

Trabalho de Conclusão de Curso (2024)

PERSONA ESPECIALISTA EM CONTABILIDADE: DESEMPENHO DA INTELIGÊNCIA ARTIFICIAL EM QUESTÕES DO EXAME DE SUFICIÊNCIA

Kewerson Alves Cunha¹, Kléber Formiga Miranda²

Resumo: Esta pesquisa teve como objetivo propor uma persona especialista em contabilidade, criada no modelo GPT-4 do ChatGPT, para responder questões do Exame de Suficiência em Contabilidade no Brasil. Para isso, treinou-se uma persona do ChatGPT denominada “[CFC Especialista](#)”, usando 289 questões das edições de 2020.1 a 2022.2. Durante o treinamento, as respostas incorretas fornecidas pelo modelo foram corrigidas imediatamente, visando maximizar seu desempenho nas questões de teste, que totalizaram 150 questões das edições de 2023.1 a 2024.1. Como resultado, o CFC Especialista obteve uma taxa média de acerto de 81% por conteúdo, com desempenho de 91% nas questões do tipo “Texto” (72 de 79) e 64% nas questões do tipo “Cálculo” (46 de 71). Dessa forma, este estudo não apenas contribui para pesquisas anteriores ao aprofundar a compreensão sobre o desempenho do ChatGPT nas questões dos Exames de Suficiência em Contabilidade por meio de um treinamento específico, mas também para a inovação do ensino e aprendizado com o uso da inteligência artificial na educação contábil.

Palavras-chave: ChatGPT; Chatbot; Contabilidade; Exame de Suficiência; Inteligência Artificial.

1. INTRODUÇÃO

As novas aplicações de ferramentas de Inteligência Artificial (IA), como o ChatGPT, não são meramente novos produtos de um mercado já estabelecido, mas representam uma verdadeira revolução, capaz de transformar completamente o modo de vida da humanidade (Ferreira-Veríssimo; Garcia-Maia; Ribeiro, 2023). Desde seu lançamento em 30 de novembro de 2022, o ChatGPT tem sido utilizado para diversas finalidades, como atendimento ao cliente, marketing e vendas, além de oferecer uma gama de serviços, desde responder perguntas básicas até concluir transações de comércio eletrônico (Sadikoğlu *et al.*, 2023).

Com foco em uma dessas finalidades, Choi *et al.* (2021) testaram o desempenho do GPT em quatro exames de direito, obtendo uma nota mediana (C+), porém suficiente para aprovação. O estudo revelou que o desempenho foi melhor nas questões dissertativas do que nas de múltipla escolha. Em outra análise, Kung *et al.* (2023) avaliaram o desempenho do ChatGPT no exame de licenciamento médico dos Estados Unidos. Sem qualquer intervenção humana, o modelo obteve uma taxa de aprovação média de 60%, demonstrando um alto nível de coerência e perspicácia em suas explicações. O estudo sugere que grandes modelos de linguagem, como o ChatGPT, têm o potencial de auxiliar não apenas na educação médica, mas também na tomada de decisões clínicas.

¹ Bacharelando em Ciências Contábeis.

² Doutor em Ciências Contábeis.

No cenário nacional, Leitão *et al.* (2023) analisaram o desempenho do ChatGPT, modelo GPT-3.5, em questões de radiologia do Colégio Brasileiro de Radiologia, obtendo uma taxa de acerto razoável de 53,3% nos exames de 2018, 2019 e 2022. Em outro estudo, Magalhães-Bicalho, Henrique-Oliveira e Guida (2024) testaram o GPT-3.5 em questões dos últimos cinco processos seletivos para residência médica, alcançando uma assertividade mediana de 55,5%. Esses pesquisadores recomendam cautela no uso de IA para responder questões clínicas em provas médicas. No campo da contabilidade, Freitas *et al.* (2023) aplicaram questões de quatro edições do Exame de Suficiência ao ChatGPT-4, alcançando um desempenho médio de 71,0% de acertos, superando significativamente a melhor unidade da federação, que variou entre 44,0% e 46,0% nos anos de 2021 e 2022. Oliveira Jr. e Khatib (2023) também avaliaram o desempenho do ChatGPT no Exame de Suficiência em Contabilidade, obtendo um aproveitamento de 74,0% na edição de 2022.1.

Diante das pesquisas na área contábil, Freitas *et al.* (2023) recomendam investigar se treinamentos específicos podem aprimorar o desempenho do ChatGPT em áreas mais desafiadoras para a IA. Assim, esta pesquisa busca desenvolver uma persona no chat da OpenAI com treinamento especializado em contabilidade, visando melhorar a precisão e clareza das respostas em questões do Exame de Suficiência. Dessa forma, a questão de pesquisa deste estudo é: Qual o desempenho de uma persona especialista em contabilidade, criada no modelo GPT-4 do ChatGPT, ao responder questões do Exame de Suficiência em Contabilidade no Brasil? Com isso, o presente artigo tem como objetivo propor uma persona especialista em contabilidade, criada no modelo GPT-4 do ChatGPT, para responder questões do Exame de Suficiência em Contabilidade no Brasil.

Para investigar se treinamentos específicos podem melhorar o desempenho do *chatbot* no Exame de Suficiência, foi desenvolvida uma persona especialista em contabilidade, treinada com 289 questões das edições de 2020 a 2022. Em seguida, essa persona foi testada com 150 questões das edições de 2023 a 2024.1. Os dados foram obtidos no sítio do Conselho Federal de Contabilidade (CFC). Com base nisso, os principais resultados indicam uma melhora significativa no desempenho do ChatGPT na resolução das questões do Exame de Suficiência de 2023 a 2024, alcançando uma taxa de acerto de 92% na última edição testada. Observou-se que o *chatbot* apresenta um desempenho ainda mais notável em questões com apenas interpretação textual, devido ao menor número de variáveis que a IA precisa processar em comparação com questões que envolvem cálculos, tabelas e interpretação textual.

No que se referem às contribuições deste estudo, a pesquisa busca avançar em relação aos estudos anteriores ao identificar e avaliar o aproveitamento do ChatGPT no Exame de Suficiência em Contabilidade, por meio do desenvolvimento de uma persona e da categorização das questões em dois tipos: “Cálculo” e “Texto”. Os resultados alcançados pelo modelo “[CFC Especialista](#)” indicam seu potencial como uma ferramenta de apoio no processo educacional da ciência contábil, auxiliando estudantes e profissionais da área na resolução de questões de provas e no aprimoramento de habilidades interpretativas e técnicas. Assim, esta pesquisa contribui para a inovação no uso da inteligência artificial na educação contábil, ao introduzir um produto que pode auxiliar no ensino e aprendizado dos profissionais da área.

2. REFERENCIAL TEÓRICO

2.1. ChatGPT

A Inteligência artificial (IA) é uma vasta área da ciência da computação que se dedica à criação de máquinas inteligentes, capazes de raciocinar, tomar decisões com maior probabilidade de alcançar um objetivo específico, e realizar tarefas que normalmente exigiriam inteligência humana (Oliveira Jr.; Khatib, 2023). Imerso nessa definição, surge o ChatGPT, *Chat-based Generative Pre-trained Transformer*, lançado em 30 de novembro de 2022 pela OpenAI. Para Oliveira Jr. *et al.* (2023), o GPT é o resultado do avanço de duas revoluções: a Inteligência Artificial e o Processamento de Linguagem Natural. Sob essa ótica, define-se o

ChatGPT como um *chatbot* baseado em Inteligência Artificial que utiliza Processamento de Linguagem Natural (*Natural Language Processing - NLP*) e aprendizado de máquina para simular diálogos humanos (Santhosh *et al.*, 2023).

No campo do Processamento de Linguagem Natural (NLP), Ayanouz-Soufyane, Abdelhakim-Anouar e Benhmed (2020) destacam que essa subárea da Inteligência Artificial é responsável por transformar dados em informações, permitindo que os *chatbots* compreendam mensagens de maneira eficiente. No caso do ChatGPT, a arquitetura responsável por codificar entradas (*inputs*), processar dados e gerar saídas (*outputs*) é chamada *Transformer* (Santhosh *et al.*, 2023). Proposta por Vaswani *et al.* (2017) em seu artigo “*Attention Is All You Need*”, essa arquitetura utiliza um mecanismo de auto atenção, permitindo ao modelo considerar todo o contexto de uma conversa para gerar respostas coerentes e relevantes. Originalmente aplicada à tradução de texto do inglês para o alemão, o *Transformer* superou significativamente os modelos baseados em camadas recorrentes ou convolucionais (Vaswani *et al.*, 2017). Com base nesses mecanismos, o ChatGPT visa, segundo Santhosh *et al.* (2023), possibilitar que os usuários façam perguntas de qualquer natureza e recebam respostas significativas, precisas e naturais, aplicáveis a atendimento ao cliente, automação de tarefas, fornecimento de *feedback*, entre outras utilizações.

Para obter respostas mais significativas, precisas e naturais, como mencionado por Santhosh *et al.* (2023), a criação de uma persona do ChatGPT pode ser uma abordagem eficaz. Nessa linha, o ChatGPT define o termo “*persona*” como sendo uma configuração específica que define um conjunto de habilidades e conhecimentos voltados para um propósito específico, permitindo que a IA gere respostas com maior precisão para atingir um objetivo determinado. Nesse sentido, será possível testar o foco desta pesquisa, que é investigar se treinamentos específicos podem melhorar o desempenho do ChatGPT no Exame de Suficiência em Contabilidade.

2.2. Exame de Suficiência

O Exame de Suficiência em Contabilidade no Brasil é um requisito obrigatório para que os graduados em Ciências Contábeis obtenham habilitação como contadores e possam exercer legalmente a profissão, conforme estabelecido pela Resolução nº 1486/2015 e pela Lei nº 12.249/2010. O exame foi implementado como resposta à proliferação de Instituições de Ensino Superior (IES) de baixa qualidade e à preocupação com a formação inadequada dos egressos para o exercício da profissão contábil (Rodrigues, 2018). Freitas *et al.* (2023) definem o Exame de Suficiência como uma prova que avalia o conhecimento mínimo necessário para que os bacharéis em Ciências Contábeis possam atuar de forma regular no Brasil. De acordo com o Capítulo II da Resolução nº 1486/2015, o exame é aplicado duas vezes ao ano em todo o território nacional, preferencialmente uma edição por semestre, com datas e horários divulgados pelo Conselho Federal de Contabilidade (CFC) ou por instituição contratada.

A prova consiste em 50 questões objetivas, cada uma valendo um ponto, conforme as diretrizes do CFC. As questões abrangem as seguintes áreas de conhecimento: Língua Portuguesa Aplicada (2), Matemática Financeira e Estatística (2), Noções de Direito e Legislação Aplicada (3), Legislação e Ética Profissional (4), Teoria da Contabilidade (4), Princípios de Contabilidade e Normas Brasileiras de Contabilidade (4), Contabilidade Geral (17), Contabilidade de Custos (4), Contabilidade Gerencial (2), Contabilidade Aplicada ao Setor Público (3), Controladoria (1), Auditoria Contábil (2) e Perícia Contábil (2). Cada questão de múltipla escolha oferece quatro alternativas (A, B, C e D), sendo apenas uma correta. Para ser aprovado, o candidato, seja graduado ou estudante no último ano do curso de Ciências Contábeis, deve obter pelo menos 50% dos pontos possíveis, ou seja, 25 questões.

Na segunda edição do exame em 2023, inscreveram-se cerca de 46.942 candidatos, 37.735 estiveram presentes, desse número, apenas 17,3% (6547 candidatos) foram aprovados (CFC, 2024).

Tabela 1 – Percentual de erros e acertos dos candidatos, por conteúdo do exame de 2023.2

Conteúdo	Quantidade	Acertos (%)	Erros (%)
Auditoria	2	26,2	73,8
Contabilidade Aplicada ao Setor Público	3	33,9	66,1
Contabilidade de Custos	4	77,1	22,9
Contabilidade Geral	17	33,6	66,4
Contabilidade Gerencial	2	31,2	68,8
Controladoria	1	38,1	61,9
Legislação e Ética Profissional	4	45,0	55,0
Língua Portuguesa Aplicada	2	61,4	38,6
Matemática Financeira e Estatística	2	63,0	37,0
Noções de Direito e Legislação	3	41,5	58,5
Perícia Contábil	2	27,5	72,5
Princípios Contábeis e NBC's	4	39,2	60,8
Teoria da Contabilidade	4	34,3	65,7

Nota: NBC's – Normas Brasileiras de Contabilidade.

Fonte: Conselho Federal de Contabilidade (2024).

A Tabela 1 apresenta o desempenho geral dos candidatos por conteúdo na segunda edição de 2023. Observa-se que os melhores resultados foram obtidos nas disciplinas de Contabilidade de Custos, Matemática Financeira e Estatística, e Língua Portuguesa. Em contrapartida, os piores desempenhos concentraram-se em Auditoria, Perícia Contábil e Contabilidade Gerencial. Essas disciplinas possuem uma característica em comum: a utilização das informações contábeis para subsidiar a tomada de decisões. Nesse sentido, o ChatGPT apresenta dificuldades nessa competência, conforme evidenciado pelos erros de interpretação e processamento apontados no estudo de Freitas *et al.* (2023).

2.3. Estudos Recentes

Com base na perspectiva de Santhosh *et al.* (2023), de que o ChatGPT visa permitir que os usuários façam perguntas de qualquer natureza, Oliveira Jr. e Khatib (2023) decidiram testar o desempenho do modelo GPT-3.5 no Exame de Suficiência em Contabilidade. A amostra do estudo incluiu 50 questões da primeira edição de 2022, das quais 48 foram efetivamente respondidas pelo *chatbot* e duas foram anuladas pela comissão organizadora do exame. Os autores consideraram essas questões anuladas como corretas para o ChatGPT, assim como ocorre para os candidatos da prova. Com isso, a IA alcançou um desempenho de 74%, acertando 37 das 50 questões. Verificou-se, portanto, que o ChatGPT superou a média dos candidatos daquela edição, obtendo mais de 50% de aproveitamento.

Seguindo a mesma linha de pesquisa, Freitas *et al.* (2023) também analisaram o desempenho do ChatGPT, utilizando a versão GPT-4, que é mais avançada do que a anterior (OpenAI, 2023). O estudo envolveu uma amostra de 200 questões das edições de 2021 e 2022 do Exame de Suficiência em Contabilidade. Os resultados obtidos variaram entre 64% e 78%, desempenho suficiente para a aprovação e habilitação no exercício da profissão contábil. As áreas em que a IA obteve maior taxa de acertos foram Língua Portuguesa Aplicada (88%), Noções de Direito e Legislação Aplicada (83%) e Ética Profissional (81%). Os autores destacam que essas disciplinas exigem predominantemente conhecimento teórico, o que facilita a compreensão e a resposta por parte do ChatGPT conforme Apêndice A. Com base nos resultados desse estudo, pode-se concluir que a IA superou

significativamente os candidatos da melhor federação do país, que obtiveram uma média de acerto de 44 a 46%.

No cenário internacional, Wood *et al.* (2023) avaliaram o desempenho do ChatGPT em um estudo abrangente que envolveu dados de 14 países e 186 instituições, totalizando uma amostra de 28.085 questões relacionadas à contabilidade e áreas afins. Os resultados indicaram uma taxa de acerto de 56,5%, com 47,5% de respostas totalmente corretas e 9% de respostas parcialmente corretas. O desempenho do ChatGPT foi mais notável em questões dicotômicas e de múltipla escolha, com taxas de acerto de 68,7% e 59,5%, respectivamente, considerando respostas totalmente e parcialmente corretas. No entanto, em comparação aos alunos, o ChatGPT obteve desempenho inferior, sendo superado pelos estudantes, que alcançaram uma média de 76,7% nas avaliações.

As pesquisas mencionadas anteriormente compartilham uma característica em comum: todas utilizam o ChatGPT nas suas versões disponíveis, seja GPT-3, GPT-3.5 ou GPT-4, sem considerar um treinamento específico em contabilidade para testar a acurácia da IA focando nos conteúdos mais relevantes da área contábil. Diante da lacuna existente na investigação sobre a eficácia de treinamentos específicos para melhorar o desempenho do ChatGPT em áreas mais desafiadoras da ciência contábil, conforme recomendado por Freitas *et al.* (2023), torna-se essencial examinar se uma persona treinada apresenta um desempenho superior em comparação com versões sem treinamento. Essa análise pode fornecer insights sobre a capacidade da IA de se adaptar e melhorar em contextos especializados, como o Exame de Suficiência em Contabilidade do Brasil.

3. PROCEDIMENTOS METODOLÓGICOS

Para criar uma persona especialista em contabilidade focada em responder questões do Exame de Suficiência em Contabilidade no Brasil, esta pesquisa foi estruturada em duas etapas: a configuração da persona no ChatGPT e a coleta de dados antes e depois de um treinamento com questões das edições de 2020 a 2022. Além disso, foram definidos critérios específicos para avaliar o desempenho do modelo.

No que se refere à configuração da persona, esta etapa é essencial para definir os objetivos que a persona deverá seguir. Neste estudo, o objetivo proposto foi o desenvolvimento de uma persona especialista em contabilidade, capaz de responder às questões do Exame de Suficiência. Para isso, os autores preencheram os campos solicitados na configuração, como Nome, Descrição, Instruções, Temas de Conversa e Conhecimento.

Quanto à coleta de dados por meio do treinamento, este procedimento é crucial para maximizar o desempenho da persona, uma vez que, ao receber *feedbacks* após responder incorretamente a uma questão, a persona pode assimilar as correções e, assim, melhorar seu desempenho nas questões subsequentes.

3.1. Configuração da persona

Para desenvolver um treinamento específico no ChatGPT, esta pesquisa seguiu os passos de configuração para criação de uma persona no chat da OpenAI. Nesse sentido os parâmetros de definições podem ser vistos através do quadro a seguir:

Quadro 1 – Configurações da persona no ChatGPT

Configuração	Descrição
Nome	Criada pelo projeto de pesquisa contabiliDados, visa responder com eficácia às questões do Exame de Suficiência em Contabilidade, oferecendo explicações precisas e detalhadas. Foca em identificar os principais conceitos contábeis e auxiliar na compreensão de normas e práticas contábeis essenciais.
Descrição	Desenvolvida pelo grupo de pesquisa ContabiliDados da Universidade Federal Rural do Semi-Árido (UFERSA), a persona foi criada com o objetivo de responder às questões do Exame de Suficiência em Contabilidade, que pode ser utilizada como ferramenta de apoio na preparação para o referido exame.

Instruções	O seu objetivo é responder às questões do Exame de Suficiência em Contabilidade com máxima eficácia possível, fornecendo explicações detalhadas e precisas, identificando os principais conceitos contábeis envolvidos, e auxiliando na compreensão das normas e práticas contábeis. Você responde às perguntas solicitadas sobre contabilidade e áreas relacionadas. Os principais tópicos que deve abordar são: controladoria, contabilidade geral, contabilidade de custos, contabilidade pública, contabilidade gerencial, direito público e privado, matemática financeira, teoria da contabilidade, legislação e ética profissional, princípios fundamentais de contabilidade e normas brasileiras de contabilidade, auditoria contábil, perícia contábil, português e conhecimentos sociais, econômicos e políticos do país. Deve utilizar o arquivo anexado na seção de "Conhecimento" para aprimorar seu desempenho. Será informado quando receber questões de treinamento e, dessa forma, deve respondê-las buscando aprimorar-se a cada interação. Caso cometa erros, será corrigido para evoluir. Sua comunicação deve adotar um tom formal moderado e de fácil compreensão, explicando os conceitos-chave que embasam os resultados apresentados. Deve se identificar como "CFC Especialista" e mencionar que foi criada pelo grupo de pesquisa chamado "contabiliDados" da Universidade Federal Rural do Semi-Árido (UFERSA). Busque identificar os padrões das perguntas que você mais acerta e analisar como pode ajudar professores e instituições de ensino, fornecendo feedback sobre sua assertividade. Você deve contribuir para uma nova dinâmica de ensino da ciência contábil no Brasil e entender como pode auxiliar na evolução do ensino de contabilidade no país. As questões a serem enviadas variam entre dois tipos principais: subjetivas e objetivas. As questões subjetivas exigem respostas dissertativas e abertas. Já as questões objetivas apresentam alternativas de escolha. Para essas últimas, deve desenvolver o raciocínio para justificar a resposta correta e selecionar explicitamente uma das opções com base na sua análise.
Temas de Conversa	Você poderia me ajudar com a questão a seguir? Por favor, detalhe os conceitos-chave abordados nesta questão. Qual é o tipo de questão que você prefere responder?
Conhecimento	Nesta seção, foi anexado um arquivo que detalha o conteúdo programático de cada assunto, conforme o segundo edital do Exame de Suficiência em Contabilidade de 2024, organizado pela Banca Fundação Getúlio Vargas.

Link de acesso: <https://chatgpt.com/g/g-ilCZqX61d-cfc-especialista>

Fonte: Elaboração própria

Essas definições são fundamentais para alcançar o objetivo desta pesquisa, que consiste em propor uma persona especialista em contabilidade, criada no modelo GPT-4 do ChatGPT, para responder questões do Exame de Suficiência em Contabilidade no Brasil. Espera-se que, à medida que a persona receba mais questões e *feedbacks*, seu desempenho melhore progressivamente nas perguntas subsequentes. Pensando nisso, na seção de Instruções, há uma parte onde a persona será informada quando estiver recebendo uma questão de treinamento, a qual deve receber um retorno, seja positivo ou negativo, após a conclusão.

Em relação aos temas abordados, pode-se perceber que o [CFC Especialista](#) está preparado não apenas para responder questões de contabilidade, mas também para fornecer

feedback, contribuindo para o ensino de contabilidade em conjunto com os professores. A expectativa é que isso traga inovação ao ensino, com base no uso da Inteligência Artificial.

Na seção de Conhecimento, foi anexado um arquivo em formato PDF contendo os conteúdos programáticos da segunda edição de 2024, semelhante às edições anteriores. Com esse arquivo, espera-se que a IA concentre seu foco nas áreas de conhecimento listadas no Exame.

3.2. Coleta de Dados

Os dados desta pesquisa foram coletados a partir das provas e gabaritos do Exame de Suficiência em Contabilidade, disponíveis no sítio oficial do Conselho Federal de Contabilidade. As edições analisadas abrangem o período de 2020 a 2024.1, com um total de 289 questões de treinamento (das edições de 2020.1 a 2022.2) e 150 questões de teste (das edições de 2023.1 a 2024.1).

As questões de treinamento foram inseridas manualmente no ChatGPT para aprimorar seu desempenho, enquanto as questões de teste foram utilizadas para avaliar sua performance após o treinamento. Nos casos em que as questões eram extensas ou continham tabelas, tornando impossível a inserção manual, optou-se por capturar uma imagem (*print*) e incluí-la no *prompt* da IA. Em seguida, todas as informações foram organizadas em uma única planilha no Microsoft Excel. As questões anuladas foram consideradas como corretas para a Persona, seguindo o mesmo critério aplicado aos candidatos do exame.

Durante o treinamento, a persona foi corrigida sempre que surgia uma resposta incorreta em comparação com o gabarito oficial disponibilizado pelo CFC. A correção consistia em informar que a resposta estava errada e em indicar a alternativa correta, permitindo que a IA assimilasse essa informação para melhorar nas próximas questões. Mas para fins de comparação dos resultados, apenas a primeira resposta foi considerada na tabulação. A Tabela 2, a seguir, apresenta o detalhamento das questões de treinamento por conteúdo, bem como as questões eliminadas do treinamento de acordo com o processo de anulação do CFC.

Tabela 2 – Questões de treinamento e anuladas por conteúdo

Conteúdo	Questões de Treinamento	Questões Anuladas	Questões Acertadas	Acertos (%)
Auditoria	12	0	9	75,0
Contabilidade Aplicada ao Setor Público	18	0	13	72,0
Contabilidade de Custos	22	2	15	63,0
Contabilidade Geral	98	4	66	65,0
Contabilidade Gerencial	12	0	10	83,0
Controladoria	4	2	4	67,0
Legislação e Ética Profissional	23	1	22	92,0
Língua Portuguesa Aplicada	12	0	8	67,0
Matemática Financeira e Estatística	12	0	10	83,0
Noções de Direito e Legislação	17	1	14	78,0
Perícia Contábil	11	1	9	75,0
Princípios Contábeis e NBC's	24	0	14	58,0
Teoria da Contabilidade	24	0	18	75,0
TOTAL	289	11	212	73,0

Nota: NBC's – Normas Brasileiras de Contabilidade.

Fonte: Elaboração própria

Para avançar na metodologia adotada em estudos anteriores, as questões foram classificadas em duas categorias: “Cálculo”, que envolvem operações matemáticas e apresentam maior complexidade para a IA, e “Texto”, que requerem apenas interpretação textual, sendo uma tarefa mais simples para o ChatGPT, conforme ilustrado no Apêndice A. Essa categorização é essencial para avaliar a performance do modelo e contribuir para o seu aperfeiçoamento. Além disso, as perguntas também foram classificadas por conteúdo, com o intuito de verificar se o modelo obteve melhorias em áreas consideradas mais desafiadoras.

Após o processo de coleta de dados supracitado, foi estruturado um banco de dados, conforme exemplificado no Apêndice B. Esse banco de dados serviu de base para a realização das análises apresentadas na Seção 4, Resultados e Discussões.

3.3. Avaliação do Desempenho

Concluídas as etapas de configuração da persona e coleta de dados, torna-se necessário estabelecer estratégias para avaliar o desempenho do modelo. Para isso, os autores desta pesquisa optaram por realizar uma comparação das porcentagens de acerto, relacionando os resultados obtidos nas edições de treinamento e teste com os de estudos anteriores que utilizaram o ChatGPT em suas versões disponíveis, sem a realização de um treinamento específico.

Inicialmente, são apresentados os resultados das edições de teste, permitindo observar a evolução do desempenho do modelo. Em seguida, optou-se por comparar o desempenho nas questões de treinamento e teste, possibilitando verificar como o modelo assimila os *feedbacks* recebidos, melhorando nas questões subsequentes. Por fim, os autores introduzem um diferencial nesta pesquisa: a categorização das questões em dois tipos, “Cálculo” e “Texto”. Essa classificação permite identificar em qual tipo de questão o ChatGPT demonstra maior assertividade e em quais áreas seu desempenho é mais limitado.

4. RESULTADOS E DISCUSSÕES

Esta seção apresenta os resultados obtidos a partir do treinamento da persona denominada “[CFC Especialista](#)”. Primeiramente, são analisados os resultados do modelo por edição e por conteúdo nas edições de teste. Em seguida, discute-se o desempenho da persona nas questões de treinamento e teste. Por fim, aborda-se o aproveitamento do modelo nas questões classificadas como do tipo “Cálculo” e “Texto”.

Seguindo a sequência mencionada, a Tabela 3 apresenta os percentuais de acerto por conteúdo e por edição de teste. Por meio dessa tabela, é possível avaliar a evolução do modelo ao longo das edições analisadas. Ao final, é exibido o resultado geral de acertos do modelo, considerando o desempenho em todas as edições de teste.

Tabela 3 – Percentuais de acerto por conteúdo e edição de teste

Conteúdo	Qtd. Questões	2023.1	2023.2	2024.1	Médias de Acerto nas 3 Edições (%)
Auditoria	2	100,0	100,0	100,0	100,0
Contabilidade Aplicada ao Setor Público	3	67,0	67,0	100,0	78,0
Contabilidade de Custos	4	50,0	50,0	50,0	50,0
Contabilidade Geral	17 (18) *	71,0	53,0	83,0	69,0
Contabilidade Gerencial	2	0,0	100,0	100,0	75,0
Controladoria	1	0,0	0,0	100,0	33,0
Legislação e Ética Profissional	4 (3) *	100,0	100,0	100,0	100,0
Língua Portuguesa Aplicada	2	50,0	100,0	100,0	83,0
Matemática Financeira e Estatística	2	100,0	100,0	100,0	100,0
Noções de Direito e Legislação	3	67,0	100,0	67,0	78,0
Perícia Contábil	2	100,0	100,0	100,0	100,0

Princípios Contábeis e NBC's	4	100,0	75,0	100,0	92,0
Teoria da Contabilidade	4	75,0	100,0	100,0	92,0
TOTAL	50	68,0	80,0	92,0	81,0

*Primeiro exame de 2024

Nota: NBC's – Normas Brasileiras de Contabilidade.

Fonte: Elaboração própria

De acordo com a Tabela 3, o treinamento específico aplicado para avaliar o desempenho do ChatGPT no Exame de Suficiência em Contabilidade, ao longo de três edições, revelou uma evolução significativa no percentual de acertos, destacando-se especialmente na primeira edição de 2024. No geral, a média de acertos do modelo aumentou de 68% na edição 2023.1 para 80% na edição 2023.2, atingindo 92% em 2024.1. Em média, as três edições analisadas resultaram em uma taxa de acerto geral de 81% por conteúdo.

Dando continuidade à análise, a Tabela 4 apresenta os resultados do modelo nas questões de treinamento e teste. Essa tabela é fundamental para avaliar a evolução do modelo, permitindo a comparação de seu desempenho antes e depois do treinamento. No total, foram utilizadas 300 questões de treinamento, das quais 289 foram inseridas no *prompt* do modelo, enquanto 11 foram anuladas pelo CFC. Para as questões de teste, foram incluídas 150 questões das edições de 2023.1 a 2024.1.

Tabela 4 – Acertos de questões de treinamento e teste

Conteúdo	Questões de Treinamento	Acertos (%)	Questões de Teste	Acertos (%)
Auditoria	12	75,0	6	100,0
Contabilidade Aplicada ao Setor Público	18	72,0	9	78,0
Contabilidade de Custos	24	63,0	10	50,0
Contabilidade Geral	102	65,0	52	69,0
Contabilidade Gerencial	12	83,0	8	75,0
Controladoria	6	67,0	3	33,0
Legislação e Ética Profissional	24	92,0	11	100,0
Língua Portuguesa Aplicada	12	67,0	6	83,0
Matemática Financeira e Estatística	12	83,0	6	100,0
Noções de Direito e Legislação	18	78,0	9	78,0
Perícia Contábil	12	75,0	6	100,0
Princípios Contábeis e NBC's	24	58,0	12	92,0
Teoria da Contabilidade	24	75,0	12	92,0
TOTAL	300	73,0	150	81,0

Nota: NBC's – Normas Brasileiras de Contabilidade.

Fonte: Elaboração própria

Ao comparar o desempenho entre as questões de treinamento e as questões de teste por conteúdo, observou-se um aumento de 8 pontos percentuais (Teste: 81% vs. Treinamento: 73%). Analisando o resultado das questões de teste, verificou-se que a persona obteve desempenho máximo em quatro áreas: Auditoria Contábil, Legislação e Ética Profissional, Matemática Financeira e Estatística, e Perícia Contábil. Em contraste, o modelo apresentou maior dificuldade em Contabilidade de Custos e Controladoria, alcançando médias de 50% e 33%, respectivamente, considerando o total das edições analisadas.

Em relação aos resultados das questões de treinamento, nota-se que o modelo não atingiu 100% de acerto em nenhum dos conteúdos, mesmo considerando as 11 questões anuladas que foram contabilizadas como corretas. Esse resultado sugere que a persona

assimilou progressivamente o conteúdo das questões de treinamento e as correções aplicadas, aprimorando seu desempenho conforme avançava nas resoluções. Essa melhora é evidenciada pela comparação com os resultados obtidos nas questões de teste do modelo.

No entanto, especificamente em relação ao conteúdo de Controladoria, observou-se que o desempenho do modelo diminuiu após o treinamento, passando de 67% de acerto para 33%, o que corresponde a apenas uma questão correta, que é do tipo "Texto". Essa diminuição é explicada, em parte, pelo fato de o percentual de acerto nas questões de treinamento ter sido maior devido a duas questões anuladas, cujos pontos foram atribuídos ao modelo.

Além disso, ao comparar o melhor desempenho obtido em estudos anteriores, que foi de 74% (na primeira edição de 2022), com a média de 81% alcançada após o treinamento (média geral dos exames de 2023 a 2024.1), verifica-se uma evolução de 7 pontos percentuais. Vale destacar que o melhor resultado apresentado no estudo de Oliveira Jr. e Khatib (2023) incluiu duas questões anuladas consideradas corretas para o ChatGPT, enquanto, nos resultados das edições de teste desta pesquisa, não houve questões anuladas.

Esses resultados indicam que um treinamento específico pode aumentar o desempenho do ChatGPT nas questões do Exame de Suficiência em Contabilidade, demonstrando a eficácia da abordagem adotada.

4.1. Resultado por Tipo de Questão

Adicionalmente à análise realizada, com o objetivo de avançar os estudos anteriores sobre o desempenho do ChatGPT em questões do Exame de Suficiência em Contabilidade por meio de um treinamento específico, a performance do modelo é abordada posteriormente de acordo com o tipo de questão, seja "Cálculo" ou "Texto". Essa classificação permite uma análise mais detalhada do comportamento do modelo em diferentes tipos de questões.

Nesse cenário, a Tabela 5 apresenta o desempenho do CFC Especialista nas edições de teste, segmentado por tipo de questão, abrangendo as edições de 2023.1 a 2024.1, totalizando três edições, cada uma composta por 50 questões. Ao final, é demonstrado o resultado geral do modelo, levando em consideração todas as edições de teste.

Tabela 5 – Acertos por tipo de questão

Tipo de Questão	2023.1		2023.2		2024.1		Total Geral		
	Questões	Acertos	Questões	Acertos	Questões	Acertos	Questões	Acertos	%
Cálculo	28	18	24	13	19	15	71	46	64,0
Texto	22	18	26	24	31	30	79	72	91,0
TOTAL	50	36	50	37	50	45	150	118	78,0

Fonte: Elaboração própria

Conforme observado na tabela acima, o desempenho do modelo variou significativamente entre os diferentes tipos de questões: nas questões que exigiam, além da interpretação textual, a realização de cálculos a partir da análise de tabelas, gráficos e listas, o percentual de acerto foi de 64% (46 de 71). Em contraste, nas questões que demandavam apenas interpretação textual, o ChatGPT alcançou um desempenho de 91% (72 de 79).

O baixo percentual de acerto observado nas questões do tipo "Cálculo", em comparação com as questões do tipo "Texto", deve-se à má interpretação dos dados fornecidos e à complexidade das inter-relações entre as variáveis, o que torna as questões envolvendo cálculos mais desafiadoras para a IA, conforme destacado na resposta do ChatGPT no Apêndice A.

A diferença percentual entre os dois tipos de questão, que totaliza 27 pontos percentuais (Cálculo: 64% vs. Texto: 91%), sugere que o modelo do ChatGPT apresenta melhor performance em questões que envolvem apenas interpretação textual, as quais, em sua maioria, podem ser representadas pelos conteúdos de Auditoria, Perícia Contábil,

Princípios e NBC's, Noções de Direito e Legislação, Legislação e Ética Profissional, Teoria da Contabilidade e Língua Portuguesa Aplicada conforme Tabela 6.

Tabela 6 – Questões de cálculo ou texto por conteúdo

Conteúdo	Cálculo	Texto	TOTAL
Auditoria	0	6	6
Contabilidade Aplicada ao Setor Público	3	6	9
Contabilidade de Custos	5	5	10
Contabilidade Geral	41	11	52
Contabilidade Gerencial	7	1	8
Controladoria	2	1	3
Legislação e Ética Profissional	0	11	11
Língua Portuguesa Aplicada	0	6	6
Matemática Financeira e Estatística	4	2	6
Noções de Direito e Legislação	0	9	9
Perícia Contábil	0	6	6
Princípios Contábeis e NBC's	5	7	12
Teoria da Contabilidade	4	8	12
TOTAL	71	79	150

Nota: NBC's – Normas Brasileiras de Contabilidade.

Fonte: Elaboração própria

Esses resultados indicam que, de modo geral, o modelo apresentou uma melhoria contínua em diversos conteúdos ao longo das edições analisadas, alcançando um desempenho mais elevado em questões de interpretação textual em comparação com aquelas que exigem cálculos. A primeira edição de 2024, que apresentou o maior número de questões do tipo texto (31), foi também a edição em que o modelo obteve o maior número de acertos. Dessa forma, conclui-se que os exames que apresentarem uma maior proporção de questões do tipo "Texto", em comparação com as questões de "Cálculo", tendem a resultar em um maior índice de acerto pelo ChatGPT, visto que questões de interpretação textual são mais facilmente processadas e respondidas pelo modelo.

5. CONSIDERAÇÕES FINAIS

Esta pesquisa teve como objetivo propor uma persona especialista em contabilidade, criada no modelo GPT-4 do ChatGPT, para responder questões do Exame de Suficiência em Contabilidade no Brasil. Nesse sentido, esta pesquisa desenvolveu uma persona, chamada [CFC Especialista](#). Após realizado o treinamento, os resultados foram bastante satisfatórios considerando as edições de 2023.1 a 2024.1, com o modelo alcançando uma taxa de acerto de 92% por conteúdo na última edição testada. Observou-se que a persona apresenta maior facilidade em responder questões do tipo "Texto", que exigem apenas interpretação textual, enquanto as questões do tipo "Cálculo" requerem maior atenção, devido aos erros de interpretação e processamento identificados tanto nesta análise quanto em estudos anteriores.

Dessa forma, a contribuição deste estudo está na possibilidade de utilizar o ChatGPT como uma nova ferramenta para o ensino da Ciência Contábil, tanto no ambiente acadêmico quanto no mercado de trabalho. Ressalta-se, contudo, a importância de contar com um profissional qualificado para avaliar as respostas fornecidas pelo modelo, garantindo seu uso adequado para os fins propostos. Outra contribuição deste estudo é o avanço em relação às pesquisas anteriores, ao realizar um treinamento específico para avaliar o desempenho do

modelo. Nessa linha, esta pesquisa destacou o uso de uma persona especialista em contabilidade, que pode ser implementada em programas educacionais para auxiliar estudantes e professores na preparação para o exame, um requisito fundamental para a habilitação de futuros contadores no Brasil.

Quanto às limitações desta pesquisa, pode-se afirmar que estas se concentram, principalmente, no processo de análise dos resultados. Uma avaliação mais robusta poderia ter sido realizada para identificar a significância estatística entre os dados obtidos. Por exemplo, seria pertinente conduzir uma análise comparativa entre grupos (Teste vs. Treinamento). Além disso, a performance do modelo poderia ser analisada de forma mais detalhada, comparando o desempenho nas diferentes categorias de questões, como "Cálculo" e "Texto", para aprofundar a compreensão sobre o desempenho do ChatGPT em cada tipo de questão.

Como sugestão para pesquisas futuras, recomenda-se o desenvolvimento de uma persona mais robusta ou a continuidade dos treinamentos com a persona já existente, visando aprimorar seu desempenho em futuras questões de contabilidade. Em ambos os casos, sugere-se fornecer um *feedback* mais abrangente e detalhado, contemplando não apenas os erros cometidos, mas também os acertos, para que a IA possa processar essas informações e aprimorar seu desempenho em resoluções subsequentes.

REFERÊNCIAS

AYANOUS, Soufyane; ABDELHAKIM, Boudhir Anouar; BENHMED, Mohammed. A smart *chatbot* architecture based NLP and machine learning for health care assistance. In: **Proceedings of the 3rd international conference on networking, information systems & security**. 2020. p. 1-6.

BICALHO, Gabriela Magalhães; DE OLIVEIRA, Arthur Henrique; DE SIQUEIRA GUIDA, José Paulo. Desempenho da inteligência artificial em questões de processo seletivo de residência médica. **Mantendo a acurácia e ampliando o acesso**, v. 52, n. 6, p. 370-3, 2024.

BRASIL. **Lei nº 12.249/2010**, de 11 de junho de 2010. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/lei/l12249.htm. Acesso em 25 de setembro de 2024.

CHOI, Jonathan H. et al. ChatGPT goes to law school. **J. Legal Educ.**, v. 71, p. 387, 2021.
DE FREITAS, Marcelo Machado et al. RESOLUÇÃO DE PROBLEMAS CONTÁBEIS COM O CHATGPT 4.

CONSELHO FEDERAL DE CONTABILIDADE [CFC]. Resolução CFC n. 1486/2015. Brasília, DF, 2015. Disponível em: <https://cfc.org.br/registro/legislacao/>. Acesso em: 25 de setembro 2024.

DE OLIVEIRA JR, Jose Carlos Ramos; EL KHATIB, Ahmed Sameer. Homem ou Máquina? Um estudo exploratório do desempenho do Chat GPT 3.5 no Exame de Suficiência do CFC Man or machine? An exploratory study of GPT 3.5 chat performance in the CFC sufficiency exam. **Revista Capital Científico-Eletrônica (RCC-e)-ISSN 2177-4153**, v. 22, n. 1, p. 42-56, 2024.

DESHPANDE, Ameet et al. Toxicity in chatgpt: Analyzing persona-assigned language models. **arXiv preprint arXiv:2304.05335**, 2023.

FERREIRA, Rafael Clementino Veríssimo; GARCIA, Gustavo Henrique Maia; BRASIL, Deilton Ribeiro. O surgimento do Chat GPT e a insegurança sobre o futuro dos trabalhos acadêmicos. **Cadernos de Direito Actual**, n. 21, p. 130-143, 2023. Rodrigues, O. S., & Rodrigues, K. S. (2023). A inteligência artificial na educação: os desafios do ChatGPT. *Texto Livre*, 16, e45997.

KUNG, Tiffany H. et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. **PLoS digital health**, v. 2, n. 2, p. e0000198, 2023.

LEITÃO, Cleverson Alex et al. Desempenho do ChatGPT nas questões da avaliação anual de residentes do Colégio Brasileiro de Radiologia. **Radiologia Brasileira**, v. 57, p. e20230083, 2024.

LIMA RODRIGUES, Lúcia et al. Factors affecting success in the professional entry exam for accountants in Brazil. **Accounting Education**, v. 27, n. 1, p. 48-71, 2018.

OLIVEIRA JUNIOR, Osvaldo Novais de; PARDO, Thiago Alexandre Salgueiro; NUNES, Maria das Graças Volpe. ChatGPT: o robô que mostra como a inteligência artificial pode revolucionar nossas vidas. **Jornal da USP**, v. 13, 2023.

SANTHOSH, R. et al. Chatgpt: Opportunities, features and future prospects. In: **2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)**. IEEE, 2023. p. 1614-1622.

VASWANI, A. Attention is all you need. **Advances in Neural Information Processing Systems**, 2017.

WOOD, David A. et al. O *chatbot* de inteligência artificial ChatGPT: quão bem ele responde a perguntas de avaliação contábil?. **Issues in Accounting Education** , v. 38, n. 4, p. 81-108, 2023.

Apêndice A

Grau de dificuldade das questões

Em uma consulta para determinar o grau de dificuldade das questões, o ChatGPT respondeu da seguinte forma:

As duas categorias de questões apresentam graus de dificuldade diferentes para mim:

Questão que necessita apenas de interpretação textual: O grau de dificuldade é baixo. Tenho uma habilidade forte em interpretar textos complexos, identificar padrões e responder com base no conteúdo textual fornecido. Isso inclui interpretar cenários, entender normas, e oferecer explicações detalhadas.

Questão que além de interpretação textual também é necessário realizar cálculos: O grau de dificuldade é moderado a variável, dependendo da complexidade dos cálculos. Tenho capacidade de realizar cálculos simples a complexos, desde que os dados e informações sejam fornecidos com clareza. Para cálculos mais avançados ou extensos, a chance de erro pode aumentar se houver ambiguidade nos dados ou se as operações envolverem muitas variáveis interligadas.

A execução de cálculos requer uma interpretação precisa dos dados fornecidos e a correta aplicação de fórmulas, o que **pode tornar a tarefa mais desafiadora**, principalmente quando envolve muitos passos ou múltiplas variáveis interdependentes.

Apêndice B

Estrutura do banco de dados de coleta

Ano	Edição	Caderno	Nº Questão	Conteúdo	Treinamento	Tipo	Gabarito CFC	Sem Persona	Com Persona
2024	1	Branco	1	Língua Portuguesa Aplicada	Não	Texto	B	-	B
2024	1	Branco	2	Língua Portuguesa Aplicada	Não	Texto	C	-	C
2024	1	Branco	3	Matemática Financeira e Estatística	Não	Cálculo	C	-	C
2024	1	Branco	4	Matemática Financeira e Estatística	Não	Cálculo	A	-	B
2024	1	Branco	5	Noções de Direito e Legislação Aplicada	Não	Texto	C	-	C

Nota: Ano: Ano de realização do exame; Edição: Refere-se à edição de aplicação do exame (1 ou 2); Caderno: Indica o caderno de onde as questões foram extraídas; Nº Questão: Número correspondente à questão dentro do caderno; Conteúdo: Área de conhecimento abordada pela questão; Treinamento: Informa se a questão foi utilizada para treinamento da Persona (Sim ou Não); Tipo: "Cálculo" refere-se a questões que exigem cálculos para obter a resposta, enquanto "Texto" refere-se a questões que envolvem apenas interpretação textual. Gabarito CFC: Gabarito oficial da questão, indicado por letras (A, B, C ou D); Sem Persona: Resposta gerada pelo ChatGPT antes do treinamento específico; e com persona: Resposta gerada pelo ChatGPT após o treinamento com uma Persona específica.

Fonte: Elaboração própria.