utilizar ferramentas de IA, como ChatGPT e que precisam ser utilizadas em conjunto com a análise humana.

No Brasil, Nunes et al. (2023) exploraram a capacidade dos *Chat GPT 3.5* e *4.0* na resolução de testes de múltipla escolha no Exame Nacional do Ensino Médio (ENEM) e concluíram que, na edição de 2022, o modelo GPT-4 alcançou uma precisão de 87%, superando em grande parte o GPT-3.5, que obteve em 76% de aproveitamento.

Em geral, esses estudos destacam tanto o potencial do ChatGPT quanto suas limitações e preocupações éticas, enfatizando a necessidade de mais pesquisas sobre segurança e ética, bem como a importância de regulamentação adequada para garantir a utilização responsável da tecnologia gerativa de linguagem.

3 Metodologia

Para alcançar os objetivos desta pesquisa, foram utilizados os modelos GPT-3 que se trata de um conjunto de modelos que podem entender e gerar linguagem natural e GPT 3.5, um conjunto de modelos que melhoram o GPT-3 e podem entender, bem como gerar linguagem natural ou código. Na condução desta pesquisa, todas as questões do 1º Exame de Suficiência de 2022 (Exame de Suficiência) do CFC foram respondidas pelo ChatGPT 3 e 3.5. As questões do 1º Exame de Exame de Qualificação Técnica de 2023 (EQT) foram respondidas pelo ChatGPT 3.5.

Todas as questões foram inseridas na página de acesso do ChatGPT no site <https://chat.openai.com/> com o objetivo de emular a utilização por um usuário sem grandes conhecimentos em linguagem de programação e para que este não precisasse interagir com a API por meio de solicitações HTTP de qualquer linguagem.

As perguntas foram inseridas no idioma português, pois o *chatbot* afirma que as perguntas podem ser inseridas em qualquer idioma.

3.1 Os exames do Conselho Federal de Contabilidade

Com a finalidade evitar qualquer viés em suas respostas e obtê-las sem treinamento prévio da IA, este artigo utilizou de questões e posteriores a 2021.

O 1º Exame de Suficiência de 2022 (Exame de Suficiência) que é composto de uma prova com 50 questões de múltipla escolha e com as condições e áreas de conhecimentos a seguir: i) Contabilidade Geral; ii) Contabilidade de Custos; iii) Contabilidade Aplicada ao Setor Público; iv) Contabilidade Gerencial; v) Controladoria; vi) Teoria da Contabilidade; vii) Legislação e Ética Profissional; viii) Princípios de Contabilidade e Normas Brasileiras de Contabilidade; ix) Auditoria Contábil; x) Perícia Contábil; xi) Noções de Direito; xii) Matemática Financeira e Estatística; e xiii) Língua Portuguesa. (Conselho Federal de Contabilidade, 2015).

Já para a análise do desempenho no exame de qualificação, este estudo se baseou nas provas do 1º Exame de Exame de Qualificação Técnica de 2023 que são compostas de 25 questões de múltipla escolha e 2 questões dissertativas. Sendo: i) a prova de Qualificação Técnica Geral (QTG), que exige conhecimentos do contador em Normas Brasileiras de Contabilidade, Técnicas e Profissionais, editadas pelo CFC, Legislação Societária e Língua Portuguesa Aplicada; ii) a prova específica para atuação do contador em auditoria independente nas instituições reguladas pela CVM, que exige conhecimentos em Legislação e Normas do Mercado de Capitais, além de outras normas pertinentes aos auditores independentes e Legislação e Normas da B3 S.A. - Brasil, Bolsa, Balcão, Governança Corporativa e segmentos especiais de listagem; iii) a prova específica para atuação do contador em auditoria independente nas instituições autorizadas a funcionar pelo BCB, que exige conhecimentos em legislação e normas aplicáveis a essas instituições, operações realizadas por elas, contabilidade de instituições financeiras e demais entidades autorizadas a funcionar pelo BCB e legislação e normas emanadas pelo Conselho Monetário Nacional (CMN); iv) a prova específica para atuação do contador em auditoria independente nas sociedades supervisionadas pela Susep, que exige conhecimentos em legislação e normas aplicáveis a essas sociedades, operações realizadas por elas e contabilidade de sociedades e demais entidades supervisionadas pela Susep; e v) a prova específica

para atuação do contador em auditoria independente nas entidades supervisionadas pela Previc, que exige conhecimentos em legislação e normas aplicáveis a essas entidades, operações realizadas por elas e contabilidade de entidades supervisionadas por esta autarquia.

Para o desenvolvimento deste trabalho, as questões de múltipla escolha foram inseridas no *chatbot* e solicitado que fosse apresentada uma única alternativa correta. Já para as questões dissertativas, foi solicitado que o ChatGPT elaborasse um texto nos parâmetros requeridos na questão.

Importante destacar que, diante das recomendações o *chatbot*, as questões dos Exames foram inseridas e formatadas no MS Word® e as tabelas presentes no exame para explicação das transações, ou eventos, foram tabuladas de forma lógica para inserção no ChatGPT.

3.1.1. Questões dissertativas - CNAI

Como explicado na seção 3.1, o EQT de 2023 tem 2 questões dissertativas em que as respostas são avaliadas levando em consideração o domínio do conteúdo dos temas abordados e da Língua Portuguesa, bem como a capacidade do examinando de expor os assuntos com clareza, concisão, precisão, coerência e objetividade, além de demonstrar adequação à norma exposta na questão. Além disso, a capacidade de organização do texto, análise e síntese dos fatos e ideias também são consideradas na avaliação (Conselho Federal de Contabilidade, 2023).

Desta forma, as questões dissertativas das provas do EQT serão avaliadas quanto a sua objetividade, precisão, coerência, coesão, plágio, gramática e lógica das frases e dos parágrafos. Para a verificação de plágio, será utilizado a ferramenta CopySpider disponível de forma gratuita no site <https://copyspider.com.br/main/pt-br> e apresenta as frases que são identificadas como plagiado e que já estejam na internet, juntamente com a porcentagem correspondente.

4 Análise dos resultados

Nesta seção são apresentadas as respostas obtidas pelo ChatGPT, o conteúdo dos dados coletados e os resultados obtidos. Os resultados serão apresentados de forma clara e concisa, com suporte de tabelas e outras visualizações dos dados.

4.1 Desempenho do ChatGPT no exame de Suficiência

A Tabela 1 apresenta o aproveitamento geral do ChatGPT ao responder as questões do Exame de Suficiência e do EQT.

Tabela 1
Aproveitamento geral

Questões	Exame de Suficiência ^a	Exame de Qualificação				
		QTG	CVM	BCB	SUSEP	PREVIC
Corretas	37	16	14	13	14	20
Incorretas	13	9	11	12	11	5
% Corretas	74%	64%	56%	52%	56%	80%
% Incorretas	26%	36%	44%	48%	44%	20%

Nota.^a Houve duas questões anuladas no gabarito final. Desta forma, foi atribuído como questão correta para cálculo do desempenho da IA, assim como seria atribuído a participantes desta edição.

A Tabela 2 apresenta o resumo do aproveitamento geral dos participantes do Exame de Suficiência e do ChatGPT por tipo de questão.

Tabela 2**Resumo do aproveitamento no Exame de Suficiência**

Exame Tipo	Questões	CFC ^a				ChatGPT			
		Acertos	%	Erros	%	Acertos	%	Erros	%
AUD	2	56.196	84,0%	10.673	16,0%	2	100,0%	0	0,0%
CASP	3	45.736	45,6%	54.486	54,4%	2	66,7%	1	33,3%
CC	4	39.098	29,3%	94.437	70,7%	2	50,0%	2	50,0%
CG	17	232.285	40,9%	335.513	59,1%	11	64,7%	6	35,3%
CGER	2	12.069	18,1%	54.683	81,9%	1	50,0%	1	50,0%
CTRL	1	16.188	48,5%	17.201	51,5%	0	0,0%	1	100,0%
LEP	4	79.784	59,7%	53.819	40,3%	4	100,0%	0	0,0%
PORT	2	15.120	22,6%	51.637	77,4%	2	100,0%	0	0,0%
MFE	2	11.886	17,8%	54.902	82,2%	2	100,0%	0	0,0%
DIR	3	46.218	46,1%	54.005	53,9%	3	100,0%	0	0,0%
PERCON	2	21.297	31,9%	45.475	68,1%	2	100,0%	0	0,0%
PCNBC	4	42.414	31,8%	90.960	68,2%	2	50,0%	2	50,0%
TC	4	88.385	66,1%	45.283	33,9%	4	100,0%	0	0,0%
Total	50	706.676	42,3%	963.074	57,7%	37	74,0%	13	26,0%

Nota. AUD = Auditoria Contábil; CASP = Contabilidade Aplicada ao Setor Público; CC = Contabilidade de Custos; CG = Contabilidade Geral; CGER = Contabilidade Gerencial; CTRL = Controladoria; LEP = Legislação e Ética Profissional; PORT = Língua Portuguesa Aplicada; MFE = Matemática Financeira e Estatística; DIR = Noções de Direito e Legislação Aplicada; PERCON = Perícia Contábil; PCNBC = Princípios de Contabilidade e Normas Brasileiras de Contabilidade; TC = Teoria da Contabilidade.

^a As informações relativas estatísticas do exame de suficiência foram extraídas do site do CFC (<https://drive.google.com/file/d/1g8f0mJIF-gwij--vTHhk3GulBqs86tF8/view>).

As questões 34 e 45 foram anuladas. Desta forma, foi atribuído como questão correta, assim como seria atribuído a participantes desta edição

De acordo com os dados apresentados nas Tabelas 1 e 2 é possível observar que em todos os exames, o ChatGPT obteve aproveitamento superior à 50%. No Exame de Suficiência, exceto no tema controladoria, o ChatGPT foi melhor em todas as demais nas áreas se comparado com os dados estatístico apurados pelo CFC. Com o aproveitamento de 74%, o *chatbot* teria passado no exame de suficiência. Já nos exames de Qualificação Técnica, o ChatGPT obteve aproveitamento de 64%, 56%, 52%, 56% e 80% nas certificações para o QTG, CVM, BCB, SUSEP e PREVIC, respectivamente.

Embora possam não ser estatísticas comparáveis em decorrência da diferença de metodologia aplicada nas pesquisas, diferença dos testes e nos modelos utilizados, esse número supera os acertos do ChatGPT obtidos na pesquisa de Bommarito et al. (2023) em que o modelo responde 57,6% das perguntas corretamente das questões do CPA, sendo possível observar como o aproveitamento da IA foi satisfatório na obtenção de respostas.

Todavia, embora as respostas do ChatGPT tenham sido insatisfatórias, elas mostram a promessa da IA e seu potencial para revolucionar a forma como acessamos, criamos e consumimos informações e serviços. Porém, há grandes preocupações regulatórias e éticas, como o papel dos trabalhadores do conhecimento e a possível exploração de dados gerados por IA (Perlman, 2022).

Nesse sentido, destaca-se que para todas as respostas incorretas que o ChatGPT apresentou nessa pesquisa, ele apresentou justificativas coerentes e embasamento ou referências corretas para a questão inserida na plataforma, evidenciando a possibilidade de indução ao erro do usuário conforme demonstrado por outros pesquisadores.

Segundo Alshurafat (2023), além da habilidade de entender e interpretar comandos com facilidade, tecnologias como as aplicadas no ChatGPT podem auxiliar na avaliação e organização das informações no momento do preparo das demonstrações financeiras e outras informações com avaliação, reduzindo o risco de erro humano e reduzindo o tempo demandado quando comparado com procedimentos manuais. Ainda assim, mesmo diante de tais facilidades, o autor apresenta o desafio de assegurar que as informações apresentadas pelo *chatbot* sejam precisas e consistentes e, desta forma, afirma que o *chatbot* precisaria de um treinamento com dados de alta qualidade e a habilidade de entender e responder a questões complexas.

4.2 Diferenças observadas entre o ChatGPT 3 e 3.5

De acordo com Oliveira Junior (no prelo), o ChatGPT obteve um aproveitamento de 58% no 1º Exame de Suficiência de 2022, tendo acertado 27 perguntas e obtido os melhores resultados nas áreas de AUD, CG, CGER, CTRL, LEP, DIR, PERCON e TC quando comparado as médias nacionais do Exame divulgadas pelo CFC. A Tabela 3 apresenta as diferenças no 1º Exame de Suficiência de 2022 entre os modelos GTP 3 e GTP 3.5.

Tabela 3
Diferenças entre os modelos no 1º Exame de Suficiência de 2022

Exame Tipo	Questões	GPT-3 ^a		GPT-3.5		Erros	%	Erros	%
		Acertos	%	Acertos	%				
AUD	2	2	100,0%	0	0,0%	2	100,0%	0	0,0%
CASP	3	1	33,3%	2	66,7%	2	66,7%	1	33,3%
CC	4	1	25,0%	3	75,0%	2	50,0%	2	50,0%
CG	17	11	64,7%	6	35,3%	11	64,7%	6	35,3%
CGER	2	2	100,0%	0	0,0%	1	50,0%	1	50,0%
CTRL	1	1	100,0%	0	0,0%	0	0,0%	1	100,0%
LEP	4	3	75,0%	1	25,0%	4	100,0%	0	0,0%
PORT	2	0	0,0%	2	100,0%	2	100,0%	0	0,0%
MFE	2	0	0,0%	2	100,0%	2	100,0%	0	0,0%
DIR	3	3	100,0%	0	0,0%	3	100,0%	0	0,0%
PERCON	2	1	50,0%	1	50,0%	2	100,0%	0	0,0%
PCNBC	4	1	25,0%	3	75,0%	2	50,0%	2	50,0%
TC	4	3	75,0%	1	25,0%	4	100,0%	0	0,0%
Total	50	29	58,0%	21	42,0%	37	74,0%	13	26,0%

Nota. ^a Oliveira Junior (2023).

Na Tabela 3 observa-se que, exceto nas áreas de Contabilidade Geral, Contabilidade Gerencial e Controladoria, o ChatGPT-3.5 obteve resultados superiores quando comparado ao ChatGPT-3. Isso demonstra uma melhora no desempenho da IA, inclusive para desempenhar análises mais profundas apresentadas na no Exame de Suficiência.

4.3 Questões dissertativas do EQT

Nesta seção será apresentada as respostas elaboradas pelo ChatGPT para as 2 questões dissertativas no EQT. Essas questões visam avaliar as habilidades do candidato, bem como do conhecimento dos temas abordados, do domínio da Língua Portuguesa, da capacidade de expressão clara, concisa, precisa, coerente e objetiva considerando a norma específica apresentada na questão. São fatores considerados

na avaliação também, a capacidade de organização do texto, análise e síntese das informações e ideias também (Conselho Federal de Contabilidade, 2023).

De acordo com Vianna (1976), as provas de redação são amplamente utilizadas para avaliar a capacidade de expressão escrita e possuem méritos inquestionáveis. No entanto, apesar de várias tentativas de solução, essas provas ainda apresentam algumas dificuldades que representam um desafio para a maioria dos especialistas. Além disso, a literatura brasileira sobre instrumentos de medidas educacionais é limitada e há uma escassez de estudos empíricos sobre esse tipo específico de prova, apesar de seu amplo uso no contexto educacional recente.

Um texto precisa fazer sentido e na sua totalidade ser coeso e ao ser escrito, é preciso atentar-se ao significado de das palavras para evitar os erros, as impropriedades na concatenação lógica das frases e dos parágrafos (Melo & Pagnan, 2008).

Isto posto, para esta pesquisa, será utilizado como critério para avaliação da qualidade da redação do *chatbot*, objetividade, precisão, coerência, coesão, plágio, gramática e lógica das frases e dos parágrafos. Cabe ressaltar que as questões exigem que a o texto seja elaborado considerando uma quantidade mínima de 15 linhas e máxima de 30 linha, mas essa exigência não será considerada. Uma vez que esse parâmetro envolve o tamanho e a forma como o candidato elabora o texto.

Após analisar os textos, constatou-se que todos eles apresentam objetividade, precisão, coerência, coesão, correção gramatical e lógica das frases e parágrafos. No entanto, foram identificadas algumas áreas que requerem melhorias. Na questão 26 na prova do CNAI QTG, houve uma dificuldade breve em compreender o trecho "O auditor deve analisar o trabalho prévio realizado...", pois não fica claro que o auditor só pode emitir um relatório de acordo com a NBC TA 810 quando for contratado para auditar as demonstrações completas. Além disso, na questão 26 do CNAI CVM, o último parágrafo não indica claramente o momento em que o texto se encaminha para uma conclusão, o que dificulta o entendimento e prejudica a fluidez do texto. Outra observação relevante é encontrada na questão 26 do CNAI BCB, onde o primeiro parágrafo da segunda questão descreve a opinião modificada e parágrafos de ênfase como elementos elaborados pela instituição financeira auditada e não elementos preparados pelo auditor independente na elaboração do relatório.

Já na questão 26 do CNAI SUSEP, o *chatbot* afirma que "Na comunicação à Susep, a entidade deve informar a intenção de emissão da dívida subordinada com antecedência mínima de 30 dias em relação à data prevista para a sua realização. Além disso, devem ser mencionadas duas características ou condições relevantes da emissão, como por exemplo o valor total da dívida subordinada a ser emitida e a taxa de juros a ser aplicada aos títulos". No entanto, essa afirmação não está presente na resolução citada na questão.

Na análise dos textos realizada pela ferramenta CopySpider, não identificou plágio nos textos elaborados pelo *chatbot*. Sendo que os termos com similaridades encontrados pela ferramenta fazem referência a normas ou ao exposto na própria questão das provas.

Pode-se inferir então que, apesar da criação de textos originais e com objetividade, precisão, coerência, coesão, correção gramatical e lógica das frases e parágrafos, esses pontos de fragilidade destacam a necessidade de realizar revisões e melhorias no texto elaborado pelo *chatbot* para garantir a clareza, a coesão e a precisão das informações apresentadas.

5 Considerações finais

Este artigo buscou abordar brevemente a definição de pesquisas recentes que tratam da utilização da tecnologia de IA, apresentar pesquisas recentes encontradas sobre utilização do ChatGPT e seu modelo de PLN, demonstrar o desempenho do ChatGPT 3.5 ao responder as questões do exame de suficiência e de qualificação técnica do CFC, comparar o desempenho entre as versões 3 e 3.5 do *Chat GPT* no exame de suficiência e explorar as limitações e os principais desafios na utilização e no desenvolvimento deste tipo de tecnologia para o profissional de contabilidade.

Constatou-se que o ChatGPT obteve aproveitamento superior de 50% para todos os exames a que foi submetido, o que é suficiente para a aprovação de um candidato nos exames de Suficiência e Qualificação Técnica. Porém, ressalta-se que outras diversas variáveis precisam ser analisadas para a apuração de um resultado preciso, como, por exemplo, a dificuldade das questões, grau de treinamento da IA para entendimento do assunto e treinamento prévio em área específica, já que o ChatGPT pode tratar de diversos assuntos e não apenas de assuntos específicos como contabilidade, economia ou finanças.

As descobertas mostram que os modelos de linguagem se aperfeiçoaram em um curto espaço temporal desde o lançamento mundial do ChatGPT e têm o potencial de melhorar a qualidade e eficiência no trabalho de conhecimento para o profissional da contabilidade e para a área contábil no geral.

Mas que, ainda assim, é importante ter uma abordagem cética ao avaliar o nível de precisão das respostas geradas por essas tecnologias e isso ocorrerá quando as empresas aplicarem efetivamente essas tecnologias em suas atividades principais ou quando empresas de consultoria e software de gestão, que já possuem essas tecnologias integradas, as oferecerem para esses setores

Apesar do cuidado para a condução desta pesquisa, observou-se que expor a IA as questões apresentadas nos exames e não a treinar com antecedência poderia ter levado os resultados de acertos e erros, a um nível de precisão diferentes dos resultados apurados nesta pesquisa, impossibilitando uma análise mais robusta da tecnologia para a contabilidade, bem como melhorar os índices de originalidade obtidos.

O ChatGPT tem potencial para melhorar as aplicações financeiras, mas também apresenta preocupações éticas e regulatórias e tem o potencial de melhorar a pesquisa em economia e finanças. Mas com limitações, como a dependência da qualidade e variedade dos dados, a falta de experiência, a compreensão limitada do contexto e considerações éticas, que devem ser levadas em consideração ao utilizar ferramentas de IA em conjunto com a análise humana (Alshater, 2023; Zaremba & Demir, 2023).

Esta pesquisa contribui trazendo conceitos e informações sobre a Inteligência Artificial e sobre sua ascensão nos últimos anos e serve de auxílio para que usuários de plataformas de IA, como estudantes, pesquisadores e profissionais das áreas de contabilidade, economia e finanças atentem-se ao uso de ferramentas e que sempre analisem seus *outputs*. Adicionalmente, contribui-se para discussões sobre as melhores formas de utilização da ferramenta, no apoio a tomadas de decisão e, especialmente, nas atividades de contabilidade e sobre o estado da arte desta importante tecnologia.

Para as pesquisas futuras, é recomendada a utilização de outros métodos na busca de explorar a utilização de tecnologias de IA, ou, replicar a proposta desta pesquisa, utilizando-se de outros procedimentos metodológico, como por exemplo, treinamento da IA, análise estatística sobre suas respostas e submeter os textos

obtidos a outras ferramentas de plágio, enriqueceria o campo de pesquisa na área da contabilidade.

Referências

- Alshater, M. (2023). *Exploring the Role of Artificial Intelligence in Enhancing Academic Performance: A Case Study of ChatGPT*. <https://ssrn.com/abstract=4312358>
- Alshurafat, C. H. (2023). *The usefulness and challenges of chatbots for accounting professionals: application on*. <https://ssrn.com/abstract=4345921>
- Aydin, Ö., & Karaarslan, E. (2023). *Is ChatGPT Leading Generative AI? What is Beyond Expectations?* APA. <https://www.researchgate.net/publication/367525655>
- Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *SSRN*. <https://ssrn.com/abstract=4337484>
- Bommarito, J., Bommarito, M., Katz, D. M., & Katz, J. (2023). *GPT as Knowledge Worker: A Zero-Shot Evaluation of (AI)CPA Capabilities*. <http://arxiv.org/abs/2301.04408>
- Bommarito, M. J., & Katz, D. M. (2022). GPT Takes the Bar Exam. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4314839>
- Conselho Federal de Contabilidade. (2015). *Regulamenta o Exame de Suficiência como requisito para obtenção de Registro Profissional em Conselho Regional de Contabilidade (CRC)*. https://www1.cfc.org.br/sisweb/SRE/docs/Res_1486.pdf
- Conselho Federal de Contabilidade. (2020). *NBC PA 13 (R3) – Exame de Qualificação Técnica Para Auditor*. [https://www1.cfc.org.br/sisweb/SRE/docs/NBCPA13\(R3\).pdf](https://www1.cfc.org.br/sisweb/SRE/docs/NBCPA13(R3).pdf)
- Conselho Federal de Contabilidade. (2023). *Editais EQT Auditoria e Perícia Nº 1/2023 - Exame de Qualificação Técnica Para Registro no Cadastro Nacional de Auditores Independentes (CNAI) e no Cadastro Nacional de Peritos Contábeis (CNPC) do Conselho Federal de Contabilidade*. www.cesgranrio.org.br
- Copeland, B. J. (2023). Artificial intelligence (AI) | Definition, Examples, Types, Applications, Companies, & Facts | Britannica. In *Encyclopedia Britannica*. <https://www.britannica.com/technology/artificial-intelligence>
- Dowling, M., & Lucey, B. (2023). ChatGPT for (Finance) research: The Bananarama Conjecture. *Finance Research Letters*, 53. <https://doi.org/10.1016/j.frl.2023.103662>
- Farseev, A. (2023). *Is Bigger Better? Why The ChatGPT Vs. GPT-3 Vs. GPT-4 “Battle” Is Just A Family Chat*. <https://www.forbes.com/sites/forbestechcouncil/2023/02/17/is-bigger-better-why-the-chatgpt-vs-gpt-3-vs-gpt-4-battle-is-just-a-family-chat/?sh=63b43e275b65>
- Frankenfield, J., & Scott, G. (2023). *Artificial Intelligence: What It Is and How It Is Used*. <https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp>
- Garattoni, B. (2023). *O futuro da inteligência artificial – e o que vem depois do ChatGPT | Super*. <https://super.abril.com.br/tecnologia/o-futuro-da-inteligencia-artificial-e-o-que-vem-depois-do-chatgpt>
- George, B. St., & S. Gillis, A. (2023). *What is the Turing Test? | Definition from TechTarget*. <https://www.techtarget.com/searchenterpriseai/definition/Turing-test>
- Giattino, C., Mathieu, E., Broden, J., & Roser, M. (2022). Artificial Intelligence. *Our World in Data*. <https://ourworldindata.org/artificial-intelligence>
- Guo, X. (2019). Research on the Transition from Financial Accounting to Management Accounting under the Background of Artificial Intelligence. *Journal of Physics: Conference Series*, 1345(4). <https://doi.org/10.1088/1742-6596/1345/4/042031>
- Heaven, W. D. (2023). *ChatGPT is everywhere. Here’s where it came from | MIT Technology Review*. <https://www.technologyreview.com/2023/02/08/1068068/chatgpt-is-everywhere-heres-where-it-came-from/>
- Hutanu, A. (2023). *How ChatGPT works and AI, ML & NLP Fundamentals | Pentalog*. <https://www.pentalog.com/blog/tech-trends/chatgpt-fundamentals/>

- Lopes, A. (2022). *ChatGPT: conheça a inteligência artificial capaz de escrever artigos, roteiros e músicas | Exame*. <https://exame.com/tecnologia/chatgpt-conheca-a-inteligencia-artificial-capaz-de-escrever-artigos-roteiros-e-musicas/>
- Madhavi, S., & Kumar, V. (2021). QUALITATIVE REVIEW OF LITERATURE FOR ARTIFICIAL INTELLIGENCE IN TODAY'S BUSINESS SCENARIO. *Perspectives on Business Management & Economics*, IV, 66–76. www.pbme.in
- Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Parli, V., Shoham, Y., Wald, R., Clark, J., & Perrault, R. (2023). *Artificial Intelligence Index Report 2023 Introduction to the AI Index Report 2023*. https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf
- Melo, L. R. D. de, & Pagnan, C. L. (2008). *Prática de Texto: Leitura e redação* (3ª Edição). A!
- Mertens, F. (2023). *THE USE OF ARTIFICIAL INTELLIGENCE IN CORPORATE DECISION-MAKING AT BOARD LEVEL: A PRELIMINARY LEGAL ANALYSIS WP 2023-01 Working Paper Series The use of artificial intelligence in corporate decision-making at board level: A preliminary legal analysis*. <https://ssrn.com/abstract=4339413>
- Nunes, D., Primi, R., Pires, R., Lotufo, R., & Nogueira, R. (2023). *Evaluating GPT-3.5 and GPT-4 Models on Brazilian University Admission Exams*. <https://arxiv.org/abs/2303.17003v1>
- Oliveira Junior, J. C. R. de. (n.d.). ChatGPT: Desempenho do chatbot no exame de suficiência do Conselho Federal de Contabilidade. *Revista Finanças Fecap*.
- Pavlik, J. V. (2023). Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *https://doi.org/10.1177/10776958221149577*, 78(1), 84–93. <https://doi.org/10.1177/10776958221149577>
- Perlman, A. (2022). *THE IMPLICATIONS OF CHATGPT FOR LEGAL SERVICES AND SOCIETY*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4294197
- Schroer, A. (2023). *Artificial Intelligence (AI): What Is AI and How Does It Work? | Built In*. <https://builtin.com/artificial-intelligence>
- Scialom, T., Chakrabarty, T., & Muresan, S. (2022). Fine-tuned Language Models are Continual Learners. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022*, 6107–6122. <https://arxiv.org/abs/2205.12393v4>
- Thorp, H. H. (2023). ChatGPT is fun, but not an author. In *Science* (Vol. 379, Issue 6630, p. 313). American Association for the Advancement of Science. <https://doi.org/10.1126/science.adg7879>
- Vianna, H. M. (1976). Flutuações de julgamentos em provas de redação. *Cadernos de Pesquisa*, 19, 5–9.
- Zaremba, A., & Demir, E. (2023). *ChatGPT: Unlocking the Future of NLP in Finance* ChatGPT, OpenAI*. <https://ssrn.com/abstract=4323643>