Relato de Experiência

Integração de Inteligência Artificial na Avaliação Educacional: Um Relato de Experiência no Ensino Técnico a Distância

Autor Principal
PAULO HENRIQUE MENDES CARVALHO
0009-0003-5929-9596

Coautor
ELIANA CRISTINA NOGUEIRA BARION
0009-0003-7082-489X

Resumo

Este relato de experiência descreve o desenvolvimento de um roteiro sistemático para a análise de questões de provas dos cursos técnicos do Centro Paula Souza, na modalidade on-line, utilizando Inteligência Artificial (IA) como ferramenta de apoio. O estudo visa garantir a qualidade das avaliações presenciais, verificando a adequação das questões em relação ao conteúdo programático e à capacidade de mensurar o aprendizado dos alunos. A metodologia adotada inicialmente foi de abordagem quantitativa, com a extração de dados tabulados, a partir do resultado do primeiro exame presencial, aplicado no primeiro semestre de 2024, do curso Técnico em Desenvolvimento de Sistemas, seguida da análise qualitativa das questões. A ferramenta ChatGPT 4.0 Plus foi utilizada para a análise quantitativa e qualitativa das provas, aplicando como base a Teoria Clássica dos Testes (TCT), com foco nos índices de dificuldade e discriminação das questões. Os resultados evidenciaram que, embora o tempo de processamento tenha sido maior devido ao volume de dados, a IA foi eficiente na identificação de padrões de erro e acerto, analisando distratores e sugerindo possíveis melhorias nas questões. O processamento do material de estudo também foi satisfatório, permitindo a extração de exemplos relevantes. Contudo, algumas inconsistências foram identificadas nos textos gerados pela IA, o que reforça a importância da intervenção humana para garantir a precisão dos resultados. A temática é relevante e torna-se urgente investir no aprofundamento teórico-metodológico no campo da Educação, sob novas perspectivas críticas, permeadas pela reflexão, sentidos e significados. Este estudo aponta para o potencial da IA como uma ferramenta para aprimorar o processo avaliativo, porém destaca a necessidade de supervisão e validação dos dados gerados.

Palavras-chave: Análise de questões. Avaliação Educacional. Educação Profissional. Inteligência Artificial. Teoria Clássica dos Testes.

Abstract

This experience report describes the development of a systematic script for analyzing questions from technical exams at Centro Paula Souza, in the online modality, using Artificial Intelligence (AI) as a support tool. The study aims to ensure the quality of in-person assessments, verifying the adequacy of the questions in relation to the syllabus and the ability to measure student learning. The methodology initially adopted was a quantitative approach, with the extraction of tabulated data from the results of the first in-person exam, applied in the first semester of 2024, of the Technical Course in Systems Development, followed by the qualitative analysis of the questions. The ChatGPT 4.0 Plus tool was used for the quantitative and qualitative analysis of the exams, applying Classical Test

Theory (CTT) as a basis, focusing on the difficulty and discrimination indexes of the questions. The results showed that, although the processing time was longer due to the volume of data, the AI was efficient in identifying error and success patterns, analyzing distractors and suggesting possible improvements in the questions. The processing of the study material was also satisfactory, allowing the extraction of relevant examples. However, some inconsistencies were identified in the texts generated by AI, which reinforces the importance of human intervention to ensure the accuracy of the results. The topic is relevant and it is urgent to invest in theoretical and methodological deepening in the field of Education, under new critical perspectives, permeated by reflection, senses and meanings. This study points to the potential of AI as a tool to improve the evaluation process, but highlights the need for supervision and validation of the data generated.

Keywords: Question Analysis. Educational Assessment. Professional Education. Artificial Intelligence. Classical Test Theory.

1 Introdução

De acordo com os dados dos últimos censos da Associação Brasileira de Educação a Distância (ABED, 2019-2023), o número de matrículas nos cursos da modalidade a distância cresceu expressivamente. No Centro Paula Souza (CPS), o cenário não foi diferente, tendo observado aumento significativo do número de matrículas nos cursos técnicos, nesse mesmo período. Com a expansão da quantidade de vagas verificou-se a necessidade urgente de se averiguar a qualidade dos cursos oferecidos nessa modalidade e um dos aspectos que consideramos importante e que, portanto, nos debruçamos para garantir a qualidade dos cursos, foi o processo avaliativo, especialmente no que se refere às provas presenciais, utilizadas semestralmente como ferramenta de mensuração do conhecimento dos alunos.

Assim, iniciamos o processo de investigação das questões de prova, a fim de verificar se são capazes de avaliar efetivamente o aprendizado dos alunos, considerando não apenas o nível de dificuldade das questões, mas também observando se a sua elaboração condiz com as bases tecnológicas, competências e habilidades previstas no plano de curso. Essa análise vai ao encontro de Krathwohl (2002) que enfatiza a importância do alinhamento entre os objetivos de aprendizagem, às estratégias de ensino e aos métodos de avaliação.

Na sociedade do conhecimento, sabemos que as tecnologias da informação e comunicação assumem espaços e forças capazes de contribuir com as práticas pedagógicas e metodologias, com implicações nas ações do professor e sua relação com o aluno e a aprendizagem (Castells, 2002**;** Pimenta e Anastasiou, 2010). Com a popularização da Inteligência Artificial (IA), em sua forma generativa, ocorrida no início de 2023, criou-se caminhos para a análise de dados de forma automatizada.

Se a IA é assunto em todos os contextos, na educação não é diferente e tende a revolucionar o processo de ensino e aprendizagem, à medida em que os educadores aproveitam as capacidades dessas ferramentas para conceber e desenvolver ambientes de aprendizagem dinâmicos e adaptáveis que atendem às diversas necessidades de seus alunos, moldando o futuro da educação para enfrentar os desafios em constante evolução (Linares, Fuentes; Galdames, 2023).

Para Ferreira e Stephen (2023), o desafio é aprender a colaborar com algoritmos, vendo-os não como ameaças, mas como extensões das capacidades humanas.

> [...] profissionais que conseguem integrar sua expertise com as novidades da IA têm chance de liderar a próxima fase de inovação no campo criativo. Nesse cenário, a adaptação não é apenas desejável, mas essencial. (FERREIRA; STEPHEN, 2023)

A fim de auxiliar no processo de investigação das questões de prova, contamos com a ajuda da Inteligência Artificial do ChatGPT 4.0 para ganhar agilidade e eficiência nas análises dos dados e identificar melhoria na elaboração das provas. Esse trabalho, que seria extremamente custoso para desenvolvimento de novos sistemas e demoraria muito tempo se realizado manualmente, com a utilização da IA permitiu comparações precisas entre as questões e os erros e acertos dos alunos, refletindo de forma mais acurada o cenário real das avaliações.

Tendo em vista essa nova possibilidade, foi desenvolvido o objeto desse relato de experiência, como modelo para um roteiro de análise de questões de exames presenciais, primeiramente quantitativo, com a compilação dos dados e identificação automática de pontos críticos de atenção, seguida da análise qualitativa das questões, baseada na comparação com o material de estudo, buscando encontrar um coeficiente de aderência entre o material estudado e o exame realizado.

Com tais informações geradas, de forma automatizada, surgem novas possibilidades para a tomada de decisão do professor conteudista, seja em relação à atualização do material de estudo, ou à reorganização das questões para os exames subsequentes.

2 Objetivo

Diante disso, esse artigo propõe-se a desenvolver um roteiro sistemático para análise das questões das provas dos cursos técnicos, da modalidade on-line, do Centro Paula Souza, com o auxílio da Inteligência Artificial, que permita identificar a sua adequação em relação ao conteúdo abordado e à real capacidade de avaliação do aprendizado dos alunos. Assim, busca-se contribuir para a melhoria da qualidade na educação a distância, garantindo que as avaliações sejam não apenas um requisito formal, mas uma verdadeira reflexão do aprendizado significativo.

Como objetivos secundários, esse trabalho pretende identificar as necessidades de viabilidade de execução deste modelo por usuários com pouco conhecimento em Inteligências Artificiais, analisar a possibilidade de execução do roteiro por qualquer ferramenta de IA, além de validar a veracidade das informações fornecidas pela Inteligência Artificial Generativa como resultado do estudo.

3 Método

Dentre outros objetivos, a análise de questões de provas é fundamental para garantir que as avaliações reflitam as competências e habilidades definidas no plano de curso, bem como a identificação de problemas estruturais nas questões. Nesse sentido, a Teoria Clássica dos Testes (TCT) fornece uma estrutura robusta para examinar a dificuldade e a discriminação das questões, permitindo que educadores avaliem se o ensino está efetivamente orientado para o desenvolvimento das habilidades desejadas. Como mencionado por Horta Neto

(2010), a qualidade na educação é subjetiva e varia conforme os sujeitos envolvidos, o que exige a definição de critérios claros de qualidade para as avaliações.

Além disso, a divulgação dos resultados, conforme abordado por Bauer, Alavarze e Oliveira (2015), é fundamental para informar alunos e professores sobre o desempenho e a aprendizagem, possibilitando intervenções pedagógicas para tomadas de decisões.

O modelo da TCT, também conhecido como Modelo Linear Clássico, foi inicialmente desenvolvido por Spearman (Muñiz, 1994; Pasquali, 2009) e formalizado por Gulliksen (Gulliksen, 1950). Outros autores, como Lord (1959) e Novick (1966), também forneceram uma visão geral sistemática do modelo.

Com base nessas contribuições, podemos apresentar os fundamentos da TCT, que se baseiam em três elementos essenciais. O postulado fundamental da teoria é expresso na fórmula: T = V + E. Isso significa que o escore observado é igual à soma do escore verdadeiro e do erro, definido por Gulliksen (1950) como:

T = escore bruto ou empírico do sujeito, que é a soma dos escores obtidos no teste;

V = escore verdadeiro, que seria a magnitude real daquilo que o teste quer medir no sujeito e que seria o próprio T, se não houvesse erro de medida e;

E = o erro cometido nesta medida.

É fundamental destacar que o erro de medida está presente em qualquer análise empírica. O objetivo da Teoria Clássica dos Testes (TCT) é desenvolver técnicas estatísticas que ajudem a controlar ou estimar a magnitude desse erro durante a aplicação dos testes (Muñiz, 1994; Pasquali, 2009). Assim, o erro pode ser entendido como a diferença entre o escore verdadeiro (a pontuação real do indivíduo) e o escore observado (a pontuação que o indivíduo obteve no teste), também conhecido como escore empírico (Lord, 1959).

Ademais, é importante ressaltar que o erro é aleatório e não sistemático; ao realizar o teste várias vezes, um indivíduo pode obter pontuações diferentes, que podem ser maiores ou menores que a pontuação verdadeira. Sendo um evento casual, a média do erro tende a ser 0 (Muñiz, 1994).

A escolha da TCT neste estudo se justifica pela necessidade de rigor na construção e análise de resultados, especialmente pelo fato de que as avaliações em evidência não foram elaboradas com elementos da TRI. Essa abordagem permite uma avaliação mais coerente e alinhada aos objetivos educacionais do Grupo de Estudo de Educação a Distância (GEEaD) do Centro Paula Souza, promovendo uma reflexão crítica sobre o processo de ensino e aprendizagem.

Semestralmente, o GEEaD realiza os exames presenciais dos sete cursos oferecidos na modalidade on-line, compostos por trinta questões de múltipla escolha, com cinco alternativas que objetivam analisar as competências e habilidades do estudante, adquiridas durante o módulo estudado. Os resultados dessas avaliações são fundamentais para que haja uma intervenção pedagógica, a fim de averiguar a consonância das questões com o material de estudo, a metodologia do curso, bem como a mediação realizada pelos professores.

Tendo em vista a importância desta avaliação para o processo de ensino e aprendizagem desses cursos, esse relato de experiência traz uma análise

estatística destas avaliações com base na teoria clássica dos testes (TCT) que oferece uma abordagem em que se investiga o escore total do teste.

Assim, este relato de experiência adotou uma abordagem quantitativa, com a extração de dados tabulados, a partir do resultado do primeiro exame presencial, aplicado no primeiro semestre de 2024, do curso Técnico em Desenvolvimento de Sistemas, ofertado na modalidade on-line.

Após a coleta dos dados, foram estabelecidos como parâmetros deste estudo a TCT, com foco na análise da dificuldade e da discriminação das questões. Os grupos de desempenho foram divididos entre os 27% superiores e 27% inferiores, além dos 46% que representam o desempenho mediano. Também foram calculadas a média de desempenho no exame para facilitar análises posteriores.

A Ferramenta de Inteligência Artificial Chat GPT 4.0 Plus foi utilizada para extração dos índices apresentados na TCT, além da verificação automatizada dos possíveis motivos de erros e aderência da questão ao material de estudo.

A utilização de instrumentos e testes psicométricos tem sido uma forma de avaliar objetivamente os fenômenos psicológicos e, apesar de algumas críticas e resistências por parte de alguns psicólogos, o uso de testes psicológicos tem se fortalecido, pois permite estabelecer um referencial que minimiza os vieses subjetivos na percepção e no julgamento do examinador. Isso é possível, desde que o profissional possua um profundo conhecimento sobre a aplicação e a interpretação desses testes (SOUZA, 2020).

De acordo com Pasquali (2009), a psicometria fundamenta-se na teoria da medida em ciências para explicar o sentido que têm as respostas dadas pelos sujeitos a uma série de tarefas e propor técnicas de medida dos processos mentais. Assim, busca mensurar um amplo conjunto de habilidades cognitivas, frequentemente utilizando baterias de testes de inteligência que abrangem conteúdos variados. Nesse contexto, a psicometria complementa a Teoria Clássica dos Testes (TCT) ao fornecer ferramentas para a implementação de análises estatísticas robustas.

Baseado nos conceitos da psicometria, para a extração dos dados do resultado do exame presencial, foram selecionadas as turmas do primeiro módulo do curso Técnico em Desenvolvimento de Sistemas, devida à quantidade de amostras ser considerada grande em relação aos cursos oferecidos pelo GEEaD, sendo 588 amostras, o que indica que a confiabilidade das informações encontradas pode ser enquadrada como muito alta.

> Considerando, nesse momento, somente o índice de dificuldade, pode-se dizer que um bom item é aquele que possui alta variância, visto que o objetivo do teste é explicitar as variações que existem entre os indivíduos, itens com alta variância irão contribuir para uma maior variância do escore do teste, já que uma das parcelas da variância do escore é a soma da variância dos itens individuais. Isso irá permitir uma maior discriminação dos indivíduos em função dos escores (PRIMI, 2012, pág. 305).

Das trinta questões que compunham o exame presencial, objeto de análise desse estudo, a questão dezenove foi anulada após a divulgação do gabarito, atribuindo pontuação nessa questão para todos os alunos que realizaram essa prova. Os dados das amostras foram extraídos após a digitalização das folhas de respostas preenchidas pelos alunos durante o exame presencial. Os dados que revelam a identificação da amostra foram removidos em consonância aos termos

da LGPD (Lei Geral de Proteção de Dados). Após as adequações necessárias para o estudo, os seguintes dados foram planilhados para análise:

Tabela 1: Dicionário de variáveis do resultado de provas

Nome	Tipo	Tamanho	Descrição
nome_curso	texto	50	Indica o curso no qual o dado foi extraído.
leitura	texto	60	Mostra a resposta dada para cada uma das 30 questões na forma de letras, sendo, “A”, “B”, “C”, “D” ou “E”. Os dados estão separados por vírgula. Questões não respondidas são caracterizadas por espaço vazio e questões com múltiplas respostas foram consideradas anuladas, sendo removidas da análise final.
correcao	texto	60	Exibe a resposta correta para cada uma das 30 questões. Os dados estão separados por vírgula, sendo que o valor “1”, indica o acerto, e “0”, o erro.
pontos	numero	2	Pontuação obtida no exame, podendo ser o resultado um número de 1 a 30, devido a questão 19 ter sido anulada.
q1 a q30	caractere	1	Cada uma dessas colunas representa uma questão específica do exame, mostrando a resposta fornecida pelo aluno.
cq1 a cq30	caractere	1	Colunas que indicam a correção de cada questão, mostrando se a resposta foi correta "1" ou incorreta "0".

Fonte: GEEaD (2024).

A partir do tratamento dos dados necessários para a análise, utilizamos o ChatGPT como ferramenta de apoio, por ser uma Inteligência Artificial do tipo GPT(Generative Pre-Trained Transformer), assim, com a sua característica de ser uma IA pré treinada, dispensa-se o prévio treinamento, possibilitando que pessoas com pouco ou nenhum conhecimento em Inteligência Artificial possa manipulá-lo e, futuramente, executar o roteiro pretendido.

Além disto, para este estudo, optou-se pela versão 4o Plus da ferramenta, devido a sua habilidade de processar documentos completos, sem a necessidade de inserir manualmente em seu prompt[1] que possibilita o envio de até 128000 caracteres, com isto, uma amostra de resultado grande poderia facilmente ultrapassar o limite de caracteres do prompt. É importante ressaltar que a ferramenta também possibilita o envio de instruções anteriores para a personalização dos resultados obtidos.

Com a seleção da ferramenta, definiu-se que seria necessário a criação de um roteiro prévio para a análise das amostras, com o objetivo de padronizar a saída e definir um fluxo de análise coeso. Assim, foi definido o seguinte roteiro com 27 indicadores:

[1] Campo para entrada de texto para processamento das ferramentas de Inteligência Artificial do tipo GPT.

1. Responda somente em português, mesmo quando perguntado em inglês. Quando precisar trabalhar com números flutuantes, trabalhe com no máximo duas casas decimais.
2. Receba um arquivo com o resultado de uma prova. Ele poderá ser um arquivo .csv ou .xlsx (planilha). Caso não seja enviado, não faça nada sem recebê-lo.
3. Utilize sempre a teoria clássica dos testes (TCT) como base de cálculo de resultados.
4. Desconsidere todos os registros em que o campo ausente esteja com valor "A" e onde o campo pontos estiver 0.
5. Você deverá separar os dados por curso, que é o campo nome_curso.
6. Antes de seguir os cálculos, identifique todas os cursos presentes no documento, analisando previamente todas as linhas do documento. Analise o documento por inteiro antes de perguntar qual curso que o usuário quer analisar. Aponte todos os cursos identificados e pergunte se o usuário quer que analise um em específico, ou que faça a análise de todos os cursos.
7. Se o usuário apontar todos, faça uma análise para cada curso individualmente. Lembre-o que pode demorar um pouco pelo volume de dados, que seria melhor em uma próxima vez solicitar um curso isolado.
8. Os campos com números são acertos e erros, por questão, respectivamente, sendo 0 erro e 1 acerto.
9. Os campos com Letras de A a E são as respostas que foram fornecidas por questão.
10. Indique a melhor e pior pontuação, a média e a quantidade de provas no(s) curso(s), quantos ausentes e quantos zerados foram descartados.
11. Calcule e aponte o valor das faixas dos 27% superior e inferior. O mediano e o total.
12. Aponte a porcentagem e quantidade de alunos que ficou em cada um dos destes itens (superior; mediano; inferior).
13. Aponte o cálculo que você fará antes de iniciar o trabalho.
14. O índice de dificuldade é dado pela fórmula:
    indice_dificuldade = 1 - (acertos_totais / total_respostas_validas)
    Neste caso:
    acertos_totais: soma dos acertos em uma questão para todos os alunos válidos.
    total_respostas_validas: número total de respostas válidas (alunos que responderam).
15. O índice de discriminação é calculado usando a fórmula:
    indice_discriminacao = (acertos_superior - acertos_inferior) / (total_respostas_superior + total_respostas_inferior)
    Sendo que:
    acertos_superior: número de acertos no grupo superior (27%).
    acertos_inferior: número de acertos no grupo inferior (27%).
    total_respostas_superior: número de respostas válidas no grupo superior.
    total_respostas_inferior: número de respostas válidas no grupo inferior.
16. Aponte o índice de dificuldade e discriminação de cada questão. Além do índice, aponte para cada questão se a dificuldade é muito fácil, fácil, média, difícil ou muito difícil. Aponte como discriminação se ela é muito fraca, fraca, boa, muito boa, ótima ou excelente.

17. Aponte qual é o gabarito da prova, de acordo com a sua conclusão da análise. Neste caso, analise, minuciosamente, se a letra correspondente a questão está de acordo com a casa de número 1 do resultado.
18. Faça um gráfico com o desempenho dos alunos por questão considerando a quantidade de acertos.
19. Faça uma análise e indique um "ranking" de dificuldade e de discriminação das questões. O ranking deve conter todas as questões, considerando desde a mais fácil até a mais difícil. Especifique o número da questão, o indice_dificuldade e a respectiva dificuldade.
    O índice de dificuldade a ser considerado deverá ser:
    Índices de dificuldade
    Muito Fácil 0,000 - 0,149
    Fácil 0,150 - 0,349
    Média 0,350 - 0,649
    Difícil 0,650 - 0,849
    Muito Difícil 0,850 - 1,000

    Índices de discriminação
    Muito Fraca 0,000 - 0,099
    Fraca 0,100 - 0,199
    Boa 0,200 - 0,399
    Muito Boa 0,400 - 0,599
    Ótima 0,600 - 0,799
    Excelente 0,800 - 1,000
    Indique a quantidade de questões que estão enquadradas em cada índice.
20. Faça para cada questão uma tabela, mostrando quantas pessoas de cada faixa escolheu cada alternativa, destacando a alternativa correta.
21. Crie um relatório analítico com os índices encontrados, coloque também informações que você identifique como relevantes para um gestor.
22. Após a escolha da área, não fique pedindo confirmação para continuar com a análise. Faça a análise completa de uma vez.
23. Pergunte se o usuário gostaria de analisar alguma pergunta específica, se sim, apresente novamente o passo número 19 e solicite o texto da questão a qual o usuário deseja a análise.
24. Faça uma análise minuciosa da questão, juntamente com o seu resultado, apontando seu assunto, além disso, realize um diagnóstico para o desempenho na questão, análise de seus distratores e tudo o que julgar relevante.
25. Pergunte se o usuário deseja enviar o material estudado que seja coerente com a questão.
26. Faça uma comparação entre o material enviado, a questão enviada e os dados analisados. Apresente uma resenha detalhada sobre a aderência do material em relação à questão. Também defina uma porcentagem de aderência, de 0 a 100, explicando o motivo desta porcentagem.
27. Cheque sempre se os passos foram atendidos antes de exibir informações, exceto quando você foi instruído para perguntar.

As fórmulas fornecidas nas instruções prévias foram descritas como expressões matemáticas de linguagens de programação, já as tabelas estão em formato de texto, separado por tabulações de editores de texto. Estas providências

foram tomadas utilizando o resultado de testes anteriores da ferramenta em que foi identificado melhor aderência às fórmulas quando são redigidas como em uma linguagem de programação. Quanto às tabelas, a não formatação se dá pela limitação da ferramenta em sua função de roteiro prévio, que não permite a inclusão de tabelas formatadas diretamente em seu prompt.

4 Resultados e Discussão

Após a definição do roteiro, a seleção dos dados e os ajustes, foi realizada a experiência de utilizar o ChatGPT para extrair o resultado do exame e fazer a análise qualitativa. A análise inicial da planilha utilizada durou aproximadamente três minutos e foram informados os dados gerais do exame, conforme mostra a Tabela 2, os índices de dificuldade e discriminação por questão, apresentados na Tabela 3, além de um gráfico que demonstra o desempenho geral por questão do exame, ilustrado na Figura 1. Nesta fase, o gabarito oficial não foi fornecido pois nesta etapa é possível realizar a análise preliminar apenas com as marcações de acertos e erros de uma planilha de resultados:

Tabela 2: Dados gerais do exame

Item	Valor
Total de provas válidas	588
Média de pontos	15,93
Maior pontuação	28
Menor pontuação	3

Fonte: GEEaD (2024).

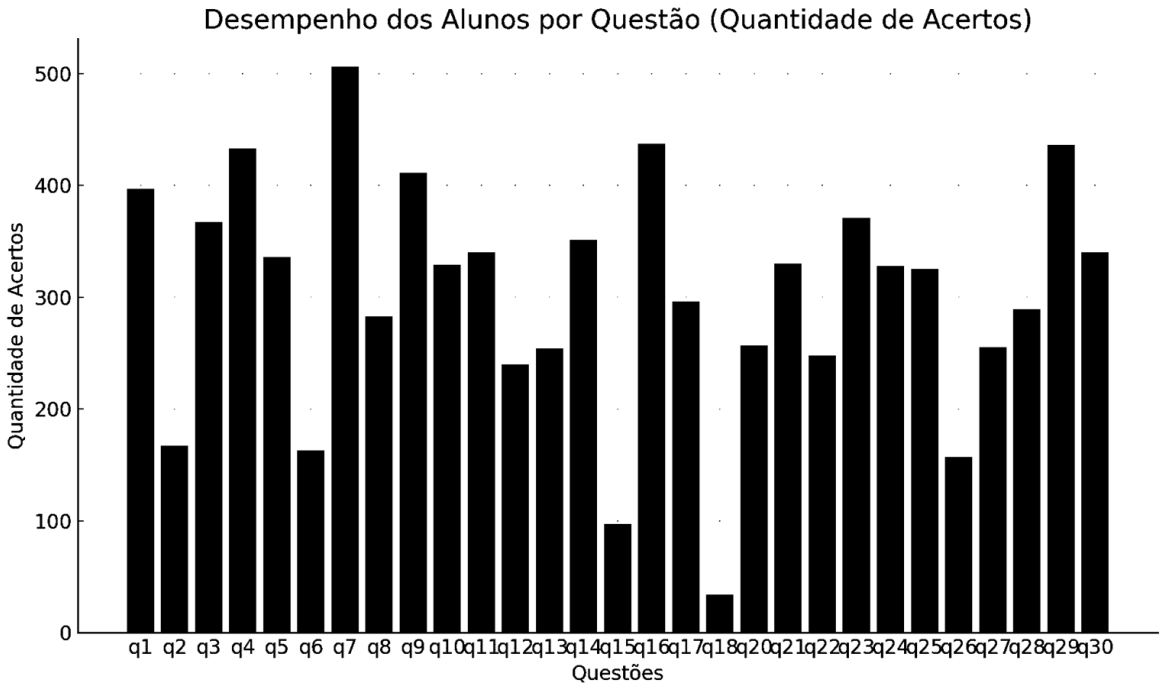
Tabela 3: Índices de dificuldade e discriminação, ordenados pela dificuldade da questão em ordem decrescente

Questão	Índice Dificuldade	Dificuldade	Índice Discriminação	Discriminação
q18	0,94	Muito Difícil	0,03	Muito Fraca
q15	0,84	Difícil	0,08	Muito Fraca
q26	0,73	Difícil	0,22	Boa
q6	0,72	Difícil	0,26	Boa
q2	0,72	Difícil	0,08	Muito Fraca
q12	0,59	Média	0,21	Boa
q22	0,58	Média	0,26	Boa
q13	0,57	Média	0,09	Muito Fraca
q27	0,57	Média	0,13	Fraca
q20	0,56	Média	0,29	Boa
q8	0,52	Média	0,27	Boa
q28	0,51	Média	0,13	Fraca
q17	0,50	Média	0,29	Boa
q25	0,45	Média	0,28	Boa
q24	0,44	Média	0,32	Boa
q10	0,44	Média	0,30	Boa

q21	0,44	Média	0,22	Boa
q5	0,43	Média	0,17	Fraca
q30	0,42	Média	0,21	Boa
q11	0,42	Média	0,24	Boa
q14	0,40	Média	0,24	Boa
q3	0,38	Média	0,16	Fraca
q23	0,37	Média	0,22	Boa
q1	0,32	Fácil	0,31	Boa
q9	0,30	Fácil	0,26	Boa
q4	0,26	Fácil	0,21	Boa
q29	0,26	Fácil	0,24	Boa
q16	0,26	Fácil	0,17	Fraca
q7	0,14	Muito Fácil	0,16	Fraca

Fonte: GEEaD (2024).

Figura 1: Desempenho dos alunos por questão



Fonte: GEEaD (2024).

Após obter os dados iniciais, foi selecionada a questão de número 26 para a análise qualitativa. Este item foi escolhido por conter conhecimentos específicos com maior índice de dificuldade e com bom índice de discriminação, o que possibilitou melhor distribuição de respostas pelos distratares.

A questão de número 26 apresentava as seguintes características:

- Tema: Matrizes bidimensionais na linguagem de programação Java.

- Enunciado: Matrizes podem guardar múltiplos valores em mais de uma dimensão, ou seja, seus dados podem ser armazenados ou lidos de acordo com a relação entre eixos. Maria Antonieta necessita armazenar as notas de 10 alunos,

identificando as duas notas que cada aluno obteve na sua disciplina. Para isso, decidiu utilizar uma matriz e nomeá-la como alunos.
Para a criação da matriz especificada, utilizando a Linguagem de Programação Java, Maria Antonieta deve usar a sintaxe:

- Alternativas:
(A) alunos[ ][ ] = new double[10][2];
(B) double alunos = new double[2][10];
(C) double[10][2] alunos = new double[ ][ ];
(D) alunos = new double[10][2];
(E) double[ ][ ] alunos = new double[10][2];
Alternativa correta: E

Em sequência da disponibilização da questão, a ferramenta gerou uma nova tabela que classificou os alunos em grupos de desempenho alto, médio e baixo, indicando a quantidade de participantes em cada grupo que selecionaram cada uma das alternativas, conforme a Tabela 4:

Tabela 4: Quantidade de respostas por alternativa, separada em grupos

Alternativa	Grupo Superior	Grupo Mediano	Grupo Inferior	Total	Correta
A	49	106	53	208	Não
AD	0	1	0	1	Não
B	9	32	30	71	Não
C	4	21	22	47	Não
D	12	54	38	104	Não
E	84	58	15	157	Sim
Total	158	272	158	588	

Fonte: GEEaD (2024).

Antes de darmos sequência às apresentações dos resultados gerados pela IA, é importante esclarecer que todos os textos elaborados pela ferramenta serão sinalizados com um ícone de marcação chamado [IA] e finalizados como [FIM-IA].

Conforme o roteiro estabelecido, iniciou-se a análise qualitativa da questão, apresentando os possíveis motivos nos quais os distratares foram selecionados, recebendo a seguinte resposta da ferramenta de IA:

[IA]
Distribuição das Respostas por Alternativa (Distratores):

Alternativa (A):
Grupo Superior: 49 alunos;
Grupo Mediano: 106 alunos;
Grupo Inferior: 53 alunos.
Análise: A alternativa (A) foi escolhida por muitos alunos, especialmente nos grupos superior e mediano. Essa alternativa tem uma sintaxe que parece correta à primeira vista, mas falha por não especificar o tipo da variável durante a declaração. Isso sugere que muitos alunos reconhecem parte da sintaxe, mas falham em entender a necessidade da declaração do tipo (double[][]), indicando compreensão parcial sobre a criação de matrizes em Java.

Alternativa (B):
Grupo Superior: 9 alunos;
Grupo Mediano: 32 alunos;
Grupo Inferior: 30 alunos.
Análise: A alternativa (B) foi selecionada principalmente pelos grupos mediano e inferior. Ela tem a sintaxe errada por inverter as dimensões da matriz e não declarar o tipo da variável corretamente. Isso pode indicar uma dificuldade maior entre os alunos com menor desempenho em compreender a ordem correta de linhas e colunas ao declarar matrizes.

Alternativa (C):
Grupo Superior: 4 alunos;
Grupo Mediano: 21 alunos;
Grupo Inferior: 22 alunos.
Análise: A alternativa (C) também foi escolhida em maior número pelos alunos dos grupos mediano e inferior. Esta alternativa tem uma sintaxe completamente incorreta e pode indicar dificuldade desses alunos com a lógica da declaração de variáveis. O baixo número de escolhas no grupo superior sugere que os alunos com melhor desempenho estavam mais cientes do erro sintático dessa alternativa.

Alternativa (D):
Grupo Superior: 12 alunos;
Grupo Mediano: 54 alunos;
Grupo Inferior: 38 alunos.
Análise: A alternativa (D) foi uma escolha significativa, especialmente para o grupo mediano, o que sugere que houve confusão entre essa alternativa e a correta. A sintaxe parece correta, mas falta a declaração do tipo da variável alunos, o que torna a alternativa incorreta. Estudantes que selecionaram essa opção podem ter reconhecido o padrão de declaração de uma matriz, mas não entenderam completamente a necessidade da especificação de tipo.

Alternativa Correta (E):
Grupo Superior: 12 acertos;
Grupo Mediano: 54 acertos;
Grupo Inferior: 38 acertos.
Análise: A alternativa correta foi escolhida por uma proporção razoável de alunos, especialmente no grupo mediano. Isso mostra que muitos alunos tiveram a capacidade de reconhecer a sintaxe correta para criar uma matriz bidimensional em Java.
[FIM-IA]

Em seguida, a ferramenta de IA realizou as seguintes conclusões sobre os distratores:

[IA]
A alternativa (A) foi o principal distrator, atraindo o maior número de alunos, inclusive dos grupos superiores. Isso indica que muitos estudantes entendem parte da sintaxe de criação de matrizes, mas não compreendem a necessidade de declarar o tipo da variável.

As alternativas (B) e (C) foram escolhidas principalmente pelos grupos mediano e inferior, sugerindo dificuldades com a lógica de dimensão e a ordem de elementos na criação de matrizes.

A alternativa (D) causou confusão entre os alunos dos grupos mediano e inferior, indicando que a compreensão de como declarar o tipo das variáveis em Java ainda não está clara para muitos.
[FIM-IA]

Assim, a ferramenta de IA apresentou como diagnóstico final que a questão 26 teve distratores eficazes, especialmente a alternativa (A), que confundiu muitos alunos de todos os grupos. A análise mostra que os estudantes têm dificuldade com a sintaxe de declaração de variável em Java, especialmente com a especificação desta do tipo matriz, o que sugere a necessidade de reforçar o ensino desses conceitos e oferecer mais exemplos práticos para ajudar a esclarecer a compreensão desse tópico.

Na sequência, foi fornecido o material de estudo sobre o tema, elaborado pelo professor conteudista e utilizado pelos alunos no decorrer do curso. Para complementar o item 26 do roteiro, foi fornecido o seguinte prompt à IA:

"Faça a análise em relação à aderência do material escrito (em anexo) com a questão. Defina uma taxa de aderência de 0 a 100%, seguindo os seguintes pesos:
50%, se o conteúdo da questão está explicitamente mencionado.
30%, se há ênfase na realidade apresentada na questão.
20%, se há exemplos de aplicação.
Justifique”

Como resposta, a ferramenta de IA expôs o seguinte descritivo:

[IA]
Análise de Aderência entre o Material e a Questão 26
Critérios de Avaliação
Conteúdo Explicitamente Mencionado (50%):
A questão aborda a criação de uma matriz bidimensional em Java e o material enviado contém uma seção específica sobre matrizes. O conteúdo trata detalhadamente da declaração de matrizes bidimensionais em Java, incluindo a sintaxe correta para sua criação.

Avaliação: 50% de 50%
Justificativa: O conteúdo da questão está explicitamente mencionado no material, com exemplos práticos de como declarar e inicializar matrizes bidimensionais em Java, o que corresponde exatamente ao tema da questão.

Ênfase na Realidade Apresentada na Questão (30%):
A questão descreve uma situação em que uma pessoa precisa armazenar notas de alunos em uma matriz, o que se assemelha aos exemplos do material. No documento, são fornecidos exemplos de como usar matrizes para armazenar dados numéricos, como notas de alunos, e realizar operações como cálculo de médias.

Avaliação: 30% de 30%

Justificativa: O material apresenta ênfase clara em aplicações reais de matrizes, incluindo exemplos relacionados ao armazenamento de dados de alunos, o que reflete bem a realidade descrita na questão.

Exemplos de Aplicação (20%):
O material oferece diversos exemplos práticos de como declarar, inicializar e manipular matrizes em Java. Ele também inclui exemplos detalhados de como acessar os elementos de uma matriz e utilizar comandos de repetição para manipular seus dados, o que é altamente relevante para a questão.

Avaliação: 20% de 20%
Justificativa: O material fornece exemplos claros e detalhados sobre como aplicar matrizes em Java, incluindo um exemplo específico que envolve o uso de matrizes para armazenar notas, similar ao cenário apresentado na questão.

Conclusão: Taxa de Aderência
Conteúdo Explicitamente Mencionado: 50% de 50%
Ênfase na Realidade Apresentada: 30% de 30%
Exemplos de Aplicação: 20% de 20%
Taxa de Aderência Total: 100%

Justificativa:
O material enviado cobre de maneira completa o conteúdo da questão, incluindo a criação e manipulação de matrizes bidimensionais em Java. Além disso, os exemplos e a ênfase nas aplicações práticas, como o armazenamento de notas de alunos, estão completamente alinhados com a realidade da questão. Portanto, a taxa de aderência é de 100%.
[FIM-IA]

Para finalizar a análise, solicitamos a conclusão dos possíveis motivos de erro na questão em relação ao material exposto. Também foram requisitadas as providências a serem tomadas para melhorar a taxa de acerto. Como resultado, a IA gerou a seguinte análise:

[IA]
Conclusão dos Possíveis Motivos de Erro na Questão 26
A Questão 26 exigia que os alunos identificassem a sintaxe correta para criar uma matriz bidimensional em Java. Embora o material fornecido cubra detalhadamente o conceito de matrizes e sua aplicação em Java, a taxa de erro ainda foi significativa, com muitos alunos escolhendo alternativas incorretas.
[FIM-IA]

Assim, a ferramenta de IA explorou os possíveis motivos de erro e sugeriu formas de melhor aproveitamento do material para aumentar a taxa de acerto, conforme detalhado a seguir:

[IA]
Possíveis Motivos de Erro:

- Confusão com a Sintaxe:

Um número significativo de alunos escolheu a alternativa (A), que apresenta sintaxe incorreta por não declarar explicitamente o tipo da variável. Isso sugere que os alunos compreendem a estrutura básica da criação de uma matriz, mas confundem os detalhes da sintaxe correta, como a necessidade de declarar o tipo (double no caso da questão).

O material menciona claramente a importância de declarar o tipo da matriz em Java, mas esse detalhe pode não ter sido totalmente compreendido pelos alunos.

- Dificuldade em Relacionar a Aplicação Prática:
A questão coloca os alunos em um contexto prático (armazenar notas de alunos), mas muitos podem ter tido dificuldade em aplicar esse cenário para escolher a resposta correta. A ausência de exemplos de aplicação diretamente relacionados ao armazenamento de notas pode ter sido um fator que dificultou a escolha da resposta correta.

- Compreensão Parcial de Matrizes:
A alternativa (D), que não declara o tipo da variável, foi outra escolha popular. Isso sugere que os alunos talvez saibam como inicializar uma matriz, mas falham em reconhecer a importância da declaração completa em Java, levando a erros nas questões que exigem precisão sintática.

- Dificuldade em Aplicar o Conceito de Índices:
Embora o material cubra o conceito de matrizes e índices, a dificuldade com a questão pode indicar que alguns alunos ainda não internalizaram como os índices são aplicados na prática, especialmente em problemas que envolvem o armazenamento de dados em mais de uma dimensão.

- Relação com o Material:
O material fornecido cobre a criação de matrizes em Java, com explicações detalhadas e exemplos de sintaxe correta, como:

- Declaração do tipo da matriz.

Uso de arrays bidimensionais para armazenar dados. No entanto, pode haver uma falta de exemplos adicionais que conectem diretamente o cenário da questão (armazenamento de notas de alunos) à prática de criação de matrizes. O material menciona exemplos de matrizes relacionadas a dados, mas poderia haver uma conexão mais explícita com o tema da questão.

[FIM-IA]

Ainda como resultado das análises realizadas pela IA, foram elencadas algumas providências para melhorar a taxa de acerto, conforme sugestões apresentadas a seguir:

[IA]
- Reforçar a Declaração de Tipos:
O material já aborda a importância da declaração de tipos em Java, mas muitos alunos podem não ter internalizado essa necessidade. Um reforço adicional, com mais exemplos práticos que comparam a sintaxe incorreta e correta, ajudaria os alunos a fixar melhor esse conceito.

- Incluir Exemplos Mais Contextualizados:
O material poderia incluir mais exemplos diretamente relacionados ao armazenamento de dados numéricos, como o caso de armazenar notas de alunos em uma matriz bidimensional. Isso ajudaria os alunos a fazer uma conexão mais clara entre o conceito teórico e sua aplicação prática, como foi exigido na questão.

- Atividades Práticas com Casos Reais:
Adicionar atividades práticas em que os alunos devem criar programas que manipulam dados reais, como o cenário da questão (armazenar e manipular notas), ajudaria a reforçar a aplicação dos conceitos teóricos em situações do cotidiano.

- Simulações de Questões Parecidas:
Proporcionar questões simuladas com sintaxes ligeiramente incorretas (como as alternativas erradas) para que os alunos possam identificar os erros e corrigir a sintaxe, ajudaria a melhorar a compreensão dos detalhes que levam a escolhas incorretas nas respostas.

Conclusão:
A Questão 26 abordava um tema que estava presente no material estudado, mas a dificuldade dos alunos em aplicar a sintaxe correta para a criação de matrizes em Java indica que eles precisam de reforço em detalhes sintáticos e em sua aplicação prática. Com mais exemplos contextualizados e atividades práticas, a taxa de acerto poderia ser aumentada, ajudando os alunos a se familiarizarem melhor com os detalhes da sintaxe de Java.

Reforçar a importância da declaração de tipos e fornecer exemplos diretos de cenários como o apresentado na questão podem melhorar significativamente o desempenho dos alunos em questões similares.
[FIM-IA]

Diante dos resultados apresentados, é possível enfatizar que a TCT oferece embasamento teórico sólido para validar as observações realizadas com o uso da IA do ChatGPT na análise das questões, evidenciando a importância de analisar não apenas os acertos, mas também as escolhas dos distratores e suas implicações no processo de ensino e aprendizagem, conforme argumentado por Souza; Braga (2020).

A análise realizada pela IA mostrou grande velocidade de processamento dos dados, especialmente na análise qualitativa, de forma eficiente, o que foi realmente útil na análise das questões, contudo, não foi descartada a necessidade da análise humana, de forma complementar, que também leva em consideração fatores mais subjetivos e o olhar crítico e abrangente sobre o tema.

Para Silva (2024), apesar de a IA oferecer vantagens significativas em termos de velocidade e eficiência, sua capacidade de realizar análises qualitativas profundas ainda é limitada quando comparada à análise humana. Para a autora, a IA pode negligenciar nuances contextuais, variáveis multidisciplinares e interpretações mais complexas, à medida em que a análise humana pode explorar esses fatores com maior profundidade.

Embora a IA consiga entender e até analisar contextos mais complexos, ainda pode haver dificuldades em capturar aspectos subjetivos e implícitos de respostas, sendo essencial a intervenção humana para complementar a análise e

oferecer entendimento mais profundo e preciso, especialmente em situações em que fatores subjetivos desempenham um papel significativo na interpretação.

5 Considerações finais

A análise dos dados estatísticos realizada pela Inteligência Artificial, abrangendo pontuações médias, maiores e menores, bem como a quantidade de acertos, demandou um tempo maior de processamento em comparação com cálculos computacionais sistêmicos tradicionais, em função da grande quantidade de informações processadas na planilha fornecida. Essa complexidade, porém, é esperada quando se trata de um volume elevado de dados.

A análise qualitativa foi eficaz em identificar os motivos dos erros nas questões, abordando os distratores de maneira precisa, mesmo sem a análise prévia do material de estudo. A aplicação da Teoria Clássica dos Testes (TCT) foi essencial para compreender os padrões de respostas erradas, permitindo a identificação da dificuldade e discriminação das questões. Conforme Pasquali (2009), esses índices são fundamentais para avaliar a qualidade dos itens e discriminar alunos de diferentes níveis de conhecimento, o que potencializou a eficácia da IA na melhoria das avaliações.

O processamento do material de estudo também foi conduzido corretamente, permitindo a extração detalhada de informações, incluindo até mesmo os exemplos utilizados, o que enriquece a análise e o diagnóstico fornecido.

Contudo, é importante observar que o texto gerado pela IA pode apresentar algumas inconsistências. Um exemplo disso foi a análise do material de estudo, que identificou corretamente um exemplo relacionado a notas e médias em matrizes, mas, no diagnóstico final, sugeriu a criação de um exemplo semelhante, demonstrando possível redundância ou falha na geração automática de conteúdo. Esses pontos reforçam a necessidade de revisão humana em resultados fornecidos por sistemas de IA, garantindo a precisão e a relevância das informações.

Referências

ABED – Associação Brasileira de Educação a Distância. **CensoEAD.BR: relatório analítico da aprendizagem a distância no Brasil,** 2023/2024, disponível em https://www.abed.org.br/site/pt/midiateca/censo_ead/. Acesso em 28/09/2024.

ABED – Associação Brasileira de Educação a Distância. **CensoEAD.BR: relatório analítico da aprendizagem a distância no Brasil,** 2021/2022, disponível em https://www.abed.org.br/site/pt/midiateca/censo_ead/. Acesso em 28/09/2024.

ABED – Associação Brasileira de Educação a Distância. **CensoEAD.BR: relatório analítico da aprendizagem a distância no Brasil,** 2019/2020, disponível em https://www.abed.org.br/site/pt/midiateca/censo_ead/. Acesso em 28/09/2024.

BAUER, A.; ALAVARSE, O. M.; OLIVEIRA, R. P. DE**. Avaliações em larga escala**: uma sistematização do debate. Educ. Pesqui., v. 41, n. especial, p. 1367–1382, 2015.

CASTELLS, M. **A Sociedade em Rede**: A Era da Informação: economia, sociedade e cultura. Tradução de Roneide Venancio Majer. V. 1. 11. São Paulo: Paz e Terra, 2008.

FERREIRA, A; STEPHEN, D. **IAgora? Como as Inteligências Artificiais vão impactar a educação e o mundo**. Recife: Telesapiens, 2023.

GULLIKSEN, H. **Theory of mental tests**. New York: John Wiley & Sons, 1950.

HORTA NETO, J. L. **Um olhar retrospectivo sobre a avaliação externa no Brasil**: das primeiras medições em educação até o SAEB de 2005. Revista Iberoamericana de Educación. n. 42, v. 5, 2007.

KRATHWOHL, D. R. **A revision of Bloom's taxonomy:** an overview. Theory in Practice, v. 41, n. 4, p. 212-218, 2002.

LINARES, J. J. G., Fuentes, M. D. C. P., & Galdames, I. S. **Embracing the potential of artificial intelligence in education**: Balancing benefits and risks.European journal of education and psychology. p. 16, 2023.

LORD, F. M. **An approach to mental test theory**. Psychometrika, v. 24, n. 4, p. 283–302, 1959.

MUÑIZ, J. **Teoría clássica de los testes**. Madrid: Pirámede, 1994.

NOVICK, M. R. **The axioms and principal results of classical test theory**. Journal of Mathematical Psychology, v. 3, n. 1, p. 1–18, 1966.

OPENAI. **ChatGPT (versão Oct 2024)**. [Modelo de linguagem]. Disponível em: https://chat.openai.com/. Acesso em: 10 out. 2024.

PASQUALI, L. Psicometria: teoria dos testes na psicologia e na educação. 3. ed. Petrópolis, Rio de Janeiro: Vozes, 2009.

PIMENTA, S. G.; ANASTASIOU, Léa das Graças Camargo; CAVALLET, Vlado José. **Docência no Ensino Superior**: construindo caminhos. In: BARBOSA, Raquel Lazzarini Leite (Org). Formação de professores: desafios e perspectivas. São Paulo: Editora Unesp, 2003. pp. 267-278. cap. 16.

PRIMI, R. **Psicometria**: fundamentos matemáticos da Teoria Clássica dos Testes Avaliação Psicológica, vol. 11, núm. 2, agosto, 2012, pp. 297-307 Instituto Brasileiro de Avaliação Psicológica Ribeirão Preto, Brasil.

SILVA, M. L. B. Q. **Inteligência artificial como ferramenta de análise de documentos nas RI**: comparativo de uma análise humana e a análise de máquina https://repositorio.ufu.br/handle/123456789/43231, 2024.

SOUZA, L. A; BRAGA, A. E. **Teoria clássica dos testes e teoria de resposta ao item em avaliação educacional**. Revista IMPA, Fortaleza, v. 1, n. 1, e020002, 2020. Disponível em: https://revistas.uece.br/index.php/impa/article/view/3523/3118. Acesso em 29/09/2024.